
Powered by
architect
East
Version 6.4
Cytel
Copyright © Cytel Inc., 1994-2016. All rights reserved.
This program is protected by US and International copyright laws.

# Preface

## Acknowledgments

Welcome to East 6.2, a software package for the design, simulation and interim monitoring of clinical trials. East is the industry standard for designing adequate and well-controlled clinical trials in accordance with FDA and EMA guidances.

The current release of East (version 6.2) is powered by the Architect platform. It was developed by a team consisting of (in alphabetical order of last names): Gayatri Bartake, Sudipta Basu, Apurva Bhingare, Pushkar Borkar, Bristi Bose, Chandrashekhar Budhwant, V. P. Chandran, Aniruddha Deshmukh, Namita Deshmukh, Namrata Deshpande, Yogesh Dhanwate, Monika Ghatage, Vishal Gujar, Shashikiran Halvagal, Karen Han, Pravin Holkar, Abhijit Jadhav, Yannis Jemiai, Ashwini Joshi, Nilesh Kakade, Anthiyur Kannappan, Kapildev Koli, Yogita Kotkar, Niranjan Kshirsagar, Hrishikesh Kulkarni, Kaushal Kulkarni, Mandar Kulkarni, Mangesh Kulkarni, Shailesh Kulthe, Nilesh Lanke, Manisha Lohokare, Charles Liu, Lingyun Liu, Shashank Maratkar, Cyrus Mehta, Abdulla Mulla, Nabeela Muzammil, Seema Nair, Neelam Nakadi, Atul Paranjape, Vidyadhar Phadke, Ashvinikumar Pinjarkar, Shital Pokharkar, Vidyagouri Prayag, Misha Salganik, Makarand Salvi, Pralay Senchaudhuri, Brian Sharkey, Priyadarshan Shinde, Sheetal Solanki, Chitra Tirodkar, Amrut Vaze, Suryakant Walunj, Suchita Wageshwari, Ritika Yadav, Sanhita Yeolekar.

Other contributors who worked on previous releases of East: Ujwala Bamishte, Dhaval Bapat, Krisnaiah Byagari, Vibhavari Deo, Yogesh Deshpande, Pranab Ghosh, Ketan Godse, Aarati

Hasabnis, Jaydip Mukhopadhyay, Sandhya Paranjpe, Nabarun Saha, Abhijit Shelar.

Others who provided valuable assistance in this release are Asmita Ghatnekar, Ajay Sathe, Rhiannon Sheapare, and Tammy Sneddon.

A large number of people, other than those listed above, have helped us through the years with the various releases of East. We received valuable assistance from Sandro Pampallona, Mandar Kale, David Bristol, Yogesh Gajjar, Rajesh Mehta, Dhanashri Pathak, Vivek Pradhan, Vipul Suru, Nitin Patel, and others.

East draws on and extends the pioneering research of

> Peter Armitage, Peter Bauer, Werner Brannath, David DeMets, Tom Fleming, Christopher Jennison, Kyungmann Kim, Gordon Lan, Hans-Helge Müller, Peter O'Brien, Sandro Pampallona, Stuart Pocock, Martin Posch, Helmut Schäfer, Daniel Scharfstein, Anastasios Tsiatis, and Bruce Turnbull.

Special credit should also be given to Sue-Jane Wang, James Hung and Robert O'Neill of the Center for Drug Evaluation at the FDA. These investigators have performed original research on the design of adaptive trials and have been instrumental in creating an atmosphere of scientific rigor for the regulatory submissions of such trials.

The textbooks, "Group Sequential Methods with Applications to Clinical Trials" by Christopher Jennison and Bruce Turnbull (Chapman and Hall/CRC, 2000), and "Statistical Monitoring of Clinical Trials: A Unified Approach" by Michael Proschan, Gordon Lan, and Janet Wittes (Springer, 2006) are excellent complements to the East software.

We express our gratitude to

> David DeMets, Chris Jennison, Kyungmann Kim, Tony Lachenbruch, Anastasios Tsiatis and Bruce Turnbull

for agreeing to serve as members of the East Advisory Committee.

We thank all our beta testers for their input and obvious enthusiasm for the East software. They are acknowledged by name in Appendix **??**.

We owe a debt of gratitude to Marvin Zelen and to Swami Sarvagatananda, special people whose wisdom, encouragement and generosity have inspired Cytel for over two decades.

Finally, we dedicate this software package to our families and to the memory of our dearly departed Stephen Lagakos and Aneesh Patel.

## Our Philosophy

We would like to share with you what drives and inspires us during the research and development stages of the East software.

### Empower, do not Frustrate

We believe in making simple, easy-to-use software that empowers people.

We believe that statisticians have a strategic role to play within their organization and that by using professionally developed trial design software they will utilize their time better than if they write their own computer programs in SAS or R to create and explore complex trial designs. With the help of such software they can rapidly generate many alternative design options that accurately address the questions at hand and the goals of the project team, freeing time for strategic discussions about the choice of endpoints, population, and treatment regimens.

We believe that software should not frustrate the user's attempt to answer a question. The user experience ought to engage the statistician and inspire exploration, innovation, and the quest for the best design. To that end, we believe in the following set of principles:

- Fewer, but Important and Useful Features It is better to implement fewer, but important and useful features, in an elegant and simple-to-use manner, than to provide a host of options that confuse more than they clarify.

As Steve Jobs put it: 'Innovation is not about saying "Yes" to everything. It's about saying "No" to all but the most crucial features.'

- Just because we Can, doesn't mean we Should Just because we can provide functionality in the software, doesn't mean we should.
- Simplify, Simplify, Simplify Find and offer simple solutions - even for the most complex trial design problems.
- Don't Hurry, but continually Improve Release new solutions when they are ready to use and continually improve the commercial releases with new features, bug fixes, and better documentation.
- Provide the best Documentation and Support Our manuals are written like textbooks, to educate, clarify, and elevate the statistical knowledge of the user.
  Our support is provided by highly competent statisticians and software engineers, focusing on resolving the customer's issue, and being mindful of the speed and quality requirements. We believe that delivering *delightful customer support is essential to our company's lifeblood*.

Finally, we listen to our customers constantly and proactively through countless informal and formal interactions, software trainings, and user group meetings. This allows us to follow all the principles laid out above in the most effective manner.

# Table of Contents

## Table of Contents

# Volume 1 The East System

# 1 Installing East 6.2

## 1.1 System Requirements to run East 6.2

The minimum hardware/operating system requirements for East 6.2 are:

- A system running one of the following operating systems:
  - Windows XP (32 or 64 bit)
  - Windows 7 (32 or 64 bit)
- A minimum of 512 MB RAM (1 GB recommended)
- A hard disk with at least 300 MB of free disk space

## 1.2 Installation

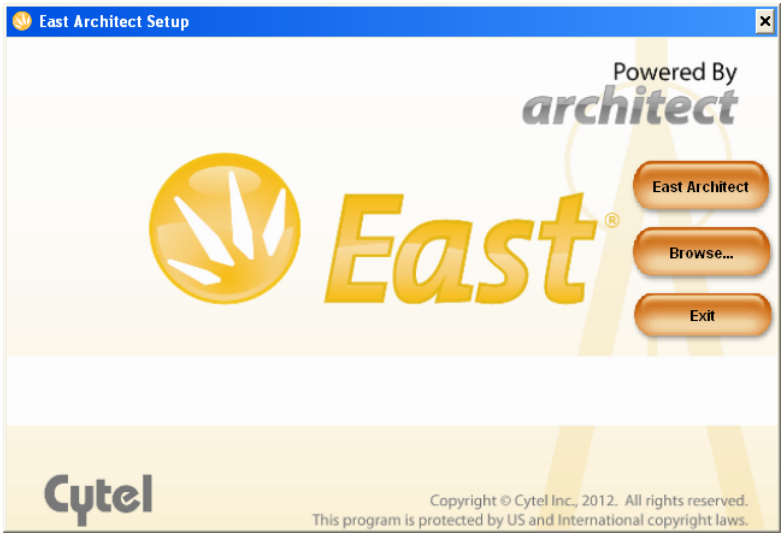
To install East 6.2, please follow these steps:

1. If any copy (including a beta or demo version) of East 6.2 is currently installed on your PC, please uninstall it or else the installation of the current version will not proceed correctly. To uninstall the earlier version of East 6.2, go to the **`Start`** Menu and select **`Programs`**→ **`East 6.2`** → **`Uninstall East 6.2`**

2. Insert the East 6.2 CD into your CD-drive.

   (a) If your Windows Autorun Default is already active, you'll see an installation screen similar to what is shown below. Follow the instructions that will appear on the

screen.




(b) If your Windows Autorun Default is not active, you won't see any installation screen. In that case, open your Windows Explorer, click on the CD Drive, and double-click on **`Setup`**. Then you will see the installation screen . Follow the instructions that will appear on the screen.

# 2 Getting Started

East has evolved over the past several years with MS Excel® as the user interface. The East on MS Excel® did not integrate directly with any other **Cytel** products. Under the **Architect** platform, East is expected to coexist and integrate seamlessly with other **Cytel** products such as SiZ, and Compass. Architect is a common platform designed to support various Cytel products. It provides a user-friendly, Windows-standard graphical environment, consisting of tabs, icons, and dialog boxes, with which you can design, simulate and analyze. Throughout the user manual, this product is referred to as East 6.

One major advantage of East 6 is the facility for creating multiple designs. This is achieved by giving multiple inputs of the parameters as either comma separated, or in a range such as **(a:b:c)** with **a** as the initial value, **b** as the last value and **c** as the step size. If you give multiple values for more than one parameter, East creates all possible combinations of the input parameters. This is an immense advancement over earlier versions of East, where you had to create one design at a time. Furthermore, one could not compare different types of designs (e.g., superiority vs. noninferiority designs). Similarly, graphical comparison of designs with different numbers of looks was difficult with earlier versions of East. All such comparisons are readily available in East 6.

We have also provided powerful data editors to create, view, and modify data. A wide variety of statistical tests are now a part of East 6, which enables you to conduct statistical analysis of interim data for continuous, discrete and time to event endpoints.

Simulations help to develop better insight into the operating characteristic of a design. In East 6, the simulation module has now been enhanced to allow fixed or random allocation to treatment and control, and different sample sizes. Such options were not possible with earlier versions of East.

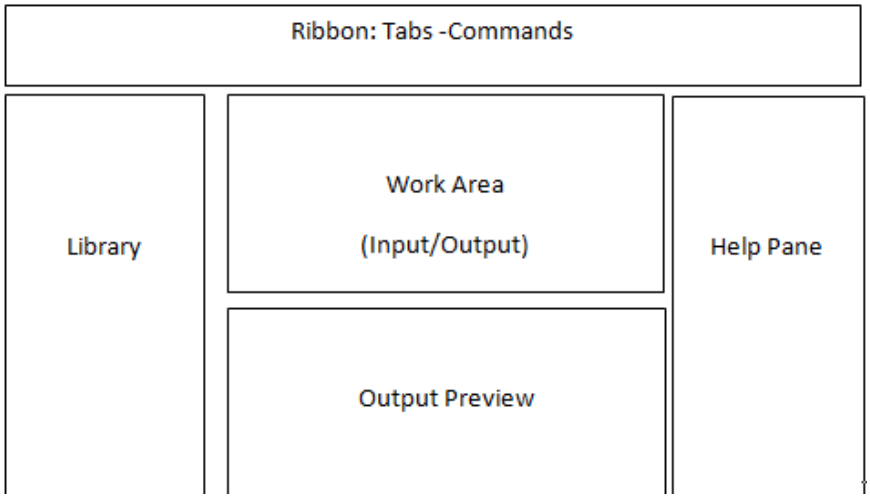
Another new feature is the option to add assumptions for accruals and dropouts. Previously this was available for survival, but has been extended to continuous and discrete endpoints in East 6. Information about accrual rates, response lag, and dropouts can be given whether designing or simulating a trial. This makes more realistic, end-to-end simulation of a trial possible.

The purpose of this chapter is to familiarize you with the East 6 user interface.

## 2.1 Workflow in Architect

In this section, the structure of Architect platform is explained. The logical workflow in which the different parts of the interface co-ordinate with each other is discussed.

The basic structure of the interface items is depicted in the following diagram.




Besides the top **Ribbon**, there are mainly four main windows in East 6 namely, (starting from left), the **Library**, the **Input / Output** window, the **Output Preview** area and the **Help Pane**. Note that both the **Library** and the **Help Pane** can be auto-hidden temporarily or throughout the session, allowing the other windows to occupy larger area on the screen for display.

Initially, **Library** shows only the **Root** node. As you work with East, multiple designs, simulation scenarios, data sets and related analysis can be managed using this panel. Various nodes for outputs and plots are created in the **Library**, facilitating work on multiple scenarios at a time. The width of the **Library** window can be adjusted for better readability.

The central part of the interface, the **Input / Output**, is the main work area where you can-

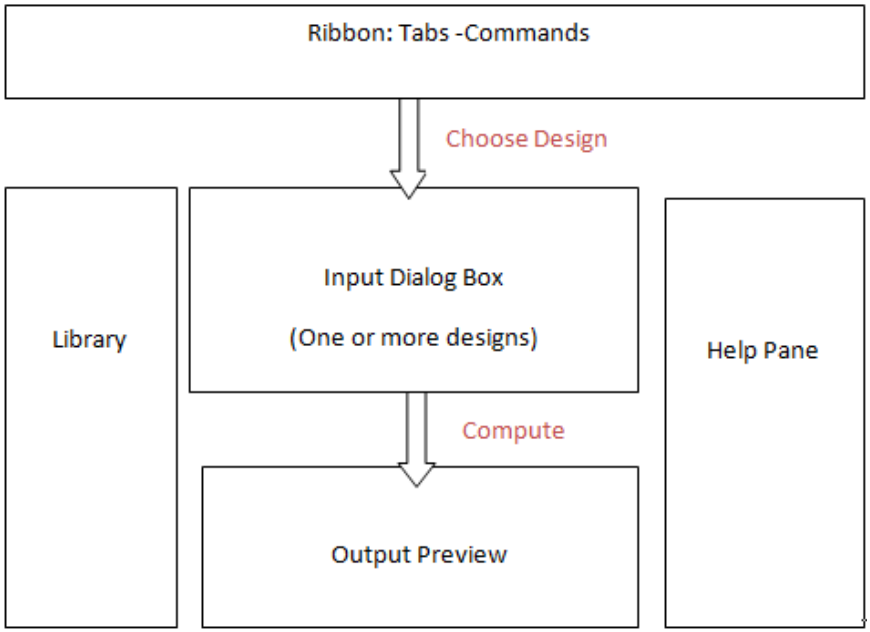
- Enter input parameters for design computation create and compare multiple designs, view plots
- Simulate a design under different scenarios
- Perform interim analysis on a group sequential design look by look and view the results, receive decisions such as stopping or continuing during the execution of a trial
- Open a data on which you want to perform analysis, enter new data, view outputs, prepare a report etc.

This is the area where the user interacts with the product most frequently.

The **Output Preview** area compiles several outputs together in a grid like structure where each row is either a design or simulation run. This area is in use only when working with Design or Simulations.

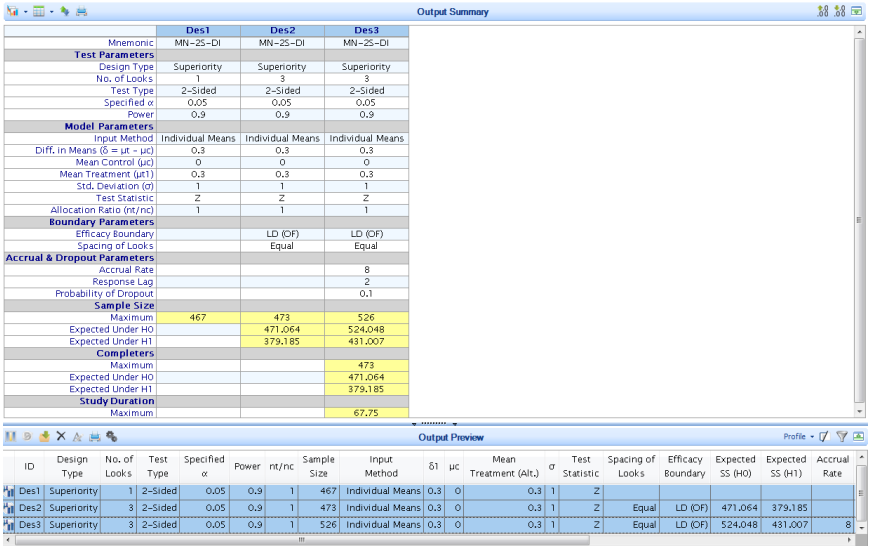
## Chapter 2: *Getting Started*

When the **Compute** or **Simulate** button is clicked, all requested designs or simulation results are computed and are listed in rows in the **Output Preview** area:




By clicking different rows of interest while simultaneously holding the **Ctrl** key, either a single or multiple designs can be displayed in the **Output Summary** in vertical manner or

side-by-side comparison can be done.



Output Summary

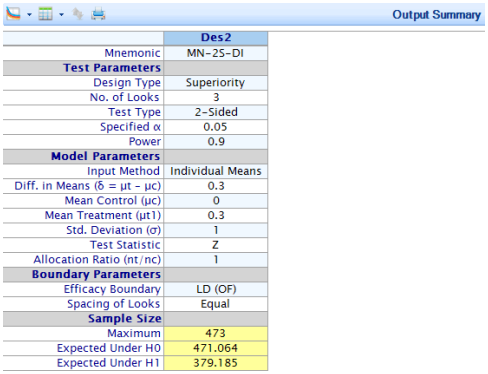
| | Des1 | Des2 | Des3 |
|---|---|---|---|
| Mnemonic | MN-2S-DI | MN-2S-DI | MN-2S-DI |
| **Test Parameters** | | | |
| Design Type | Superiority | Superiority | Superiority |
| No. of Looks | 1 | 3 | 3 |
| Test Type | 2-Sided | 2-Sided | 2-Sided |
| Specified α | 0.05 | 0.05 | 0.05 |
| Power | 0.9 | 0.9 | 0.9 |
| **Model Parameters** | | | |
| Input Method | Individual Means | Individual Means | Individual Means |
| Diff. in Means (δ = μt - μc) | 0.3 | 0.3 | 0.3 |
| Mean Control (μc) | 0 | 0 | 0 |
| Mean Treatment (μt1) | 0.3 | 0.3 | 0.3 |
| Std. Deviation (σ) | 1 | 1 | 1 |
| Test Statistic | Z | Z | Z |
| Allocation Ratio (nt/nc) | 1 | 1 | 1 |
| **Boundary Parameters** | | | |
| Efficacy Boundary | | LD (OF) | LD (OF) |
| Spacing of Looks | | Equal | Equal |
| **Accrual & Dropout Parameters** | | | |
| Accrual Rate | | | 8 |
| Response Lag | | | 2 |
| Probability of Dropout | | | 0.1 |
| **Sample Size** | | | |
| Maximum | 467 | 473 | 526 |
| Expected Under H0 | | 471.064 | 524.048 |
| Expected Under H1 | | 379.185 | 431.007 |
| **Completers** | | | |
| Maximum | | | 473 |
| Expected Under H0 | | | 471.064 |
| Expected Under H1 | | | 379.185 |
| **Study Duration** | | | |
| Maximum | | | 67.75 |

Output Preview

Profile

| ID | Design Type | No. of Looks | Test Type | Specified α | Power | nt/nc | Sample Size | Input Method | δ1 | μc | Mean Treatment (Alt.) | σ | Test Statistic | Spacing of Looks | Efficacy Boundary | Expected SS (H0) | Expected SS (H1) | Accrual Rate |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Des1 | Superiority | 1 | 2-Sided | 0.05 | 0.9 | 1 | 467 | Individual Means | 0.3 | 0 | 0.3 | 1 | Z | | | | | |
| Des2 | Superiority | 3 | 2-Sided | 0.05 | 0.9 | 1 | 473 | Individual Means | 0.3 | 0 | 0.3 | 1 | Z | Equal | LD (OF) | 471.064 | 379.185 | |
| Des3 | Superiority | 3 | 2-Sided | 0.05 | 0.9 | 1 | 526 | Individual Means | 0.3 | 0 | 0.3 | 1 | Z | Equal | LD (OF) | 524.048 | 431.007 | 8 |

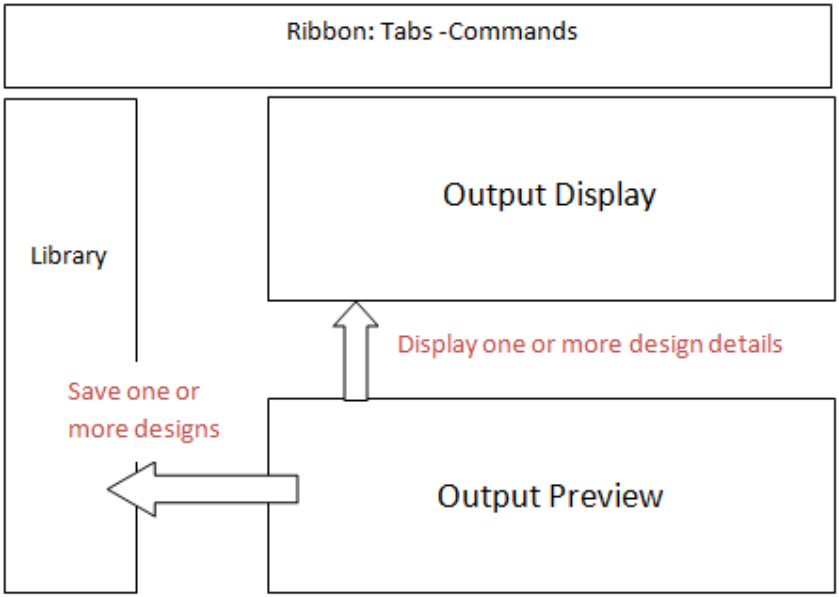
Note that the active window and the **Output Preview** can be minimized, maximized, or resized. If you want to focus on the **Output Summary**, click the  icon in the top-right corner of the main window. The Output will be maximized as shown below:



Output Summary

| | Des2 |
|---|---|
| Mnemonic | MN-2S-DI |
| **Test Parameters** | |
| Design Type | Superiority |
| No. of Looks | 3 |
| Test Type | 2-Sided |
| Specified α | 0.05 |
| Power | 0.9 |
| **Model Parameters** | |
| Input Method | Individual Means |
| Diff. in Means (δ = μt - μc) | 0.3 |
| Mean Control (μc) | 0 |
| Mean Treatment (μt1) | 0.3 |
| Std. Deviation (σ) | 1 |
| Test Statistic | Z |
| Allocation Ratio (nt/nc) | 1 |
| **Boundary Parameters** | |
| Efficacy Boundary | LD (OF) |
| Spacing of Looks | Equal |
| **Sample Size** | |
| Maximum | 473 |
| Expected Under H0 | 471.064 |
| Expected Under H1 | 379.185 |

Any of the designs/simulations in the **Output Preview** area can be saved in the **Library**, as depicted in the following workflow diagram.



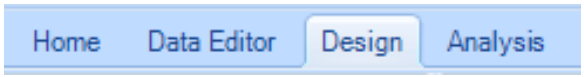

Double click any of these nodes and the detailed output of the design will be displayed. This will include all relevant input and output information. Right clicking any design node in the **Library** will allow you to perform various operations on the design such as interim monitoring and simulation.

The **Help Pane** displays the context sensitive help for the control currently under the focus. This help is available for all the controls in the **Input / Output** window. This pane also displays the design specific help which discusses the purpose of the selected test, the published literature referred while developing it and finally the user manual references to quickly look-up for more details in the East6 User Manual. This pane can be hidden or locked by clicking the pin symbol.

All the windows and features mentioned above are described in detail with the help of an illustration in the subsequent sections of this chapter.

## 2.2 A Quick Overview of User Interface

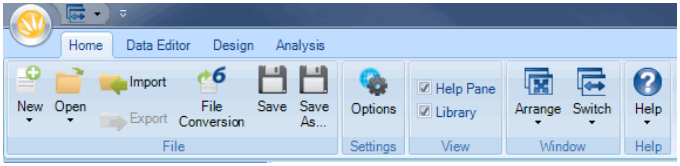
Almost all the functionalities of East 6 are invoked by selecting appropriate menu items and icons from the **Ribbon**. The interface consists of four windows as described in the previous section and four major menu items. These menu items are:




- **Home**. This menu contains typical file-related Windows sub-menus. The **Help** sub-menu provides access to this manual.
- **Data Editor**. This menu will be available once a data set is open, providing several sub-menus used to create, manage and transform data.
- **Design**. This menu provides a sub-menu for each of the study designs which can be created using East 6. The study designs are grouped according to nature of the response. The tasks like Simulations and Interim Monitoring are available for almost all the study designs under this menu.
- **Analysis**. This menu provides a sub-menu for each of the analysis procedure that can be carried out in East 6. The tests are grouped according to the nature of the response. There are also options for basic statistics and plots.

## 2.3 Home Menu

The **Home** menu contains a variety of submenus:




### 2.3.1 File


Click this icon to create new case data or crossover data. A new workbook or log can also be created.


Click this icon to open a saved data set, workbook, or log file.


Click this icon to import external files created by other programs.


Click this icon to export files in various formats.


Click this icon to save the current files or workbooks.


Click this icon to save a file or workbook with different name.

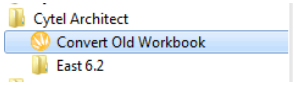
### 2.3.2 Importing workbooks from East5.4

East allows the conversion of workbooks previously created in East 5.4 (and above) to be imported into East 6 for further development. In order to open a workbook with the `.es5` extension given by previous versions of East, it must first be converted to a file with the `.cywx` extension that will be recognized by East 6. This is easily accomplished through the **Covert**

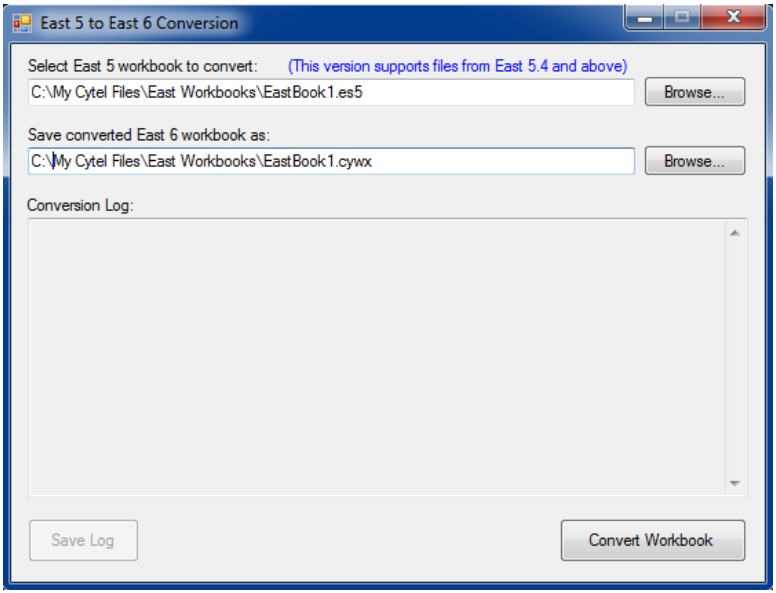
**Old Workbook** utility. Click on the 

icon under **Home** menu to see the location of this utility.

From the Windows **Start** menu under **All Programs**, select **Covert Old Workbook** located in the **Cytel Architect** folder:




We can see the following window which accepts East5.4 workbook as input and outputs a workbook of East6. Click the **Browse** buttons to choose the East 5.4 file to be converted and the file to be saved with **`.cywx`** extension of East 6 version. To start the conversion click **Convert Workbook**:

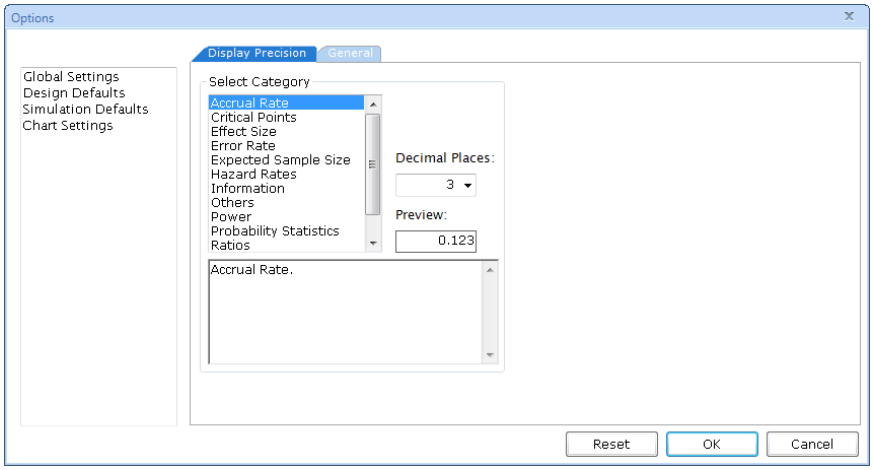



Once complete, the file can be opened as a workbook in East 6 through **`Home`**–> **`File`**–> **`Open`**.

### 2.3.3 Settings

Click the 

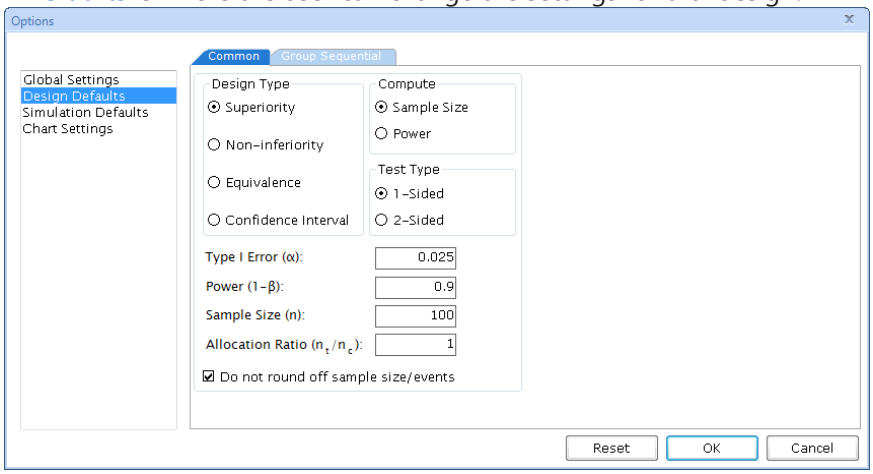
icon in the **Home** menu to adjust default values in East 6.




The options provided in the **Display Precision** tab are used to set the decimal places of numerical quantities. The settings indicated here will be applicable to all tests in East 6 under the **Design** and **Analysis** menus.

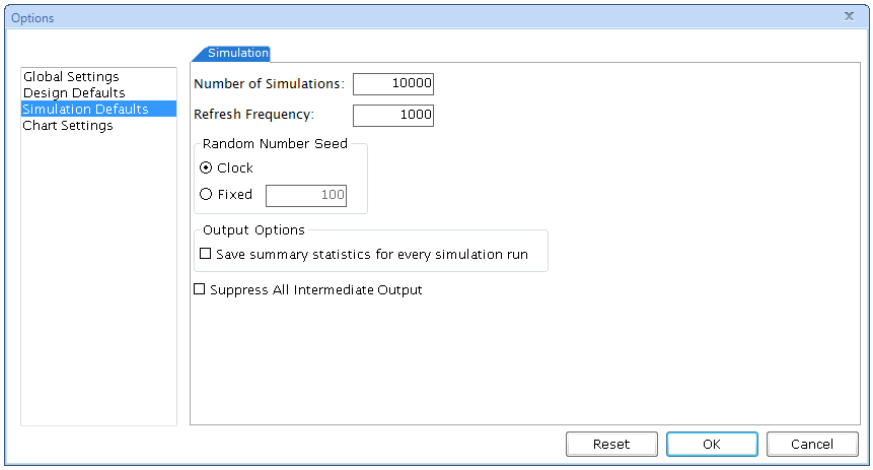
The **General** tab has the provision of adjusting the paths for storing workbooks, files, and temporary files. These paths will remain throughout the current and future sessions even after East is closed. This is the place where we need to specify the installation directory of the R software in order to use the feature of R Integration in East6.

The **Design Defaults** is where the user can change the settings for trial design:

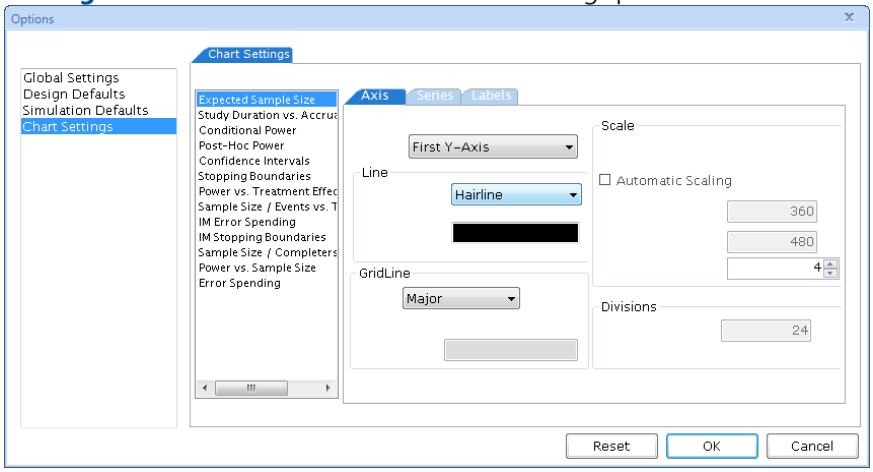



Under the **Common** tab, default values can be set for input design parameters. Under the **Group Sequential** tab, defaults are set for boundary information. When a new design is started, input fields will contain these specified defaults.

**Simulation Defaults** is where we can change the settings for simulation:

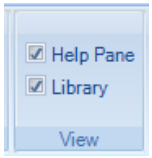

Options
Global Settings
Design Defaults
Simulation Defaults
Chart Settings
Simulation
Number of Simulations: 10000
Refresh Frequency: 1000
Random Number Seed
Clock
Fixed 100
Output Options
Save summary statistics for every simulation run
Suppress All Intermediate Output
Reset
OK
Cancel

The **Chart Settings** allows defaults to be set for the following quantities on East6 charts:




## 2.3.4 View

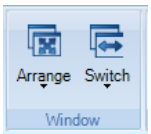
The **View** submenu consists of enabling or disabling the **Help** and **Library** panes by (un)checking the respective check boxes.




## 2.3.5 Window

The **Window** submenu contains an **Arrange** and **Switch** option. This provides the ability to view different standard arrangements of available windows (Design Input Output, Log, Details,

charts and plots) and to switch the focus from one window to another.




### 2.3.6 Help

The **Help** submenu provides the following ways to access the East6 documenatation:


- **User Manual**: Invoke the current East 6 user manual.
- **Tutorial**: Invoke the available East 6 tutorials.
- **About East 6**: Displays the current version and license information for the installed software.

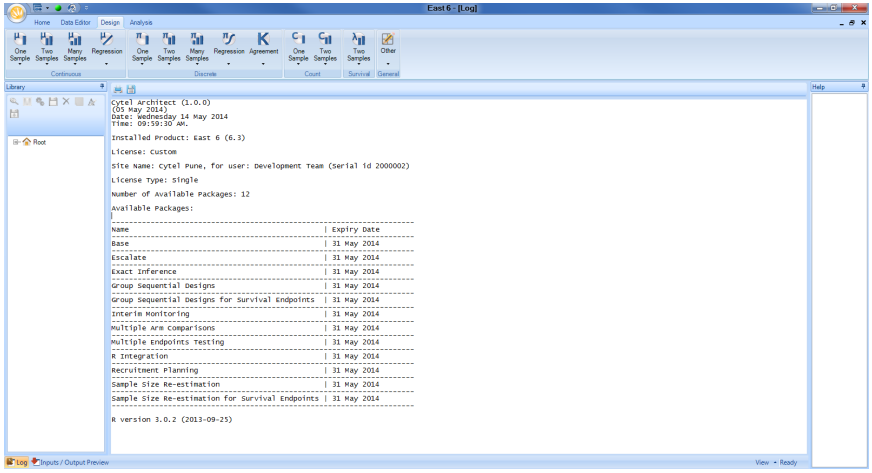
# Volume 2 Continuous Endpoints

# 3 Tutorial: Normal Endpoint

This tutorial introduces you to East on the Architect platform, using an example clinical trial to test difference of means.

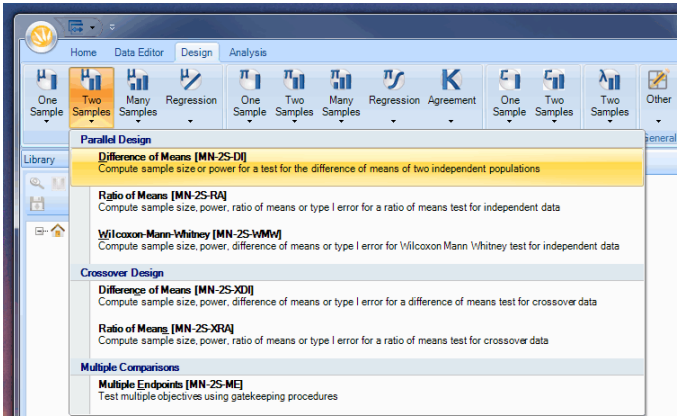
## 3.1 Fixed Sample Design

When you open East, you will see the following screen below.

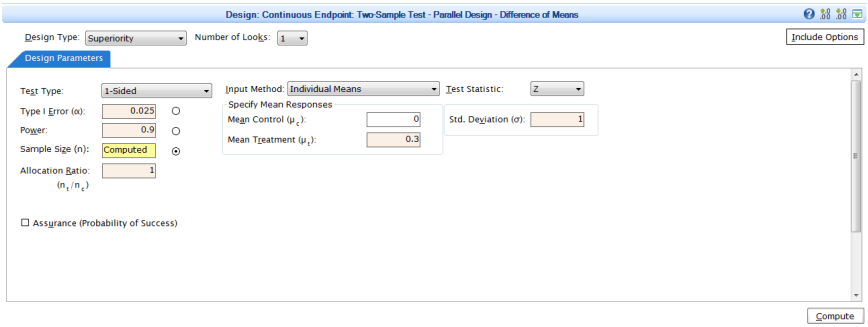



By default, the Design tab in the ribbon will be active. The items on this tab are grouped under the following categories of endpoints: Continuous, Discrete, Count, Survival, and General. Click

**Continuous: Two Samples**, and then **Parallel Design: Difference of Means**.




The following input window will appear.




By default, the radio button for **Sample Size (n)** is selected, indicating that it is the variable to be computed. The default values shown for **Type I Error** and **Power** are $0.025$ and $0.9$. Keep the same for this design. Since the default inputs provide all of the necessary input information, you are ready to compute sample size by clicking the **Compute** button. The

calculated result will appear in the **Output Preview** pane, as shown below.

| ID | Design Type | No. of Looks | Test Type | Specified α | Power | nt/nc | Sample Size | Input Method | δ1 | μc | Mean Treatment (Alt.) | σ | Test Statistic |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Des1 | Superiority | 1 | 1-Sided | 0.025 | 0.9 | 1 | 467 | Individual Means | 0.3 | 0 | 0.3 | 1 | Z |

This single row of output contains relevant details of inputs and the computed result of total sample size (and total completers) of 467. Select this row, and click  to display a summary of the design details in the upper pane (known as **Output Summary**).

| | Des1 |
|---|---|
| Mnemonic | MN-2S-DI |
| **Test Parameters** | |
| Design Type | Superiority |
| No. of Looks | 1 |
| Test Type | 1-Sided |
| Specified α | 0.025 |
| Power | 0.9 |
| **Model Parameters** | |
| Input Method | Individual Means |
| Diff. in Means (δ = μt - μc) | 0.3 |
| Mean Control (μc) | 0 |
| Mean Treatment (μt1) | 0.3 |
| Std. Deviation (σ) | 1 |
| Test Statistic | Z |
| Allocation Ratio (nt/nc) | 1 |
| **Sample Size** | |
| Maximum | 467 |

The discussion so far gives you a quick feel of the software for computing sample size for a single look design. We will describe further features in an example for a group sequential design in the next section.

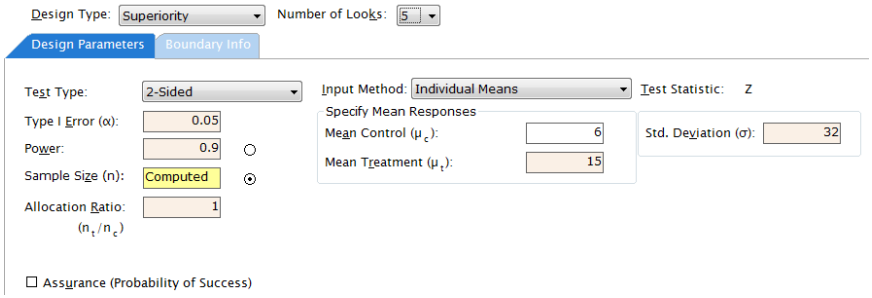
# Chapter 3: *Tutorial: Normal Endpoint*

## 3.2 Group Sequential Design for a Normal Superiority Trial

### *3.2.1 Study Background*

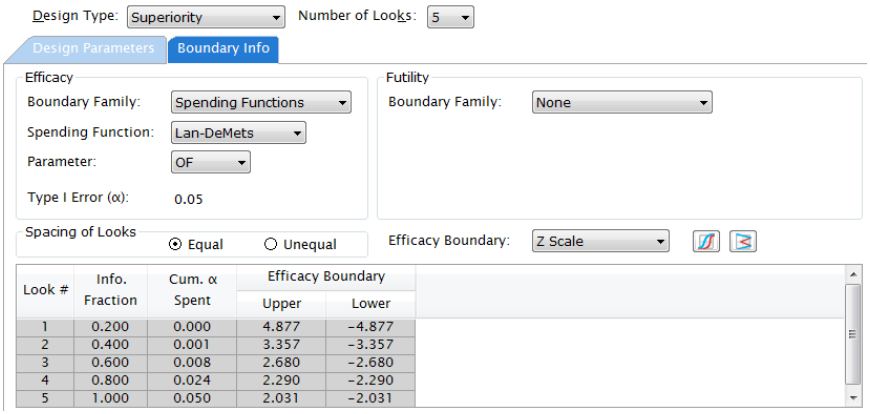
Drug X is a newly developed lipase inhibitor for obesity management that acts by inhibiting the absorption of dietary fats. The performance of this drug needs to be compared with an already marketed drug Y for the same condition. In a randomized, double-blind, trial comparing the efficacy and safety of 1 year of treatment with X to Y (each at 120 mg for three times a day), obese adults are to be randomized to receive either X or Y combined with dietary intervention for a period of one year. The endpoint is weight loss (in pounds). You are to design a trial having 90% power to detect a mean difference of 9 lbs between X and Y, assuming 15 lbs and 6 lbs weight loss in each treatment arm, respectively, and a common standard deviation of 32 lbs. The design is required to be a 2-sided test at the 5% significance level.

From the design menu choose **Continuous: Two Samples**, and then **Parallel Design: Difference of Means**. Select `2-Sided` for **Test Type**, and enter $0.05$ for **Type I Error**. Specify the **Mean Control** be $6$, the **Mean Treatment** to be $15$, and the common **Std. Deviation** to be $32$. Next, change the **Number of Looks** to be $5$. You will see a new tab, **Boundary Info**, added to the input dialog box.




Click the **Boundary Info** tab, and you will see the following screen. On this tab, you can choose whether to specify stopping boundaries for efficacy, or futility, or both. For this trial,

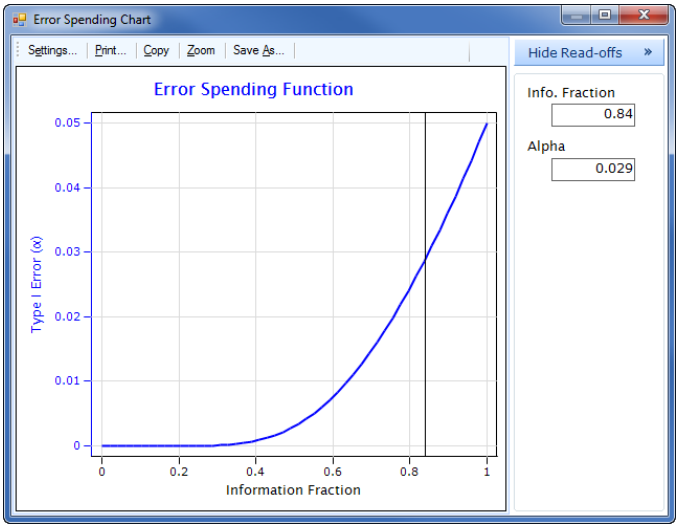
choose efficacy boundaries only, and leave all other default values. We will implement the Lan-Demets (O'Brien-Fleming) spending function, with equally spaced looks.

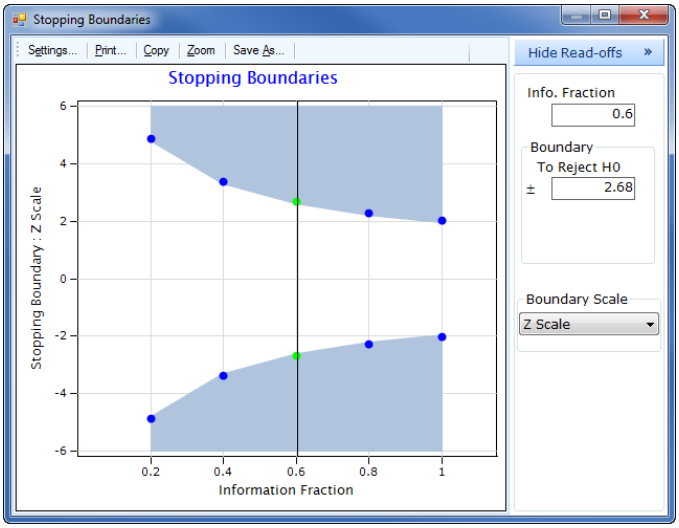



| Look # | Info. Fraction | Cum. α Spent | Efficacy Boundary | |
|---|---|---|---|---|
| | | | Upper | Lower |
| 1 | 0.200 | 0.000 | 4.877 | -4.877 |
| 2 | 0.400 | 0.001 | 3.357 | -3.357 |
| 3 | 0.600 | 0.008 | 2.680 | -2.680 |
| 4 | 0.800 | 0.024 | 2.290 | -2.290 |
| 5 | 1.000 | 0.050 | 2.031 | -2.031 |

On the **Boundary Info** tab, click on the icons  or  , to generate the following

charts.




Click **Compute**. East will show the results in the **Output Preview**.

The maximum combined sample size required under this design is 544. The expected sample sizes under H0 and H1 are 540 and 403, respectively. Click  in the **Output Preview** toolbar to save this design to Wbk1 in the **Library**. Double-click on Des1 to generate the following output.

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

| Test Parameters | |
|---|---|
| Design ID | Des2 |
| Design Type | Superiority |
| Number of Looks | 5 |
| Test Type | 2-Sided |
| Specified α | 0.05 |
| Power | 0.9 |
| **Model Parameters** | |
| Test Statistic | Z |
| Input Method | Individual Means |
| Mean Control ($\mu_c$) | 6 |
| Mean Treatment ($\mu_{tt}$) | 15 |
| $\delta = \mu_t - \mu_c$ | |
| Under H0 | 0 |
| Under H1 | 9 |
| Std. Deviation (σ) | 32 |
| Allocation Ratio ($n_t/n_c$) | 1 |
| **Boundary Parameters** | |
| Spacing of Looks | Equal |
| Efficacy Boundary | LD (OF) |

**Sample Size Information**

| | Control Arm | Treatment Arm | Total |
|---|---|---|---|
| Sample Size (n) | | | |
| Maximum | 272 | 272 | 544 |
| Expected H1 | 201.775 | 201.474 | 403.249 |
| Expected H0 | 270.224 | 270.207 | 540.43 |
| Maximum Information (I):0.133 | | | |

**Stopping Boundaries: Look by Look**

| Look # | Info. Fraction (n/n_max) | Sample Size (n) | Cumulative α Spent | Boundaries | | Boundary Crossing Probability (Incremental) | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | Under H0 | | Under H1 | |
| | | | | Efficacy Z | | Efficacy | | Efficacy | |
| | | | | Upper | Lower | Upper | Lower | Upper | Lower |
| 1 | 0.2 | 109 | 1.104E-6 | 4.872 | -4.872 | 5.519E-7 | 5.519E-7 | 3.32E-4 | 1.146E-10 |
| 2 | 0.401 | 218 | 7.981E-4 | 3.354 | -3.354 | 3.985E-4 | 3.985E-4 | 0.1 | 2.818E-8 |
| 3 | 0.599 | 326 | 0.008 | 2.682 | -2.682 | 0.003 | 0.003 | 0.344 | 8.483E-8 |
| 4 | 0.8 | 435 | 0.024 | 2.29 | -2.29 | 0.008 | 0.008 | 0.301 | 7.455E-8 |
| 5 | 1 | 544 | 0.05 | 2.031 | -2.031 | 0.013 | 0.013 | 0.154 | 4.003E-8 |

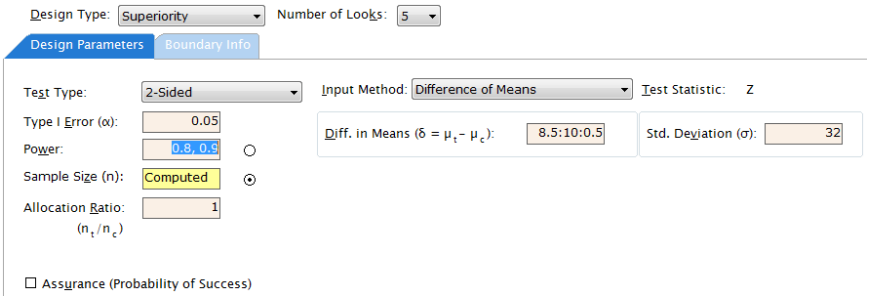
Once you have finished examining the output, close this window, and re-start East before continuing.

## 3.2.2 Creating multiple designs easily

In East, it is easy to create multiple designs by inputting multiple parameter values. In the trial described above, suppose we want to generate designs for all combinations of the following parameter values: **Power** = $0.8, 0.9$, and **Difference in Means** = $8.5, 9, 9.5, 10$. The number of such combinations is $2 \times 4 = 8$.

East can create all 8 designs by a single specification in the input dialog box. Enter the following values as shown below. Remember that the common **Std. Deviation** is 32. From the **Input Method**, select the **`Difference of Means`** option. The values of **Power** have been entered as a list of comma-separated values, while **Difference in Means** has been entered as

a colon-separated range of values: 8.5 to 10 in steps of 0.5.




Now click compute. East computes all 8 designs, and displays them in the **Output Preview** as shown below. Click  to maximize the **Output Preview**.

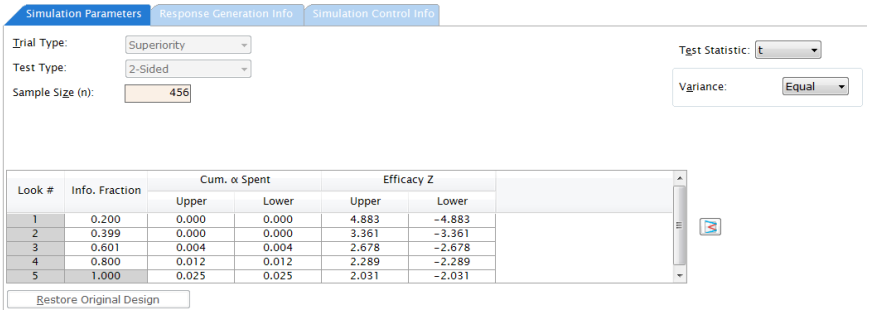
Select the first 3 rows using the Ctrl key, and click  to display a summary of the design details in the upper pane, known as the **Output Summary**.

| | Des1 | Des2 | Des3 |
|---|---|---|---|
| Mnemonic | MN-2S-DI | MN-2S-DI | MN-2S-DI |
| **Test Parameters** | | | |
| Design Type | Superiority | Superiority | Superiority |
| No. of Looks | 5 | 5 | 5 |
| Test Type | 2-Sided | 2-Sided | 2-Sided |
| Specified α | 0.05 | 0.05 | 0.05 |
| Power | 0.8 | 0.8 | 0.801 |
| **Model Parameters** | | | |
| Input Method | Difference of Means | Difference of Means | Difference of Means |
| Diff. in Means (δ = μt - μc) | 8.5 | 9 | 9.5 |
| Std. Deviation (σ) | 32 | 32 | 32 |
| Test Statistic | Z | Z | Z |
| Allocation Ratio (nt/nc) | 1 | 1 | 1 |
| **Boundary Parameters** | | | |
| Efficacy Boundary | LD (OF) | LD (OF) | LD (OF) |
| Spacing of Looks | Equal | Equal | Equal |
| **Sample Size** | | | |
| Maximum | 456 | 407 | 366 |
| Expected Under H0 | 453.004 | 404.326 | 363.595 |
| Expected Under H1 | 366.549 | 327.093 | 294.023 |

Select Des1 in the **Output Preview**, and click  toolbar to save this design in the **Library**. We will use this design for simulation and interim monitoring, as described below. Now that you have saved Des1, delete all designs from the **Output Preview** before continuing, by selecting all designs with the Shift key, and clicking  in the toolbar.

### 3.2.3 Simulation

Right-click Des1 in the **Library**, and select **Simulate**. Alternatively, you can select Des1 and click the  icon.




We will carry out a simulation of Des1 to check whether it preserves the specified power. Click **Simulate**. East will execute by default 10000 simulations with the specified inputs. Close the intermediate window after examining the results. A row labeled as Sim1 will be added in the **Output Preview**.

Click the  icon to save this simulation to the **Library**. A simulation sub-node will be added under Des1 node. Double clicking on the Sim1 node, will display the detailed

simulation output in the work area.

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

| Simulation Parameters | |
|---|---|
| Simulation ID | Sim1 |
| Design Type | Superiority |
| Number of Looks | 5 |
| Test Type | 2-Sided |
| Sample Size (n) | 456 |
| Variance | Equal |
| Test Statistic | t |
| Avg. Power at Termination | 0.802 |
| **Response Generation Parameters** | |
| Generate Data Using | Individual Means |
| Mean Control ($\mu_c$) | 0 |
| Mean Treatment ($\mu_t$) | 8.5 |
| SD Control ($\sigma_c$) | 32 |
| SD Treatment ($\sigma_t$) | 32 |
| **Simulation Control Parameters** | |
| Starting Seed | Clock |
| Number of Simulations | 10000 |

**Average Sample Size**

| Look # | Average Sample Size (n) |
|---|---|
| 1 | 91 |
| 2 | 182 |
| 3 | 274 |
| 4 | 365 |
| 5 | 456 |
| Average | 366.996 |

**Simulation Boundaries and Boundary Crossing Probabilities**

| Look # | Sample Size (n) | Boundaries | | Stopping For | | Total Simulations | |
|---|---|---|---|---|---|---|---|
| | | Efficacy | | | | | |
| | | Upper | Lower | Upper Efficacy | Lower Efficacy | Count | % |
| 1 | 91 | 4.883 | -4.883 | 2 | 0 | 2 | 0.020% |
| 2 | 182 | 3.361 | -3.361 | 609 | 0 | 609 | 6.090% |
| 3 | 274 | 2.678 | -2.678 | 2523 | 0 | 2523 | 25.230% |
| 4 | 365 | 2.289 | -2.289 | 2893 | 0 | 2893 | 28.930% |
| 5 | 456 | 2.031 | -2.031 | 1996 | 0 | 3973 | 39.730% |
| Total | | | | 8023 | 0 | 10000 | |
| % | | | | 80.230% | 0.000% | | |

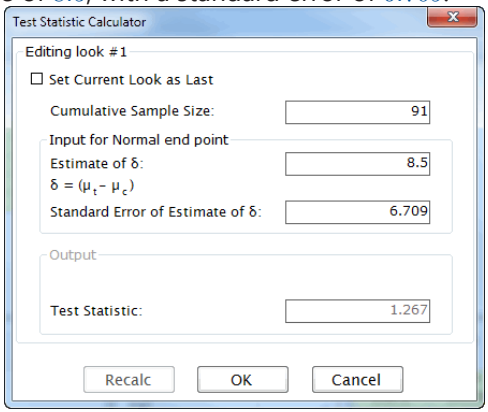
**Simulation Seed and Elapsed Time**

Starting Seed: 725323
Total Number of Simulations: 10000
Elapsed Time: 00:01:05

In 80.23% of the simulated trials, the null hypothesis was rejected. This value is very close to the specified power of 80%. The next section will explore interim monitoring with this design.

### 3.2.4 Interim Monitoring

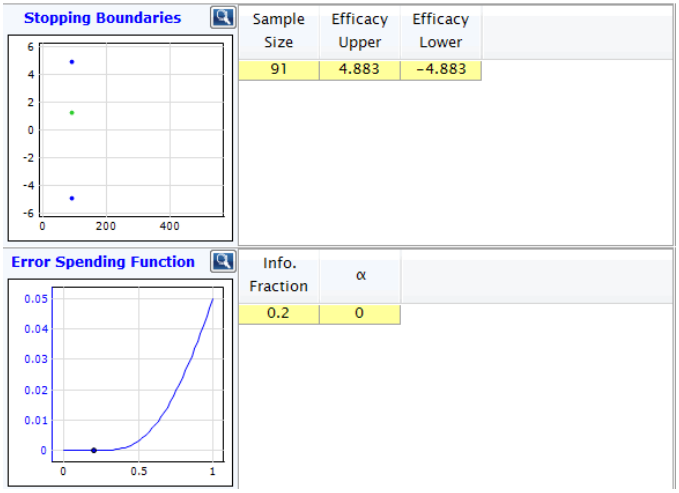
Right-click Des1 in the **Library** and select **Interim Monitoring**. Click the Enter Interim Data to open the **Test Statistic Calculator**. Suppose that after 91 subjects, at the first look, you have observed a mean difference of 8.5, with a standard error of 6.709.



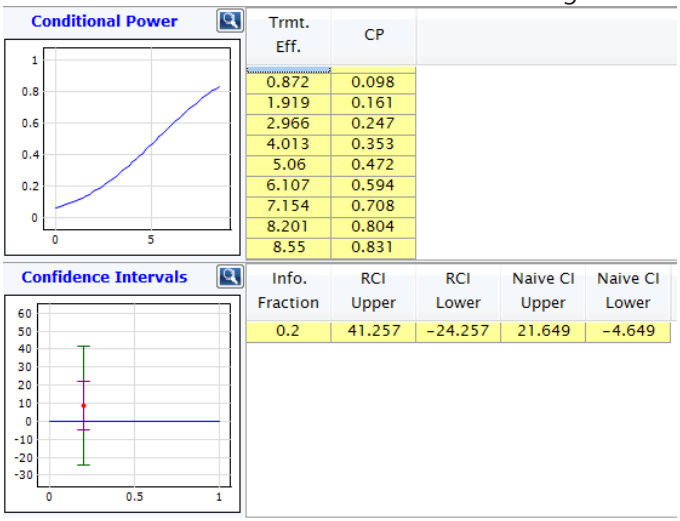

Click **OK** to update the IM Dashboard.

| Look # | Information Fraction | Cumulative Sample Size | Test Statistic | δ | Standard Error | Efficacy | | 95% RCI for δ | | Repeated p-value | CP | Predictive Power |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | Upper | Lower | Upper | Lower | | | |
| 1 | 0.2 | 91 | 1.267 | 8.5 | 6.709 | 4.883 | -4.883 | 41.257 | -24.257 | 0.932 | 0.828 | 0.673 |
| 2 | | | | | | | | | | | | |
| 3 | | | | | | | | | | | | |
| 4 | | | | | | | | | | | | |
| 5 | | | | | | | | | | | | |

The **Stopping Boundaries** and **Error Spending Function** charts on the left:




The **Conditional Power** and **Confidence Intervals** charts on the right:




Suppose that after 182 subjects, at the second look, you have observed a mean difference of

16, with a standard error of 4.744. Click **Recalc**, and then **OK** to update the IM Dashboard. In this case, a boundary has been crossed, and the following window appears.

Click **Stop** to complete the trial. The IM Dashboard will be updated accordingly, and a table for **Final Inference** will be displayed as shown below.

Final Inference

| | |
|---|---|
| Final Outputs at Look # | 2 |
| Adj. p-value | 0.001 |
| Adj. Pt. Est. for δ | 16 |
| **Adj. 95% CI for δ** | |
| Upper Confidence Bound | 25.298 |
| Lower Confidence Bound | 6.702 |
| Post-Hoc Power | |

# 4 Normal Superiority One-Sample

To compare a new process or treatment to a well-established control, a single-sample study may suffice for preliminary information prior to a full-scale investigation. This single sample may either consist of a random sample of observations from a single treatment when the mean is to be compared to a specified constant or a random sample of paired differences or ratio between two treatments. The former is presented in Section (4.1) and the latter is discussed in Section (4.2) and Section (4.3).

## 4.1 Single Mean

- 4.1.1 Trial Design
- 4.1.2 Simulation
- 4.1.3 Interim Monitoring
- 4.1.4 Trial Design Using a t-Test (Single Look)

The problem of comparing the mean of the distribution of observations from a single random sample to a specified constant is considered. For example, when developing a new drug for treatment of a disease, there should be evidence of efficacy. For this single-sample problem, it is desired to compare the unknown mean $\mu$ to a fixed value $\mu_0$. The null hypothesis $H_0$: $\mu = \mu_0$ is tested against the two-sided alternative hypothesis $H_1$: $\mu \neq \mu_0$ or a one-sided alternative hypothesis $H_1$: $\mu < \mu_0$ or $H_1$: $\mu > \mu_0$. The power of the test is computed at a specified value of $\mu = \mu_1$ and standard deviation $\sigma$.

Let $\hat{\mu}_j$ denote the estimate of $\mu$ based on $n_j$ observations, up to and including the $j$-th look, $j = 1, ..., K$, with a maximum of $K$ looks. The test statistic at the $j$-th look is based on the value specified by the null hypothesis, namely

$$Z_j = n_j^{1/2}(\hat{\mu}_j - \mu_0)/\hat{\sigma}_j, \tag{4.1}$$

where $\hat{\sigma}_j^2$ is the sample variance based on $n_j$ observations.

### 4.1.1 Trial Design

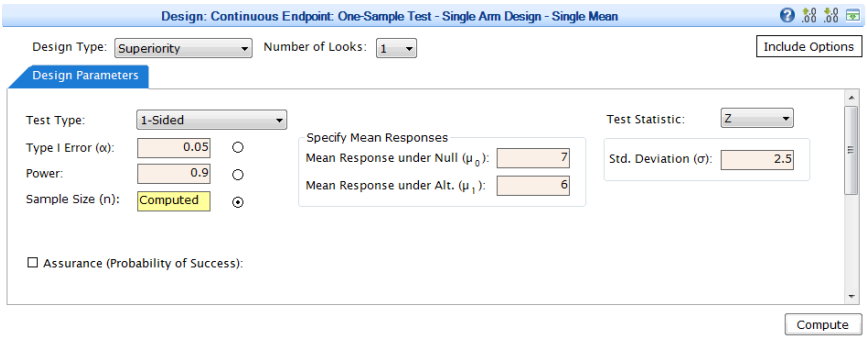
Consider the situation where treatment for a certain infectious disorder is expected to result in a decrease in the length of hospital stay. Suppose that hospital records were reviewed and it was determined that, based on this historical data, the average hospital stay is approximately 7 days. It is hoped that the new treatment can decrease this to less than 6 days. It is assumed that the standard deviation is $\sigma = 2.5$ days.The null hypothesis $H_0$: $\mu = 7 (= \mu_0)$ is tested against the alternative hypothesis $H_1$: $\mu < 7$.

First, click **Continuous: One Sample** on the **Design** tab and then click **Single Arm Design: Single Mean** as shown below.

This will launch a new input window.

**Single-Look Design**

We want to determine the sample size required to have power of 90% when $\mu = 6 (= \mu_1)$, using a test with a one-sided type-1 error rate of 0.05. Choose **Test Type** as `1-Sided`. Specify **Mean Response under Null ($\mu_0$)** as $7$, **Mean Response under Alt. ($\mu_1$)** as $6$ and **Std. Deviation** $(\sigma)$ as $2.5$. The upper pane should appear as below:




Click **Compute**. This will calculate the sample size for this design and the output is shown as a

row in the **Output Preview**. The computed sample size is 54 subjects.

Output Preview

| | ID | Design Type | No. of Looks | Test Type | Specified α | Power | Sample Size | σ | μ0 | μ1 | Test Statistic |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Des1 | Superiority | 1 | 1-Sided | 0.05 | 0.902 | 54 | 2.5 | 7 | 6 | Z |

This design has default name Des 1. Select this design by clicking anywhere along the row and click in the **Output Preview** toolbar. Some of the design details will be displayed in the upper pane, labeled as **Output Summary**.

Output Summary

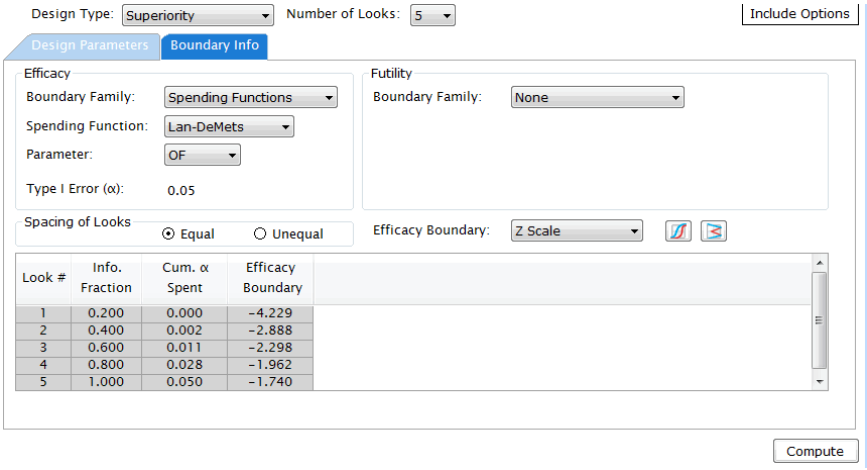
| | Des1 |
|---|---|
| Mnemonic | MN-1S-SM |
| **Test Parameters** | |
| Design Type | Superiority |
| No. of Looks | 1 |
| Test Type | 1-Sided |
| Specified α | 0.05 |
| Power | 0.902 |
| **Model Parameters** | |
| Std. Deviation (σ) | 2.5 |
| Mean Response under Null (μ0) | 7 |
| Mean Response under Alt. (μ1) | 6 |
| Test Statistic | Z |
| **Sample Size** | |
| Maximum | 54 |

In the **Output Preview** toolbar select Des 1, click to save this design to Wbk1 in the **Library**.

***Five-Look Design***

To allow the opportunity to stop early and proceed with a full-scale plan, five equally-spaced analyses are planned, using the Lan-DeMets (O'Brien-Fleming) stopping boundary. Create a new design by right-clicking Des 1 in the **Library**, and selecting **Edit Design**. In the Input, change the **Number of Looks** from 1 to 5, to generate a study with four interim looks and a final analysis. A new tab for **Boundary Info** should appear. Click this tab to reveal the

stopping boundary parameters. By default, the **Spacing of Looks** is set to **Equal**, which means that the interim analyses will be equally spaced in terms of the number of patients accrued between looks. The left side contains details for the **Efficacy** boundary, and the right side contains details for the **Futility** boundary. By default, there is an efficacy boundary (to reject $H_0$) selected, but no futility boundary (to reject $H_1$). The **Boundary Family** specified is of the **Spending Functions** type. The default **Spending Function** is the **Lan-DeMets** (Lan & DeMets, 1983), with **Parameter** as **OF** (O'Brien-Fleming), which generates boundaries that are very similar, though not identical, to the classical stopping boundaries of O'Brien and Fleming (1979). For a detailed description of the different spending functions and stopping boundaries available in East refer to Chapter **??**. The cumulative alpha spent and the boundary values are displayed below.




Click **Compute**. The maximum and expected sample sizes are highlighted in yellow in the **Output Preview**. Save this design in the current workbook by selecting the corresponding row in the **Output Preview** and clicking on the **Output Preview** toolbar. To compare Des 1 and Des 2, select both rows in **Output Preview** using the Ctrl key and click in the

**Output Preview** toolbar. This will display both designs in the **Output Summary** pane.

| | Des1 | Des2 |
|---|---|---|
| Mnemonic | MN-1S-SM | MN-1S-SM |
| **Test Parameters** | | |
| Design Type | Superiority | Superiority |
| No. of Looks | 1 | 5 |
| Test Type | 1-Sided | 1-Sided |
| Specified α | 0.05 | 0.05 |
| Power | 0.902 | 0.903 |
| **Model Parameters** | | |
| Std. Deviation (σ) | 2.5 | 2.5 |
| Mean Response under Null (μ0) | 7 | 7 |
| Mean Response under Alt. (μ1) | 6 | 6 |
| Test Statistic | Z | Z |
| **Boundary Parameters** | | |
| Spacing of Looks | | Equal |
| Efficacy Boundary | | LD (OF) |
| **Sample Size** | | |
| Maximum | 54 | 56 |
| Expected Under H0 | | 55.531 |
| Expected Under H1 | | 39.897 |

Des 2 results in a maximum of 56 subjects in order to attain 90% power, with an expected sample size of 40 under the alternative hypothesis. In order to see the stopping probabilities, double-click Des 2 in the **Library**.

**Stopping Boundaries: Look by Look**

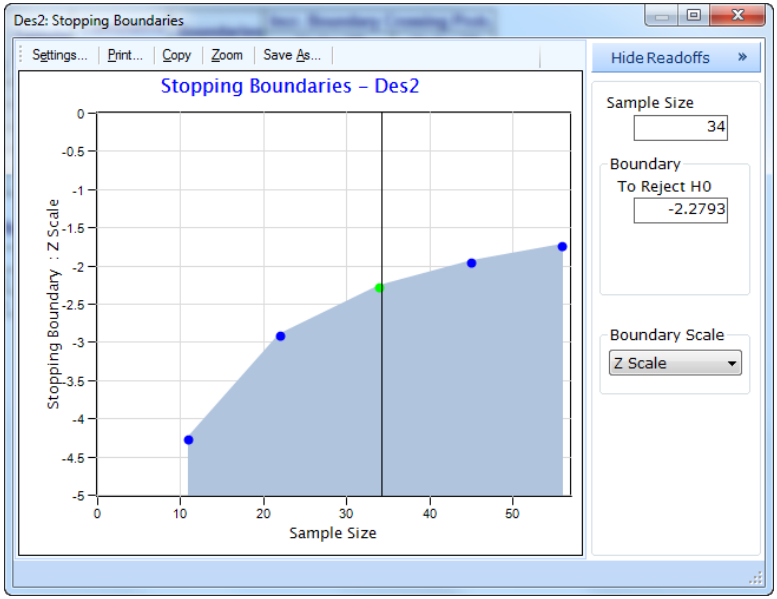
| Look # | Info. Fraction (n/n_max) | Sample Size (n) | Cumulative α Spent | Boundaries | Boundary Crossing Probability (Incremental) | |
|---|---|---|---|---|---|---|
| | | | | | Under H0 | Under H1 |
| | | | | Efficacy Z | Efficacy | Efficacy |
| 1 | 0.196 | 11 | 9.767E-6 | -4.27 | 9.767E-6 | 0.002 |
| 2 | 0.393 | 22 | 0.002 | -2.918 | 0.002 | 0.147 |
| 3 | 0.607 | 34 | 0.012 | -2.279 | 0.01 | 0.376 |
| 4 | 0.804 | 45 | 0.029 | -1.958 | 0.017 | 0.25 |
| 5 | 1 | 56 | 0.05 | -1.741 | 0.021 | 0.128 |

The clear advantage of this sequential design resides in the relatively high cumulative probability of stopping by the third look if the alternative is true, with a sample size of 34 patients, which is well below the requirements for a fixed sample study (54 patients). Close the Output window before continuing.

***Examining stopping boundaries and spending functions***

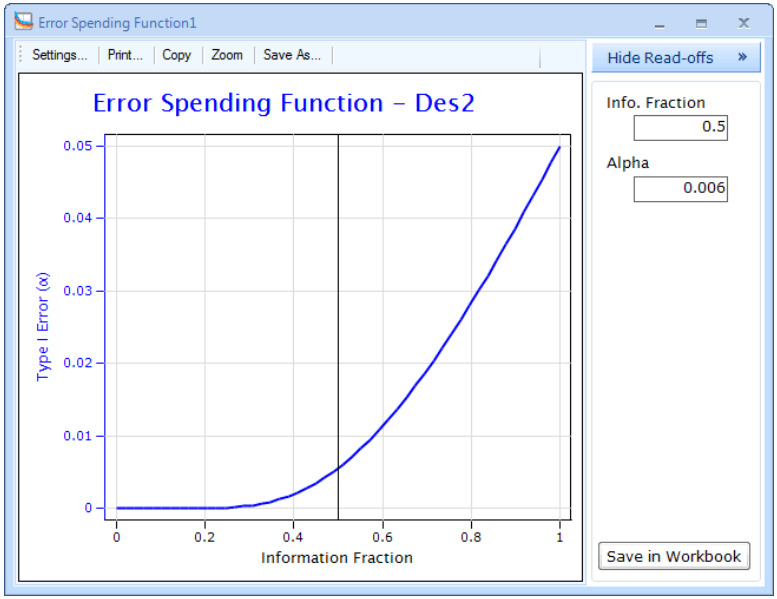
You can plot the boundary values of Des 2 by clicking on the **Library** toolbar, and then

clicking **Stopping Boundaries**. The following chart will appear:




You can choose different boundary scales from the drop down box located in the right hand side. The available boundary scales are Z scale, Score Scale, $\mu/\sigma$ Scale and $p$-value scale. To plot the error spending function for Des 2, select Des 2 in the **Library**, click in the

toolbar, and then click **Error Spending**. The following chart will appear:




The above spending function is according to Lan and DeMets (1983) with O'Brien-Fleming flavor and for one-sided tests has the following functional form:

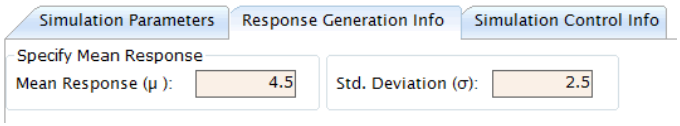
$$\alpha(t) = 2 - 2\Phi\left(\frac{Z_{\alpha/2}}{\sqrt{t}}\right)$$

Observe that very little of the total type-1 error is spent early on, but more is spent rapidly as the information fraction increases, and reaches 0.05 at an information fraction of 1. Feel free to try other plots by clicking in the Library toolbar. Close all charts before continuing.

### 4.1.2 Simulation

Suppose we want to see the advantages of performing the interim analyses, as it relates to the chance of stopping prior to the final analysis. This examination can be conducted using

simulation. Select Des 2 in the **Library**, and click  in the toolbar. Alternatively, right-click on Des 2 and select **Simulate**. A new Simulation window will appear. For example, suppose you wish to determine how quickly this trial could be terminated if the treatment difference was much greater than expected. For example, under the alternative hypothesis, $\mu = 4.5$. Click on the **Response Generation Info** tab, and specify: **Mean Response($\mu$)** = $4.5$ and **Std. Deviation ($\sigma$)** = $2.5$.




Click **Simulate** to start the simulation. Once the simulation run has completed, East will add an additional row to the **Output Preview** labeled as Sim 1.

Select Sim 1 in the **Output Preview** and click  . Now double-click on Sim 1 in the **Library**. The simulation output details will be displayed in the upper pane.

**Simulation Boundaries and Boundary Crossing Probabilities**

| Look # | Sample Size (n) | Boundaries | Stopping For | Total Simulations | |
|---|---|---|---|---|---|
| | | Efficacy | | | |
| | | Lower | Efficacy | Count | % |
| 1 | 11 | -4.27 | 2606 | 2606 | 26.060% |
| 2 | 22 | -2.918 | 6886 | 6886 | 68.860% |
| 3 | 34 | -2.279 | 504 | 504 | 5.040% |
| 4 | 45 | -1.958 | 4 | 4 | 0.040% |
| 5 | 56 | -1.741 | 0 | 0 | 0.000% |
| Total | | | 10000 | 10000 | |
| % | | | 100.000% | | |

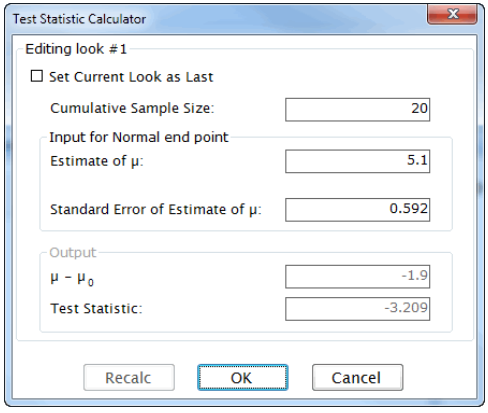
Observe that 100% simulated trials rejected the null hypothesis, and about 26% of these simulations were able to reject the null at the first look after enrolling only 11 subjects. Your numbers might differ slightly due to a different starting seed.

### 4.1.3 Interim Monitoring

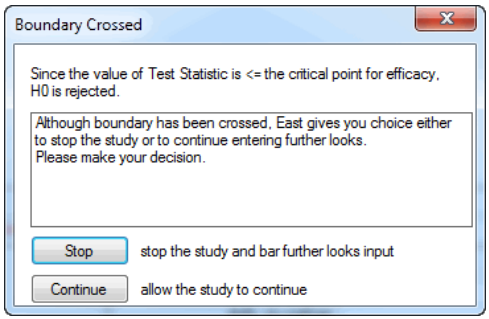
Suppose that the trial has commenced and Des 2 was implemented. Right-click Des 2 in the **Library**, and select **Interim Monitoring**.

Although we specified that there will be five equally spaced interim looks, the Lan-DeMets methodology implemented in East allows you to alter the number and spacing of these looks. Accordingly, suppose that an interim look was taken after enrolling 20 subjects and the sample mean, based on these 20 subjects, was 5.1 with a standard error of 0.592. Since $\mu_0 = 7$, based on equation (4.1) the value of the test statistic at the first look would be $Z_1 = (5.1 - 7)/0.592$ or -3.209.

Click **Enter Interim Data** on the toolbar. In the **Test Statistic Calculator**, enter the following values, and click **Recalc** and then**OK**.




Since the stopping boundary is crossed, the following dialog box appears.




Click **Stop** to take you back to the interim monitoring dashboard. For final inference, East will display the following summary information on the dashboard.

Final Inference

| | |
|---|---|
| Final Outputs at Look # | 1 |
| Adj. p-value | 0.001 |
| Adj. Pt. Est. for μ | 5.1 |
| **Adj. 90% CI for μ** | |
| Upper Confidence Bound | 6.074 |
| Lower Confidence Bound | 4.126 |
| Post-Hoc Power | |

### 4.1.4 Trial Design Using a t-Test (Single Look)

The sample size obtained to correctly power Des 1 in Section (4.1.1) relied on using a Wald-type statistic for the hypothesis test, given by equation (4.1). Due to the assumption of normal distribution for the test statistic, we have ignored the fact that the variance $\sigma$ is estimated from the sample. For large sample sizes this approximation is acceptable. However, in small samples with unknown standard deviation the test statistic

$$Z = n^{1/2}(\hat{\mu} - \mu_0)/\hat{\sigma}, \tag{4.2}$$

is distributed with student's t distribution with $(n-1)$ degrees of freedom. Here, $\hat{\sigma}^2$ denotes the sample variance based on $n$ observations.

Consider the example in Section 4.1.1 where we would like to test the null hypothesis that the

average hospital stay is 7 days, $H_0$: $\mu = 7 (= \mu_0)$, against the alternative hypothesis that is less than 7 days, $H_1$: $\mu < 7$. We will now design the same trial in a different manner, using the t distribution for the test statistic.

Right-click Des 1 in the **Library**, and select **Edit Design**. In the input window, change the **Test Stat.** from **Z** to **t**. The entries for the other fields need not be changed.

Click **Compute**. East will add an additional row to the **Output Preview** labeled as Des 3. The required sample size is 55. Select the rows corresponding to Des 1 and Des 3 and click  . This will display Des 1 and Des 3 in the **Output Summary**.

| | Des1 | Des3 |
|---|---|---|
| Mnemonic | MN-1S-SM | MN-1S-SM |
| **Test Parameters** | | |
| Design Type | Superiority | Superiority |
| No. of Looks | 1 | 1 |
| Test Type | 1-Sided | 1-Sided |
| Specified α | 0.05 | 0.05 |
| Power | 0.902 | 0.9 |
| **Model Parameters** | | |
| Std. Deviation (σ) | 2.5 | 2.5 |
| Mean Response under Null (μ0) | 7 | 7 |
| Mean Response under Alt. (μ1) | 6 | 6 |
| Test Statistic | Z | t |
| **Sample Size** | | |
| Maximum | 54 | 55 |

Des 3, which uses the t distribution, requires that we commit a combined total of 55 patients to the study, just one more compared to Des 1, which uses the normal distribution. The extra patient is needed to compensate for the extra variability due to estimation of the var$[\hat{\delta}]$.

## 4.2 Mean of Paired Differences

- 4.2.1 Trial Design
- 4.2.2 Simulation
- 4.2.3 Interim Monitoring
- 4.2.4 Trial Design Using a t-Test (Single Look)

The paired t-test is used to compare the means of two normal distributions when each observation in the random sample from one distribution is matched with a unique observation from the other distribution. Let $\mu_c$ and $\mu_t$ denote the two means to be compared and let $\sigma^2$ denote the variance of the differences.

The null hypothesis $H_0$: $\mu_c = \mu_t$ is tested against the two-sided alternative hypothesis $H_1$: $\mu_c \neq \mu_t$ or a one-sided alternative hypothesis $H_1$: $\mu_c < \mu_t$ or $H_1$: $\mu_c > \mu_t$. Let $\delta = \mu_t - \mu_c$.

The null hypothesis can be expressed as $H_0$: $\delta = 0$ and the alternative can be expressed as $H_1$: $\delta \neq 0$, $H_1$: $\delta > 0$, or $H_1$: $\delta < 0$. The power of the test is computed at specified values of $\mu_c, \mu_t,$ and $\sigma$.

Let $\hat{\mu}_{cj}$ and $\hat{\mu}_{tj}$ denote the estimates of $\mu_c$ and $\mu_t$ based on $n_j$ observations, up to and including $j$-th look, $j = 1, \ldots, K$ where a maximum of $K$ looks are to be made. The estimate of the difference at the $j$-th look is

$$\hat{\delta}_j = \hat{\mu}_{tj} - \hat{\mu}_{cj}$$

and the test statistic at the $j$-th look is

$$Z_j = n_j^{1/2} \hat{\delta}_j / \hat{\sigma}_j, \tag{4.3}$$

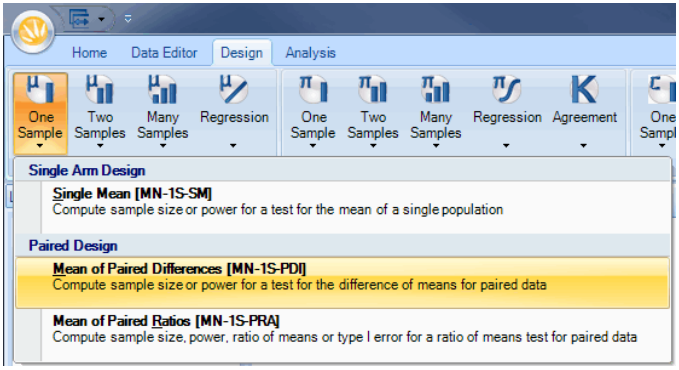
where $\hat{\sigma}_j^2$ is the sample variance of $n_j$ paired differences.

### 4.2.1 Trial Design

Consider the situation where subjects are treated once with placebo after pain is experimentally induced, and later treated with a new analgesic after pain is induced a second time. Pain is reported by the subjects using a 10 cm visual analog scale (0="no pain", . . . , 10="extreme pain"). After treatment with placebo, the average is expected to be 6 cm. After treatment with the analgesic, the average is expected to be 4 cm. It is assumed that the common standard deviation is $\sigma$ = 5 cm. The null hypothesis $H_0$: $\delta = 0$ is tested against the alternative hypothesis $H_1$: $\delta < 0$.

Start East afresh. First, **Continuous: One Sample** on the **Design** tab, and then click **Paired**

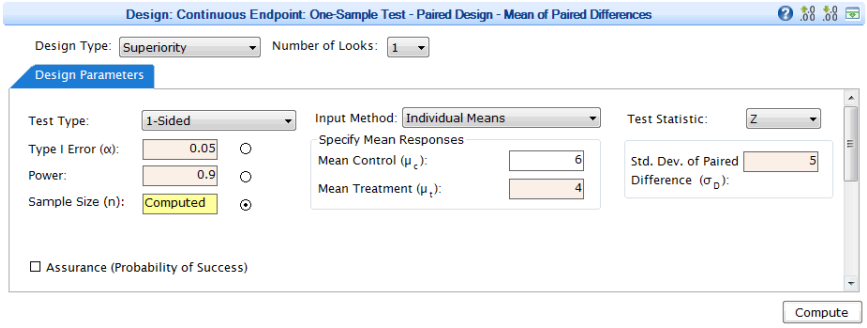
## Design: Mean of Paired Differences




This will launch a new input window.

### *Single-Look Design*

We want to determine the sample size required to have power of 90% when $\mu_c = 6$ and $\mu_t = 4$, using a test with a one-sided type-1 error rate of 0.05. Select **Test Type** as `1-Sided`, `Individual Means` for **Input Method**, and specify the **Mean Control ($\mu_c$)** as $6$ and **Mean Treatment ($\mu_t$)** as $4$. Enter **Std. Dev. of Paired Difference ($\sigma_0$)** as $5$. The upper pane should appear as below:




Click **Compute**. This will calculate the sample size for this design and the output is shown as a

row in the **Output Preview**. The computed sample size is 54 subjects.

Output Preview

| ID | Design Type | No. of Looks | Test Type | Specified α | Power | Sample Size | Test Statistic | Input Method | μc | Mean Treatment (Alt.) | δ1 | σD |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Des1 | Superiority | 1 | 1-Sided | 0.05 | 0.902 | 54 | Z | Individual Means | 6 | 4 | -2 | 5 |

This design has default name Des 1. Select this design by clicking anywhere along the row in the **Output Preview** and click . Some of the design details will be displayed in the upper pane, labeled as **Output Summary**.

Output Summary

| | Des1 |
|---|---|
| Mnemonic | MN-1S-PDI |
| **Test Parameters** | |
| Design Type | Superiority |
| No. of Looks | 1 |
| Test Type | 1-Sided |
| Specified α | 0.05 |
| Power | 0.902 |
| **Model Parameters** | |
| Test Statistic | Z |
| Input Method | Individual Means |
| Mean Control (μc) | 6 |
| Mean Treatment (μt) | 4 |
| Diff. of Means (μt - μc) | -2 |
| Std. Deviation (σD) | 5 |
| **Sample Size** | |
| Maximum | 54 |

In the **Output Preview** toolbar select Des 1, click to save this design to Wbk1 in the **Library**.

***Three-Look Design***

For the above study, suppose we wish to take up to two equally spaced interim looks and one final look as we accrue data, using the Lan-DeMets (O'Brien-Fleming) stopping boundary. Create a new design by right-clicking Des 1 in the **Library**, and **Edit Design**. In the Input, change the **Number of Looks** from 1 to 3, to generate a study with two interim looks and a final analysis.

Click **Compute**. The maximum and expected sample sizes are highlighted in yellow in the **Output Preview**. Save this design in the current workbook by selecting the corresponding

row in **Output Preview** and clicking on the **Output Preview** toolbar. To compare Des 1 and Des 2, select both rows in **Output Preview** using the Ctrl key and click . Both designs will be displayed in the **Output Summary** pane.

| | Des1 | Des2 |
|---|---|---|
| Mnemonic | MN-1S-PDI | MN-1S-PDI |
| **Test Parameters** | | |
| Design Type | Superiority | Superiority |
| No. of Looks | 1 | 3 |
| Test Type | 1-Sided | 1-Sided |
| Specified α | 0.05 | 0.05 |
| Power | 0.902 | 0.902 |
| **Model Parameters** | | |
| Test Statistic | Z | Z |
| Input Method | Individual Means | Individual Means |
| Mean Control (μc) | 6 | 6 |
| Mean Treatment (μt) | 4 | 4 |
| Diff. of Means (μt - μc) | -2 | -2 |
| Std. Deviation (σD) | 5 | 5 |
| **Boundary Parameters** | | |
| Efficacy Boundary | | LD (OF) |
| Spacing of Looks | | Equal |
| **Sample Size** | | |
| Maximum | 54 | 55 |
| Expected Under H0 | | 54.685 |
| Expected Under H1 | | 42.646 |

Des 2 results in a maximum of 55 subjects in order to attain 90% power, with an expected sample size of 43 under the alternative hypothesis. In the **Output Preview** toolbar select Des 2, click to save this design to Wbk1 in the **Library**. In order to see the stopping probabilities, double-click Des 2 in the **Library**.

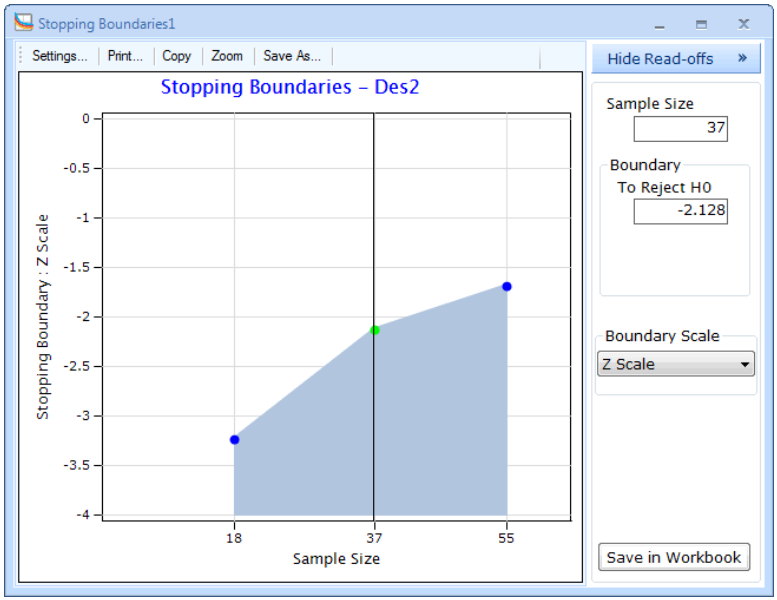
⊖ **Stopping Boundaries: Look by Look**

| Look # | Info. Fraction (n/n_max) | Sample Size (n) | Cumulative α Spent | Boundaries | Boundary Crossing Probability (Incremental) | |
|---|---|---|---|---|---|---|
| | | | | | Under H0 | Under H1 |
| | | | | Efficacy Z | Efficacy | Efficacy |
| 1 | 0.327 | 18 | 6.124E-4 | -3.233 | 6.124E-4 | 0.062 |
| 2 | 0.673 | 37 | 0.017 | -2.128 | 0.016 | 0.558 |
| 3 | 1 | 55 | 0.05 | -1.696 | 0.033 | 0.282 |

The clear advantage of this sequential design resides in the high cumulative probability of stopping by the third look if the alternative is true, with a sample size of 37 patients, which is well below the requirements for a fixed sample study (54 patients). Close the Output window

before continuing.

Select Des 2 and click on the Library toolbar. You can select one of many plots, including one for **Stopping Boundaries**:




Close this chart before continuing.

### *4.2.2 Simulation*

Select Des 2 in the **Library**, and click in the toolbar. Click on the **Response Generation Info** tab, and make sure **Mean Treatment($\mu_t$)** = $4$, **Mean Control($\mu_c$)** = $6$ and **Std. Deviation ($\sigma$)** = $5$. Click **Simulate**. Once the simulation run has completed, East will add an additional row to the **Output Preview** labeled as Sim 1.

Select Sim 1 in the **Output Preview** and click . Now double-click on Sim 1 in the

**Library**. The simulation output details will be displayed.

**Simulation: Continuous Endpoint: One-Sample Test - Paired Design - Mean of Paired Differences**

| Simulation Parameters | |
|---|---|
| Simulation ID | Sim1 |
| Design Type | Superiority |
| Number of Looks | 3 |
| Test Type | 1-Sided |
| Sample Size (n) | 55 |
| Test Statistic | t |
| Avg. Power at Termination | 0.895 |
| **Response Generation Parameters** | |
| Mean Control ($\mu_c$) | 6 |
| Mean Treatment ($\mu_t$) | 4 |
| Std. Deviation ($\sigma_D$) | 5 |
| **Simulation Control Parameters** | |
| Starting Seed | Clock |
| Number of Simulations | 10000 |

⊖**Average Sample Size**

| Look # | Average Sample Size (n) |
|---|---|
| 1 | 18 |
| 2 | 37 |
| 3 | 55 |
| Average | 41.981 |

⊖**Simulation Boundaries and Boundary Crossing Probabilities**

| Look # | Sample Size (n) | Boundaries | Stopping For | Total Simulations | |
|---|---|---|---|---|---|
| | | Efficacy | | | |
| | | Lower | Efficacy | Count | % |
| 1 | 18 | -3.233 | 950 | 950 | 9.500% |
| 2 | 37 | -2.128 | 5280 | 5280 | 52.800% |
| 3 | 55 | -1.696 | 2719 | 3770 | 37.700% |
| Total | | | 8949 | 10000 | |
| % | | | 89.490% | | |

Overall, close to 90% of simulations have rejected $H_0$. The numbers on your screen might differ slightly due to a different seed.

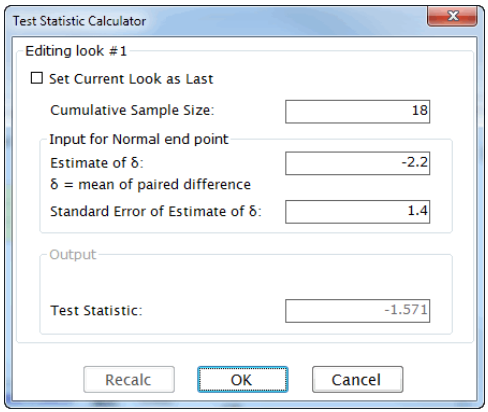
### 4.2.3 Interim Monitoring

For an ongoing study we evaluate the test statistic at an interim stage to see whether we have enough evidence to reject $H_0$. Right-click Des 2 in the Library, and select **Interim Monitoring**.

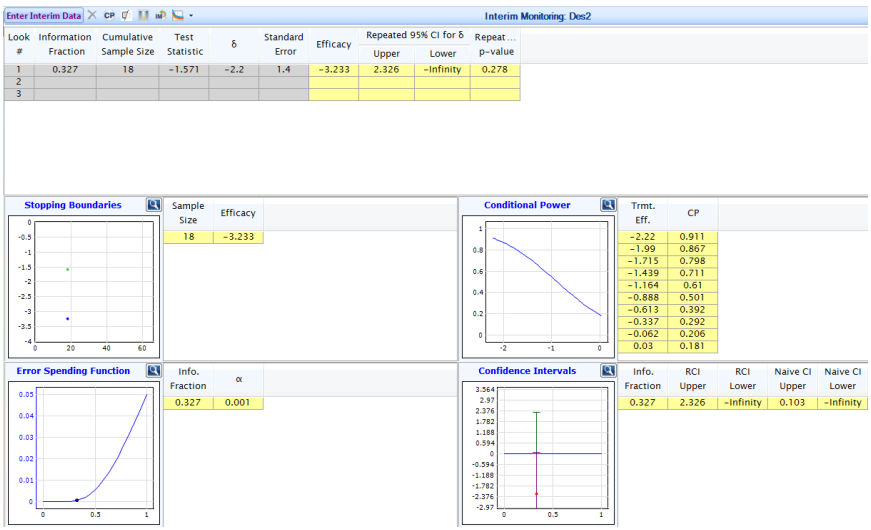
Although the design specified that there be three equally spaced interim looks, the Lan-DeMets methodology implemented in East allows you to alter the number and spacing of these looks. Suppose that an interim look was taken after enrolling 18 subjects and the sample mean, based on these subjects, was -2.2 with a standard error of 1.4. Then based on equation (4.3), the value of the test statistic at first look would be $Z_1 = (-2.2)/1.4$ or -1.571.

Click **Enter Interim Data** on the toolbar. In the **Test Statistic Calculator**, enter the following

values, and click **Recalc** and then**OK**.

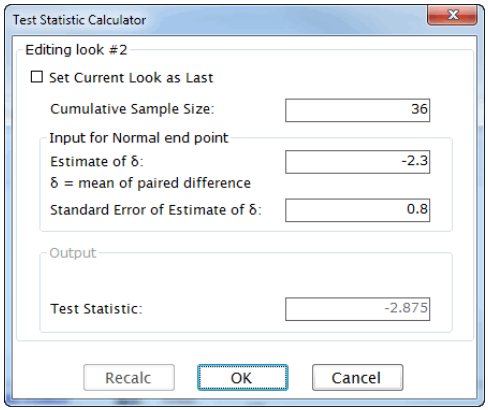



The dashboard will be updated accordingly.

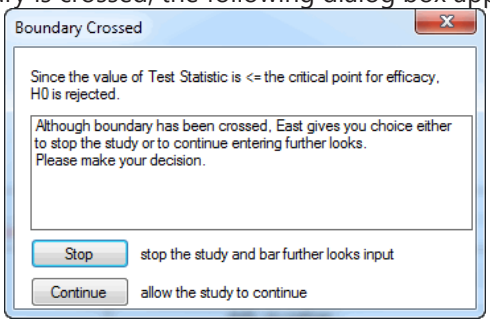



As the observed value -1.571 has not crossed the critical boundary value of -3.233, the trial continues. Now, 18 additional subjects are enrolled, and a second interim analysis with 36

subjects is conducted. Suppose that the observed difference is -2.3 with standard error as 0.8. Select the Look 2 row and click **Enter Interim Data**. Enter these values, and click **Recalc**, and then**OK**.




Since the stopping boundary is crossed, the following dialog box appears. Click on **Stop**.




For final inference, East will display the following summary information on the dashboard.

| Final Inference | |
|---|---|
| Final Outputs at Look # | 2 |
| Adj. p-value | 0.002 |
| Adj. Pt. Est. for δ | -2.287 |
| **Adj. 90% CI for δ** | |
| Upper Confidence Bound | -0.959 |
| Lower Confidence Bound | -3.607 |
| Post-Hoc Power | |

### 4.2.4 Trial Design Using a t-Test (Single Look)

The sample size obtained to correctly power the trial in Section (4.2.1) relied on using a Wald-type statistic for the hypothesis test, given by equation (4.3). However, we neglected the fact that the variance $\sigma$ is estimated by assuming that the test statistic follows a standard normal distribution. For large sample sizes, asymptotic theory supports this approximation. In a single-look design, this test statistic is calculated as

$$Z = n^{1/2}\hat{\delta}/\hat{\sigma}, \tag{4.4}$$

where $\hat{\sigma}^2$ is the sample variance based on $n$ observed paired differences. In the following calculations we take into consideration that $Z$ follows a Student's t-distribution with $(n-1)$ degrees of freedom.

Consider the example in Section 4.2.1 where we would like to test the null hypothesis that the analgesic does not reduce pain, $H_0$: $\delta = 0$, against the alternative hypothesis that the new analgesic works to reduce pain, $H_1$: $\delta < 0$. We will design this same trial using the t distribution for the test statistic.

Right-click Des 1 from the **Library**, and select **Edit Design**. Change the **Test Stat.** from **`z`** to **`t`**. The entries for the other fields need not be changed, and click **Compute**. East will add an additional row to the **Output Preview** labeled as Des 3. Select the rows corresponding to

Des 1 and Des 3. This will display Des 1 and Des 3 in the **Output Summary**.

| | Des1 | Des3 |
|---|---|---|
| Mnemonic | MN-1S-PDI | MN-1S-PDI |
| **Test Parameters** | | |
| Design Type | Superiority | Superiority |
| No. of Looks | 1 | 1 |
| Test Type | 1-Sided | 1-Sided |
| Specified α | 0.05 | 0.05 |
| Power | 0.902 | 0.9 |
| **Model Parameters** | | |
| Test Statistic | Z | t |
| Input Method | Individual Means | Individual Means |
| Mean Control (μc) | 6 | 6 |
| Mean Treatment (μt) | 4 | 4 |
| Diff. of Means (μt - μc) | -2 | -2 |
| Std. Deviation (σD) | 5 | 5 |
| **Sample Size** | | |
| Maximum | 54 | 55 |

Using the t distribution, we need one extra subject to compensate for the extra variability due to estimation of the var$[\hat{\delta}]$.

## 4.3 Ratio of Paired Means

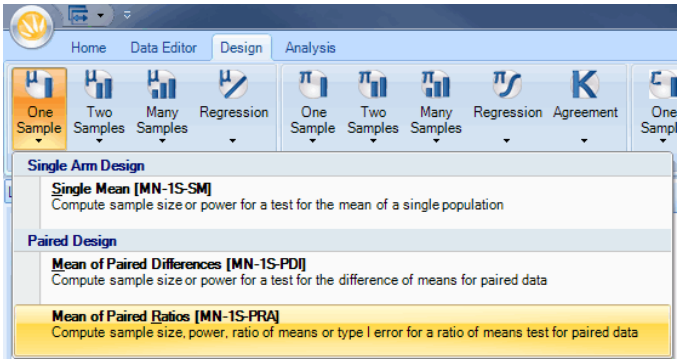
The test for ratio of paired difference is used to compare the means of two log normal distributions when each observation in the random sample from one distribution is matched with a unique observation from the other distribution. Let $\mu_c$ and $\mu_t$ denote the two means to be compared and let $\sigma_c^2$ adn $\sigma_t^2$ are the respective variances.

The null hypothesis $H_0$: $\mu_c/\mu_t = 1$ is tested against the two-sided alternative hypothesis $H_1$: $\mu_c/\mu_t \neq 1$ or a one-sided alternative hypothesis $H_1$: $\mu_c/\mu_t < 1$ or $H_1$: $\mu_c/\mu_t > 1$. Let $\rho = \mu_t/\mu_c$. Then the null hypothesis can be expressed as $H_0$: $\rho = 1$ and the alternative can be expressed as $H_1$: $\rho \neq 1$, $H_1$: $\rho > 1$, or $H_1$: $\rho < 1$. The power of the test is computed at specified values of $\mu_c, \mu_t,$ and $\sigma$. We assume that $\sigma_t/\mu_t = \sigma_c/\mu_c$ i.e., the coefficient of variation (CV) is the same under both control and treatment.

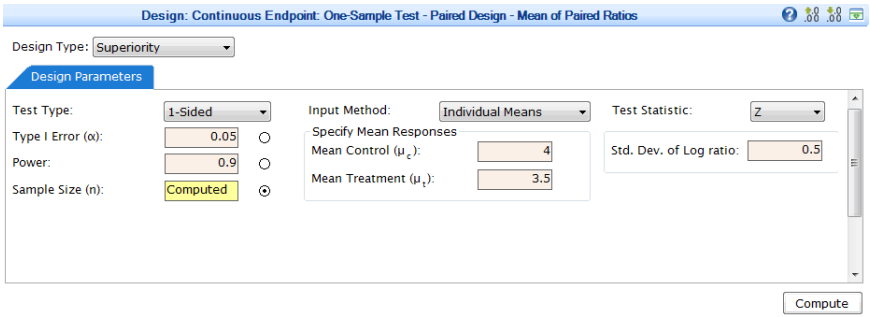
### 4.3.1 Trial Design

Start East afresh. Click **Continuous: One Sample** on the **Design** tab, and then click **Paired**

**Design: Mean of Paired Ratios** as shown below.




This will launch a new window. The upper pane of this window displays several fields with default values. Select **Test Type** as `1-Sided`, and `Individual Means` for **Input Method**. Specify the **Mean Control ($\mu_c$)** as $4$ and **Mean Treatment ($\mu_t$)** as $3.5$. Enter **Std. Dev. of Log ratio** as $0.5$. The upper pane should appear as below:




Click **Compute**. This will calculate the sample size for this design and the output is shown as a row in the **Output Preview**. The computed sample size is $121$ subjects (or pairs of observations).

This design has default name Des 1. In the **Output Preview** toolbar select Des 1, click  to save this design to Wbk1 in the **Library**.

### 4.3.2 Trial Design Using a t-test

Right-click Des 1 in the **Library** and select **Edit Design**. In the input window, change the **Test Stat.** from `z` to `t`.

Click **Compute**. East will add an additional row to the **Output Preview** labeled as Des 2.

Select the rows corresponding to Des 1 and Des 2 using the Ctrl key and click  . This will display Des 1 and Des 2 in the **Output Summary**.

| | Des1 | Des2 |
|---|---|---|
| Mnemonic | MN-1S-PRA | MN-1S-PRA |
| **Test Parameters** | | |
| Design Type | Superiority | Superiority |
| Test Type | 1-Sided | 1-Sided |
| Specified α | 0.05 | 0.05 |
| Power | 0.902 | 0.901 |
| **Model Parameters** | | |
| Mean Control (μc) | 4 | 4 |
| Mean Treatment (μt) | 3.5 | 3.5 |
| Std.Dev. of Log Ratio | 0.5 | 0.5 |
| Input Method | Individual Means | Individual Means |
| Test Statistic | Z | t |
| **Sample Size** | | |
| Maximum | 121 | 122 |

Des 2 uses the t distribution and requires that we commit a combined total of 122 patients to the study, one more compared to Des 1, which uses a normal distribution.

# 5 Normal Noninferiority Paired-Sample

Two common applications of the paired sample design include: (1) comparison of two treatments where patients are matched on demographic and baseline characteristics, and (2) two observations made from the same patient under different experimental conditions. The type of endpoint for paired noninferiority design could be difference of means or ratio of means. The former is presented in Section 5.1 and the latter is discussed in Section 5.2. For paired sample noninferiority trials, East can be used only when no interim look is planned.

## 5.1 Mean of Paired Differences

▪ 5.1.1 Trial Design ▪ 5.1.2 Trial Design Using a t-Test (Single Look) ▪ 5.1.3 Simulation

Consider a randomized clinical trial comparing an experimental treatment, T, to a control treatment, C, on the basis of outcome variable, X, with means $\mu_t$ and $\mu_c$, respectively, and with a standard deviation of paired difference as $\sigma_D^2$. Here, the null hypothesis $H_0$: $\mu_t - \mu_c \leq \delta_0$ is tested against the one-sided alternative hypothesis $H_1$: $\mu_t - \mu_c > \delta_0$. Here $\delta_0$ denotes the noninferiority margin and $\delta_0 < 0$. Let $\delta = \mu_t - \mu_c$. Then the null hypothesis can be expressed as $H_0$: $\delta \leq \delta_0$ and the alternative can be expressed as $H_1$: $\delta > \delta_0$.

Here we assume that the each paired observation on X from T and C are distributed according to a bivariate normal distribution with means as ($\mu_t$, $\mu_c$) , variances as ($\sigma_t^2$, $\sigma_c^2$) and correlation coefficient as $\rho$. Let us have $N$ such paired observations from T and C and $\hat{\mu}_c$ and $\hat{\mu}_t$ denote the estimates of $\mu_c$ and $\mu_t$ based on these $N$ pairs. Therefore, the estimate of the difference is $\hat{\delta} = \hat{\mu}_t - \hat{\mu}_c$. Denoting the standard error of $\hat{\delta}$ by $se(\hat{\delta})$, the test statistic can be defined as

$$Z = \frac{\hat{\delta} - \delta_0}{se(\hat{\delta})} \tag{5.1}$$

The test statistic $Z$ is distributed as a t distribution with $(N-1)$ degrees of freedom. For large samples, the t-distribution can be approximated by the standard normal distribution. The power of the test is computed at specified values of $\mu_c, \mu_t$, and $\sigma_D$. East allows you to analyze using both normal and t distribution.

The advantage of the paired sample noninferiority design compared to the two independent sample noninferiority design lies in the smaller $se(\hat{\delta})$ in former case. The paired sample design is more powerful than the two independent sample design: to achieve the same level of power, the paired sample design requires fewer subjects.

### 5.1.1 Trial Design

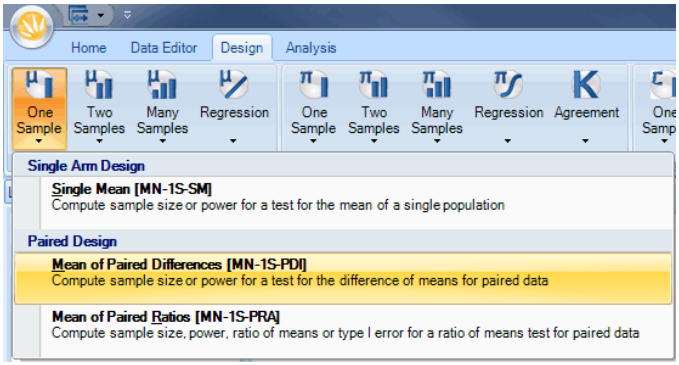
Iezzi et. al. (2011) investigated the possibility of reducing radiation dose exposure while maintaining the image quality in a prospective, single center, intra-individual study. In this study, patients underwent two consecutive multidetector computed tomography angiography (MDCTA) scans 6 months apart, one with a standard acquisition protocol (C) and another using a low dose protocol (T). Image quality was rated as an ordinal number using a rating scale ranging from 1 to 5. Let $\mu_c$ and $\mu_t$ denote the average rating of image quality for standard acquisition and low dose protocol, respectively, and $\delta = \mu_t - \mu_c$ be the difference between two means. Based on the 30 samples included in the study, $\mu_c$ and $\mu_t$ were estimated as 3.67 and 3.12, respectively. The noninferiority margin for image quality considered was $-1$. Accordingly, we will design the study to test

$$H_0 : \delta \leq -1 \qquad \text{against} \qquad H_1 : \delta > -1$$

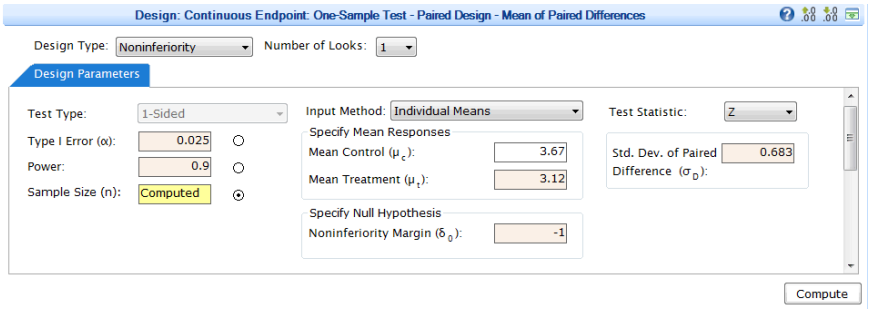
The standard deviation of paired difference was estimated as 0.683. We want to design a study with 90% power at $\mu_c = 3.67$ and $\mu_t = 3.12$ and that maintains overall one-sided type I error of 0.025.

First, click **Continuous: One Sample** on the **Design** tab and then click **Paired Design: Mean**

**of Paired Differences** as shown below.




This will launch a new window. Select **`Noninferiority`** for **Design Type**, and **`Individual Means`** for **Input Method**. Specify the **Mean Control ($\mu_c$)** as $3.67$, **Mean Treatment ($\mu_t$)** as $3.12$, and the **Std. Dev. of Paired Difference ($\sigma_D$)** as $0.683$. Finally, enter $-1$ for the **Noninferiority Margin ($\delta_0$)**. Leave all other entries with their default values. The upper pane should appear as below:




Click **Compute**. This will calculate the sample size for this design and the output is shown as a row in the **Output Preview** located in the lower pane of this window. The computed sample

size (25 subjects) is highlighted.

| ▲ | ID | Design Type | No. of Looks | Test Type | Specified α | Power | Sample Size | Test Statistic | Input Method | μc | Mean Treatment (Alt.) | δ1 | δ0 | σD |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Des1 | Noninferiority | 1 | 1-Sided | 0.025 | 0.909 | 25 | Z | Individual Means | 3.67 | 3.12 | -0.55 | -1 | 0.683 |

This design has default name Des 1. You can select this design by clicking anywhere along the row in the **Output Preview**. Select this design and click in the **Output Preview** toolbar. Some of the design details will be displayed in the upper pane, labeled as **Output Summary**.

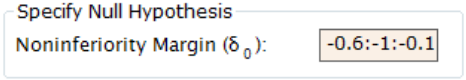
Output Summary

| | Des1 |
|---|---|
| Mnemonic | MN-1S-PDI |
| **Test Parameters** | |
| Design Type | Noninferiority |
| No. of Looks | 1 |
| Test Type | 1-Sided |
| Specified α | 0.025 |
| Power | 0.909 |
| **Model Parameters** | |
| Test Statistic | Z |
| Input Method | Individual Means |
| Mean Control (μc) | 3.67 |
| Mean Treatment (μt) | 3.12 |
| Diff. of Means (μt - μc) | -0.55 |
| Noninferiority Margin (δ0) | -1 |
| Std. Deviation (σD) | 0.683 |
| **Sample Size** | |
| Maximum | 25 |

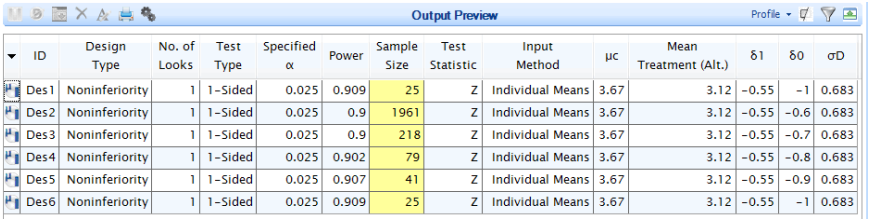
A total of 25 subjects must be enrolled in order to achieve the desired 90% power under the alternative hypothesis. In the **Output Preview** select Des 1 and click in the toolbar to save this design to Wbk1 in the **Library**.

The noninferiority margin of $-1$ considered above is the minimal margin. Since the observed difference is only little less than -0.5 we would like to calculate sample size for a range of noninferiority margins, say, $-0.6$, $-0.7$, $-0.8$, $-0.9$ and $-1$. This can be done easily in East. First select Des 1 in the **Library**, and click on the **Library** toolbar. In the Input, change the

**Noninferiority Margin ($\delta_0$)** $-0.6 : -1 : -0.1$.




Click **Compute** to generate sample sizes for different noninferiority margins. This will add 5 new rows to the **Output Preview**. There will be a single row for each of the noninferiority margins.




| ID | Design Type | No. of Looks | Test Type | Specified α | Power | Sample Size | Test Statistic | Input Method | μc | Mean Treatment (Alt.) | δ1 | δ0 | σD |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Des1 | Noninferiority | 1 | 1-Sided | 0.025 | 0.909 | 25 | Z | Individual Means | 3.67 | 3.12 | -0.55 | -1 | 0.683 |
| Des2 | Noninferiority | 1 | 1-Sided | 0.025 | 0.9 | 1961 | Z | Individual Means | 3.67 | 3.12 | -0.55 | -0.6 | 0.683 |
| Des3 | Noninferiority | 1 | 1-Sided | 0.025 | 0.9 | 218 | Z | Individual Means | 3.67 | 3.12 | -0.55 | -0.7 | 0.683 |
| Des4 | Noninferiority | 1 | 1-Sided | 0.025 | 0.902 | 79 | Z | Individual Means | 3.67 | 3.12 | -0.55 | -0.8 | 0.683 |
| Des5 | Noninferiority | 1 | 1-Sided | 0.025 | 0.907 | 41 | Z | Individual Means | 3.67 | 3.12 | -0.55 | -0.9 | 0.683 |
| Des6 | Noninferiority | 1 | 1-Sided | 0.025 | 0.909 | 25 | Z | Individual Means | 3.67 | 3.12 | -0.55 | -1 | 0.683 |

The computed sample sizes are 1961, 218, 79, 41 and 25 with noninferiority margins $-0.60$, $-0.7$, $-0.8$, $-0.9$ and $-1$, respectively. To compare all 5 designs, select last 5 rows in **Output Preview**, and click  . The 5 designs will be displayed in the **Output Summary** pane.

| | Des2 | Des3 | Des4 | Des5 | Des6 |
|---|---|---|---|---|---|
| Mnemonic | MN-1S-PDI | MN-1S-PDI | MN-1S-PDI | MN-1S-PDI | MN-1S-PDI |
| **Test Parameters** | | | | | |
| Design Type | Noninferiority | Noninferiority | Noninferiority | Noninferiority | Noninferiority |
| No. of Looks | 1 | 1 | 1 | 1 | 1 |
| Test Type | 1-Sided | 1-Sided | 1-Sided | 1-Sided | 1-Sided |
| Specified α | 0.025 | 0.025 | 0.025 | 0.025 | 0.025 |
| Power | 0.9 | 0.9 | 0.902 | 0.907 | 0.909 |
| **Model Parameters** | | | | | |
| Test Statistic | Z | Z | Z | Z | Z |
| Input Method | Individual Means | Individual Means | Individual Means | Individual Means | Individual Means |
| Mean Control (μc) | 3.67 | 3.67 | 3.67 | 3.67 | 3.67 |
| Mean Treatment (μt) | 3.12 | 3.12 | 3.12 | 3.12 | 3.12 |
| Diff. of Means (μt - μc) | -0.55 | -0.55 | -0.55 | -0.55 | -0.55 |
| Noninferiority Margin (δ0) | -0.6 | -0.7 | -0.8 | -0.9 | -1 |
| Std. Deviation (σD) | 0.683 | 0.683 | 0.683 | 0.683 | 0.683 |
| **Sample Size** | | | | | |
| Maximum | 1961 | 218 | 79 | 41 | 25 |

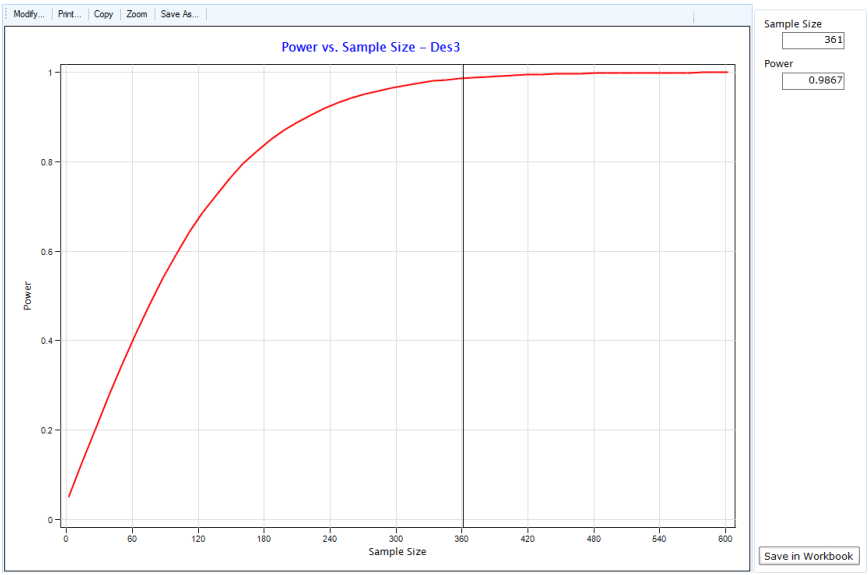
Suppose we have decided to go with Des 3 to test the noninferiority hypothesis with noninferiority margin of $-0.7$. This requires a total sample size of 218 to achieve 90% power. Select Des 3 in the **Output Preview** and click  in the toolbar to save this design to

## Chapter 5: *Normal Noninferiority Paired-Sample*

Wbk1 in the **Library**. Before we proceed we would like to delete all designs from the **Output Preview**. Select all rows and then either click  in the toolbar, or click **Delete** after right click. To delete the designs from the workbook in **Library** select the corresponding designs individually (one at a time) and then click **Delete** after right click. You can try deleting Des 1 from the **Library**.

***Plotting***

With Des 3 selected in the **Library**, click  on the **Library** toolbar, and then click **Power vs Sample Size**. The resulting power curve for this design will appear.




You can move the vertical bar along the X axis. To find out power at any sample size, move the vertical bar to that sample size and the numerical value of sample size and power will be displayed on the right of the plot.You can export this chart in one of several image formats (e.g., Bitmap or JPEG) by clicking **Save As...**. Close this chart before continuing. In a similar fashion one can see power vs delta plot by clicking  and then **Power vs Treatment Effect**.

You can obtain the tables associated with these plot by clicking , and then clicking the appropriate table. Close the plots before continuing.

### 5.1.2 Trial Design Using a t-Test (Single Look)

The sample size obtained to correctly power Des 3 relied on using a Wald-type statistic for the hypothesis test. Due to the assumption of a normal distribution for the test statistic, we have ignored the fact that the variance $\sigma$ is estimated from the sample. For large sample sizes, this approximation is acceptable. However, in small samples with unknown standard deviation, the test statistic

$$Z = (\hat{\delta} - \delta_0)/se(\hat{\sigma})$$

is distributed as Student's t distribution with $(n-1)$ degrees of freedom where $n$ is the number of paired observations.

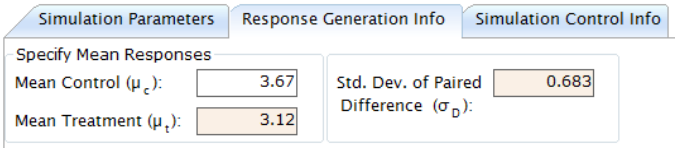
Select Des 3 from the **Library**, and click . This will take you to the input window. Now change the **Test Statistic** from `z` to `t`. The entries for the other fields need not be changed.

Click **Compute**. East will add an additional row to the **Output Preview**. The required sample size is $220$. This design uses the t distribution and it requires us to commit a combined total of 220 patients to the study, two more compared to Des 3 which uses the normal distribution. The extra couple of patients are needed to compensate for the extra variability due to estimation of the var$[\hat{\delta}]$.

### 5.1.3 Simulation

Select Des 3 in the **Library**, and click in the toolbar. Alternatively, right-click on Des 3 and select **Simulate**. A new Simulation window will appear. Click on the **Response Generation Info** tab, and specify: **Mean control** = $3.67$, **Mean Treatment** = $3.12$, and **Std.**

**Deviation of Paired Difference ($\sigma_D$)** = $0.683$.




Leave all default values, and click **Simulate**. Once the simulation run has completed, East will add an additional row to the **Output Preview** labeled as Sim 1.

Select Sim 1 in the **Output Preview** and click . Double-click Sim 1 in the **Library**, and the simulation output details will be displayed in the right pane under the **Simulation** tab.

**Simulation: Continuous Endpoint: One-Sample Test - Paired Design - Mean of Paired Differences**

| Simulation Parameters | |
|---|---|
| Simulation ID | Sim1 |
| Design Type | Noninferiority |
| Number of Looks | 1 |
| Test Type | 1-Sided |
| Sample Size (n) | 218 |
| Noninf. Margin ($\delta_0$) | -0.7 |
| Test Statistic | t |
| Avg. Power at Termination | 0.899 |
| **Response Generation Parameters** | |
| Mean Control ($\mu_c$) | 3.67 |
| Mean Treatment ($\mu_t$) | 3.12 |
| Std. Deviation ($\sigma_D$) | 0.683 |
| **Simulation Control Parameters** | |
| Starting Seed | Clock |
| Number of Simulations | 10000 |

**Average Sample Size**

| Look # | Average Sample Size (n) |
|---|---|
| 1 | 218 |
| Average | 218 |

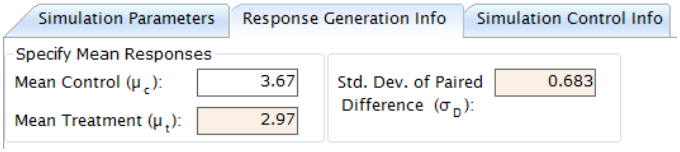
**Simulation Boundaries and Boundary Crossing Probabilities**

| Look # | Sample Size (n) | Boundaries | Stopping For | Total Simulations | |
|---|---|---|---|---|---|
| | | Efficacy | | | |
| | | Upper | Efficacy | Count | % |
| 1 | 218 | 1.96 | 8986 | 10000 | 100.000% |
| Total | | | 8986 | 10000 | |
| % | | | 89.860% | | |

Notice that the percentage of rejections out of 10000 simulated trials is consistent with the design power of 90%. The exact result of the simulations may differ slightly, depending on the seed.

Now we wish to simulate from a point that belongs to $H_0$ to check whether the chosen design maintains type I error of 5%. Right-click Sim 1 in the **Library** and select **Edit Simulation**. Go to the **Response Generation Info** tab in the upper pane and specify: **Mean control** = $3.67$,

**Mean Treatment** = $2.97$, and **Std. Deviation of Paired Difference ($\sigma_D$)** = $0.683$.




Click **Simulate**. Once the simulation run has completed, East will add an additional row to the **Output Preview** labeled as Sim 2. Select Sim 2 in the **Output Preview** and click . Now double-click on Sim 2 in the **Library**. The simulation output details will be displayed.

**Simulation: Continuous Endpoint: One-Sample Test - Paired Design - Mean of Paired Differences**

| Simulation Parameters | |
|---|---|
| Simulation ID | Sim2 |
| Design Type | Noninferiority |
| Number of Looks | 1 |
| Test Type | 1-Sided |
| Sample Size (n) | 218 |
| Noninf. Margin ($\delta_0$) | -0.7 |
| Test Statistic | t |
| Avg. Power at Termination | 0.023 |
| **Response Generation Parameters** | |
| Mean Control ($\mu_c$) | 3.67 |
| Mean Treatment ($\mu_t$) | 2.97 |
| Std. Deviation ($\sigma_D$) | 0.683 |
| **Simulation Control Parameters** | |
| Starting Seed | Fixed |
| Number of Simulations | 10000 |

**Average Sample Size**

| Look # | Average Sample Size (n) |
|---|---|
| 1 | 218 |
| Average | 218 |

**Simulation Boundaries and Boundary Crossing Probabilities**

| Look # | Sample Size (n) | Boundaries | Stopping For | Total Simulations | |
|---|---|---|---|---|---|
| | | Efficacy | | | |
| | | Upper | Efficacy | Count | % |
| 1 | 218 | 1.96 | 233 | 10000 | 100.000% |
| Total | | | 233 | 10000 | |
| % | | | 2.330% | | |

The upper efficacy stopping boundary was crossed close to the specified type I error of 2.5%. The exact result of the simulations may differ slightly, depending on the seed.

## 5.2 Ratio of Paired Means

Consider a randomized clinical trial comparing an experimental treatment, T, to a control treatment, C, on the basis of outcome variable, X, with means $\mu_t$ and $\mu_c$, respectively, and let $\sigma_t^2$ and $\sigma_c^2$ denote the respective variances. The null hypothesis $H_0$: $\mu_t/\mu_c \leq \rho_0$ is tested against the one-sided alternative hypothesis $H_1$: $\mu_t/\mu_c > \rho_0$. Here, $\rho_0$ denotes the noninferiority margin and $\rho_0 < 1$. Let $\rho = \mu_t/\mu_c$. Then the null hypothesis can be expressed as $H_0$: $\rho \leq \rho_0$ and the alternative can be expressed as $H_1$: $\rho > \rho_0$.

Let us have $N$ such paired observations from T and C and $(X_{it}, X_{ic})$ denotes the $i$th pair of observations ($i = 1, \cdots, N$). Then $\log X_{it} - \log X_{ic} = \log(X_{it}/X_{ic})$ denotes the logarithm of ratio of means for $i$th subject. We assume that the paired log-transformed observations on X from T and C, $(\log X_{it}, \log X_{ic})$ are bivariate normally distributed with common parameters. In other words, $(X_{it}, X_{ic})$ is distributed as bivariate log-normal distribution.

Denote $\log X_{it}$ by $y_{it}$, $\log X_{ic}$ by $y_{ic}$, and the corresponding difference by $\delta_{yi} = y_{it} - y_{ic}$. Assume that $\hat{\delta}_y$ denotes the sample mean for these paired differences with estimated standard error $se(\hat{\delta}_y)$. The test statistic can be defined as

$$Z = \frac{\hat{\delta}_y - \log \rho_0}{se(\hat{\delta}_y)}, \tag{5.2}$$

The test statistic $Z$ is distributed as a t distribution with $(N-1)$ degrees of freedom. For large samples, the t-distribution can be approximated by the standard normal distribution. East allows you to analyze using both normal and t distribution. The power of the test is computed at specified values of $\mu_c$, $\mu_t$, and $\sigma$.

### 5.2.1 Trial Design

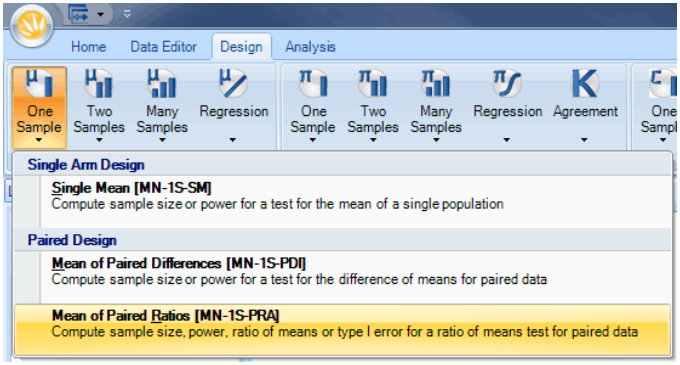
We will use the same example cited in the previous section, but will transform the difference hypothesis into the ratio hypothesis. Let $\mu_c$ and $\mu_t$ denote the average rating of image quality for standard acquisition and low dose protocol, estimated as 3.67 and 3.12, respectively. Let $\rho = \mu_t/\mu_c$ be the ratio between two means. Considering a noninferiority margin of $-0.7$ for the test of difference, we can rewrite the hypothesis mentioned in previous section as

$$H_0 : \rho \leq 0.81 \qquad \text{against} \qquad H_1 : \rho > 0.81$$

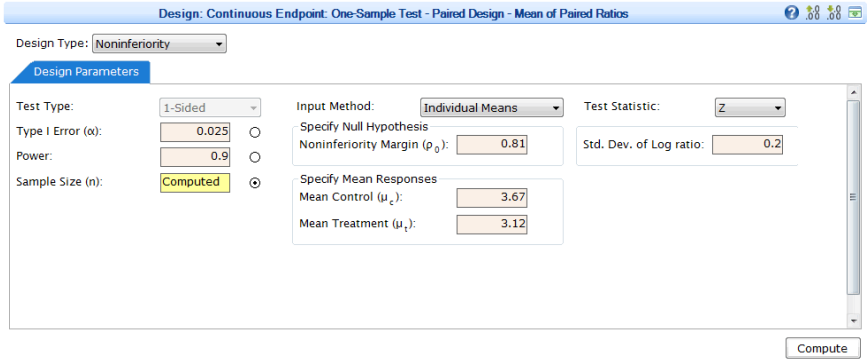
We are considering a noninferirority margin of $0.81 (= \rho_0)$. For illustration we will assume the standard deviation of log ratio as 0.20. As before, we want to design a study with 90% power at $\mu_c = 3.67$ and $\mu_t = 3.12$, and maintains overall one-sided type I error of 0.025.

Start East afresh. Click **Continuous: One Sample** on the **Design** tab and then click **Paired**

**Design: Mean of Paired Ratios** as shown below.

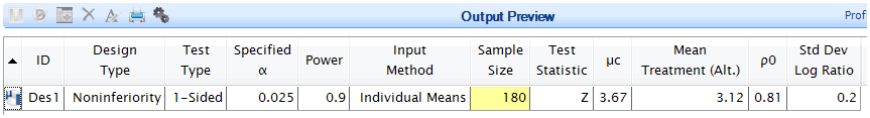



This will launch a new window. The upper pane of this window displays several fields with default values. Select **Noninferiority** for **Design Type**, and **Individual Means** for **Input Method**. Specify the **Mean Control ($\mu_c$)** as $3.67$, **Mean Treatment ($\mu_t$)** as $3.12$, and **Noninferiority margin ($\rho_0$)** as $0.81$. Enter $0.20$ for **Std. Dev. of Log Ratio**, and $0.025$ for **Type I Error ($\alpha$)**. The upper pane now should appear as below:




Click **Compute**. This will calculate the sample size for this design and the output is shown as a row in the **Output Preview** located in the lower pane of this window. The computed sample

size (180 subjects) is highlighted in yellow.




| ID | Design Type | Test Type | Specified α | Power | Input Method | Sample Size | Test Statistic | μc | Mean Treatment (Alt.) | ρ0 | Std Dev Log Ratio |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Des1 | Noninferiority | 1-Sided | 0.025 | 0.9 | Individual Means | 180 | Z | 3.67 | 3.12 | 0.81 | 0.2 |

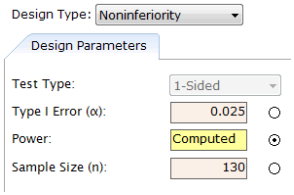
This design has default name Des 1. You can select this design by clicking anywhere along the row in the **Output Preview**. Select this design and click in the **Output Preview** toolbar. Some of the design details will be displayed in the upper pane, labeled as **Output Summary**.

| | Des1 |
|---|---|
| Mnemonic | MN-1S-PRA |
| **Test Parameters** | |
| Design Type | Noninferiority |
| Test Type | 1-Sided |
| Specified α | 0.025 |
| Power | 0.9 |
| **Model Parameters** | |
| Mean Control (μc) | 3.67 |
| Mean Treatment (μt) | 3.12 |
| Noninferiority Margin (ρ0) | 0.81 |
| Std.Dev. of Log Ratio | 0.2 |
| Input Method | Individual Means |
| Test Statistic | Z |
| **Sample Size** | |
| Maximum | 180 |

A total of 180 subjects must be enrolled in order to achieve the desired 90% power under the alternative hypothesis. In the **Output Preview** select Des 1 and click in the toolbar to save this design to Wbk1 in the **Library**.

Suppose you think enrolling 180 subjects is too much for your organization and you can go up to only 130 subjects. You want to evaluate the power of your study at sample size 130 but with the design parameters remain unaltered. In order to compute power with 130 subjects, first select the Des 1 in the **Library**, and click on the **Library** toolbar. In the Input dialog

box, first select the radiobutton for **Power**, and then enter 130 for **Sample Size**.




Now click **Compute**. This will add another row labeled as Des 2 in **Output Preview** with computed power highlighted in yellow. The design attains a power of 78.7%. Now select both the rows in **Output Preview** by pressing the Ctrl key, and click  in the **Output Preview** toolbar to see a summary of both designs in the **Output Summary**.

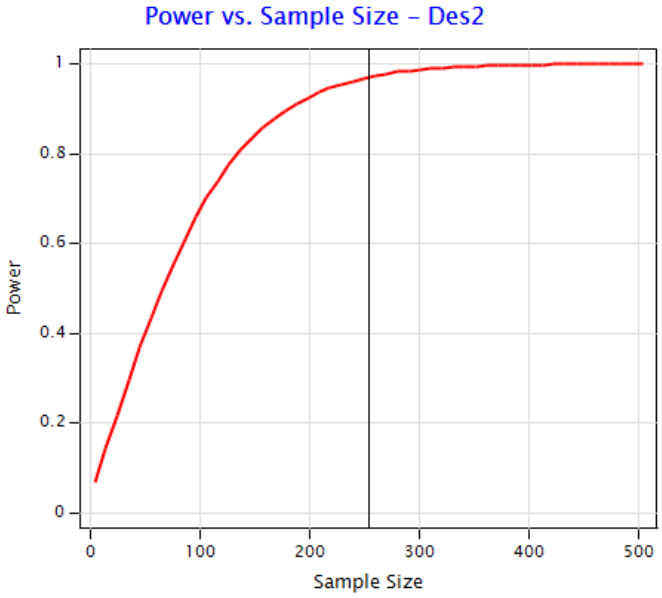
| | Des1 | Des2 |
|---|---|---|
| Mnemonic | MN-1S-PRA | MN-1S-PRA |
| **Test Parameters** | | |
| Design Type | Noninferiority | Noninferiority |
| Test Type | 1-Sided | 1-Sided |
| Specified α | 0.025 | 0.025 |
| Power | 0.9 | 0.787 |
| **Model Parameters** | | |
| Mean Control (μc) | 3.67 | 3.67 |
| Mean Treatment (μt) | 3.12 | 3.12 |
| Noninferiority Margin (ρ0) | 0.81 | 0.81 |
| Std.Dev. of Log Ratio | 0.2 | 0.2 |
| Input Method | Individual Means | Individual Means |
| Test Statistic | Z | Z |
| **Sample Size** | | |
| Maximum | 180 | 130 |

In the **Output Preview** select Des 2 and click  in the toolbar to save this design to Wbk1 in the **Library**.

***Plotting***

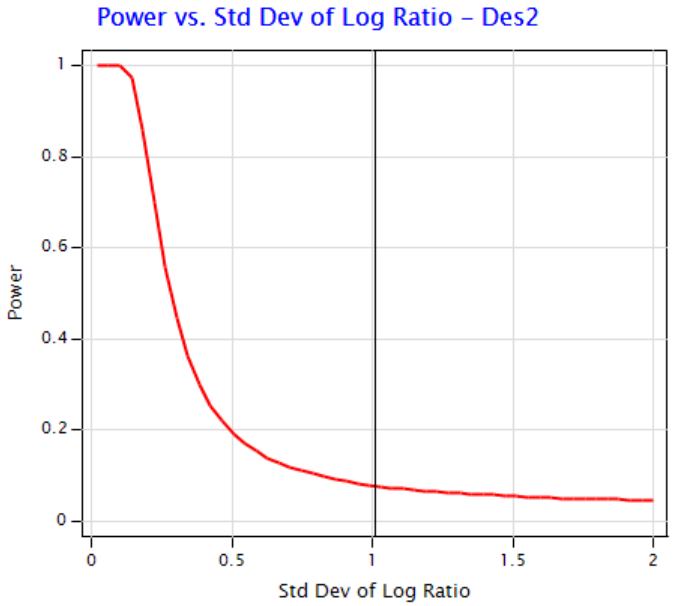
With Des 2 selected in the **Library**, click  on the **Library** toolbar, and then click **Power vs Sample Size** . The resulting power curve for this design will appear. You can move the

vertical bar along the X axis.




Suppose you would like to explore the relationship between power and standard deviation. In order to visualize this relationship, select Des 2 in the **Library**, click on the **Library** toolbar, and then click **General (User Defined Plot)**. Select `Std Dev of Log Ratio` for

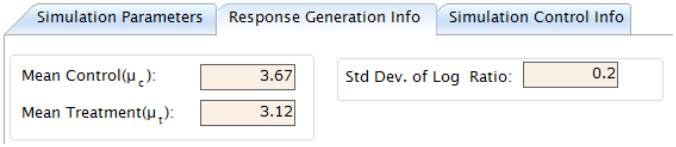
**X-Axis**. This will display the power vs. standard deviation plot.




Close the plot window before you continue.

### 5.2.2 Simulation

Select Des 2 in the **Library**, and click  in the toolbar. Alternatively, right-click on Des 2 and select **Simulate**. A new Simulation window will appear. Click on the **Response Generation Info** tab, and specify: **Mean control** = 3.67, **Mean Treatment** = 3.12, and **Std Dev of Log Ratio**= 0.2.



Click **Simulate**. Once the simulation run has completed, East will add an additional row to the **Output Preview** labeled as Sim 1.

Select Sim 1 in the **Output Preview** and click  . Now double-click on Sim 1 in the **Library**. The simulation output details will be displayed.

**Simulation: Continuous Endpoint: One-Sample Test - Paired Design - Mean of Paired Ratios**

| Simulation Parameters | |
|---|---|
| Simulation ID | Sim1 |
| Trial Type | Noninferiority |
| Test Type | 1-Sided |
| Sample Size (n) | 130 |
| Test Statistic | t |
| Noninferiority Margin ($\rho_0$) | 0.81 |
| **Response Generation Parameters** | |
| Mean Response under Control ($\mu_c$) | 3.67 |
| Mean Response under Treatment ($\mu_t$) | 3.12 |
| Simulation Std. Dev. of Log Ratio | 0.2 |
| **Simulation Control Parameters** | |
| Starting Seed | Clock |
| Number of Simulations | 10000 |

**Simulation Boundaries**

Critical Point: 1.96

**Overall Simulation Results**

| | Upper H0 | Lower H0 |
|---|---|---|
| No.of Rejections | 7832 | NA |
| % | 78.32 | NA |

Starting Seed: 6641254
Total Number of Simulations: 10000
Elapsed Time: 00:00:05

# Volume 3 Binomial and Categorical Endpoints

# 6 Binomial Superiority One-Sample

This chapter deals with the design, simulation, and interim monitoring of two types of tests involving binomial response rates. In Section 6.1, we discuss group sequential designs in which an observed binomial response rate is compared to a fixed response rate, possibly derived from historical data. Section 6.2 deals with McNemar's test for comparing matched pairs of binomial responses in a group sequential setting.

## 6.1 Binomial One Sample



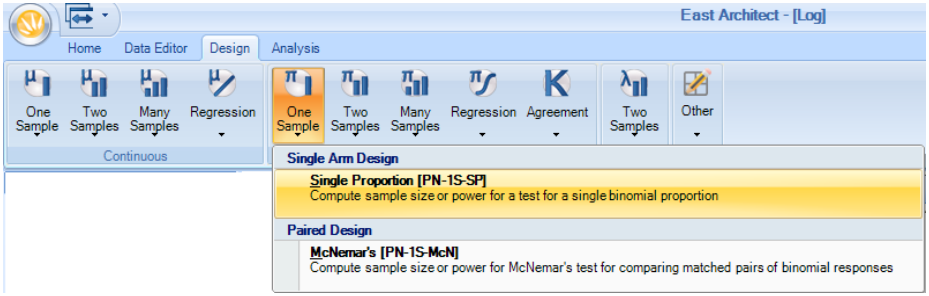
In experimental situations, where the variable of interest has a binomial distribution, it may be of interest to determine whether the response rate $\pi$ differs from a fixed value $\pi_0$. Specifically we wish to test the null hypothesis $H_0$: $\pi = \pi_0$ against the two sided alternative hypothesis $H_1$: $\pi \neq \pi_0$ or against one sided alternatives of the form $H_1$: $\pi > \pi_0$ or $H_1$: $\pi < \pi_0$. The sample size, or power, is determined for a specified value of $\pi$ which is consistent with the alternative hypothesis, denoted $\pi_1$.

### 6.1.1 Trial Design

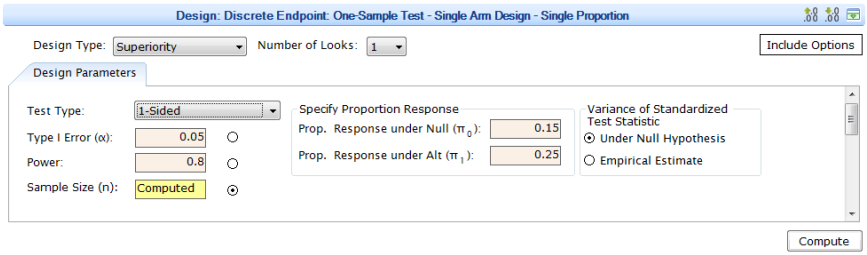
Consider the design of a single-arm oncology trial in which we wish to determine if the tumor response rate of a new cytotoxic agent is at least 15%. Thus, it is desired to test the null hypothesis $H_0$: $\pi = 0.15$ against the one-sided alternative hypothesis $H_1$: $\pi > 0.15$. We will design this trial with a one sided test that achieves 80% power at $\pi = \pi_1 = 0.25$ with a one-sided level 0.05 test.

***Single-Look Design*** To begin, click **Design** tab, then **Single Sample** under **Discrete** group,

and then click **Single Proportion**.

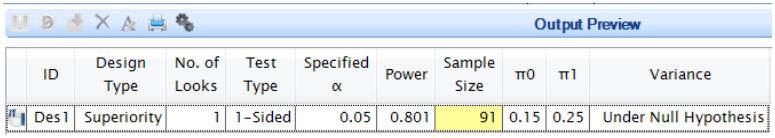



In the ensuing dialog box , choose the design parameters as shown below. We first consider a single-look design, so leave the default value for **Number of Looks** to $1$. In the drop down menu, next to **Test Type** select `1-Sided`. Enter $0.8$ for **Power**. Enter $0.15$ in the box next to **Prop. Response under Null ($\pi_0$)** and $0.25$ in the box next to **Prop. Response under Alt ($\pi_1$)**. This dialog box also asks us to specify whether we wish to standardize the test statistic (for performing the hypothesis test of the null hypothesis $H_0$: $\pi = 0.15$) with the null or the empirical variance. We will discuss the test statistic and the method of standardization in the next subsection. For the present, select the default radio button **Under Null Hypothesis**.

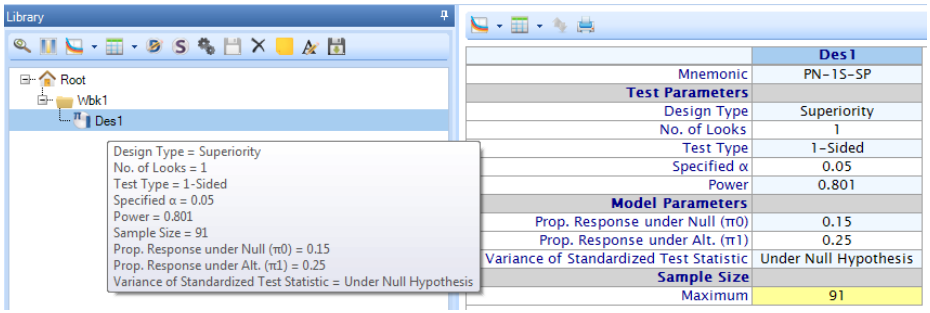



Now click **Compute**. The design is shown as a row in the Output Preview located in the lower pane of this window. The sample size required in order to achieve the desired 80% power is 91

subjects.

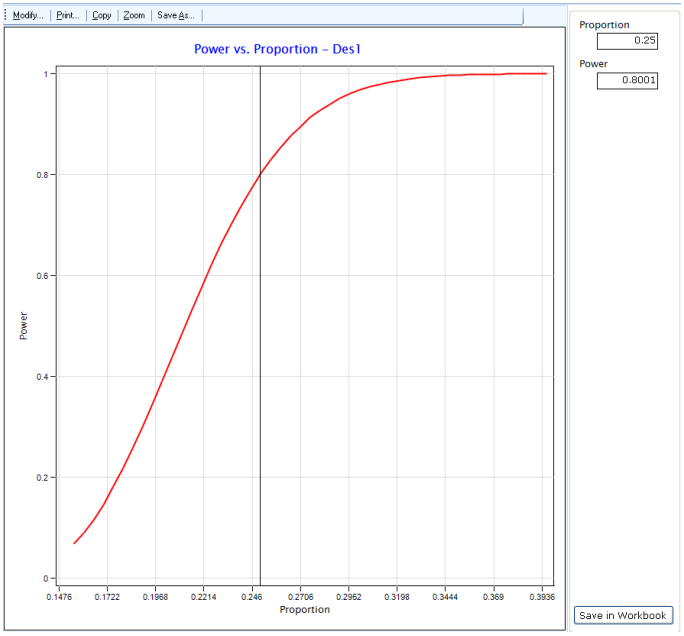



You can select this design by clicking anywhere on the row in the **Output Preview**. Click icon to get the design output summary displayed in the upper pane. In the **Output Preview** toolbar, click icon to save this design Des1 to workbook Wbk1 in the **Library**. If you hover the cursor over the node Des1 in the Library, a tooltip will appear that summarizes the input parameters of the design.



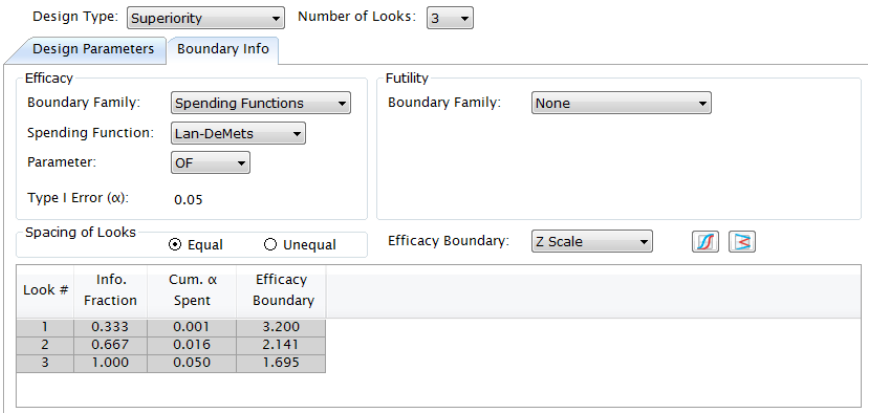

With the design Des1 selected in the Library, click icon on the Library toolbar, and then click **Power vs. Treatment Effect ($\delta$)**. The power curve for this design will be displayed. You can save this chart to the Library by clicking **Save in Workbook**. Alternatively, you can export the chart in one of several image formats (e.g., Bitmap or JPEG) by clicking **Save As...**. For now,

you may close the chart before continuing.




***Three-Look Design*** In order to reach an early decision and enter into comparative trials, let us plan to conduct this single-arm study as a group sequential trial with a maximum of 3 looks. Create a new design by selecting Des1 in the **Library**, and clicking the  icon on the **Library** toolbar. Change the **Number of Looks** from 1 to 3, to generate a study with two interim looks and a final analysis. A new tab **Boundary Info** will appear. Clicking on this tab will reveal the stopping boundary parameters. By default, the **Spacing of Looks** is set to **Equal**, which means that the interim analyses will be equally spaced in terms of the number of patients accrued between looks. The left side contains details for the **Efficacy** boundary, and the right side for the **Futility** boundary. By default, there is an efficacy boundary (to reject H0) selected, but no futility boundary (to reject H1). The **Boundary Family** specified is of the `Spending Functions` type. The default **Spending function** is the Lan-DeMets (Lan & DeMets, 1983), with **Parameter** `OF` (O'Brien-Fleming), which generates boundaries that are very similar,

though not identical, to the classical stopping boundaries of O'Brien and Fleming (1979). Technical details of these stopping boundaries are available in Appendix **??**.




Return to the design parameters by clicking **Design Parameters** tab. The dialog box requires us to make a selection in the section labeled **Variance of Standardized Test Statistic**. We are being asked to specify to East how we intend to standardize the test statistic when we actually perform the hypothesis tests at the various monitoring time points. There are two options: **Under Null Hypothesis** and **Empirical Estimate**. To understand the difference between these two options, let $\hat{\pi}_j$ denote the estimate of $\pi$ based on $n_j$ observations, up to and including the $j$ th monitoring time point.

**Under Null Hypothesis** The test statistic to be used for the interim monitoring is

$$Z_j^{(N)} = \frac{\hat{\pi}_j - \pi_0}{\sqrt{\pi_0(1-\pi_0)/n_j}}\,. \tag{6.1}$$

**Empirical** The test statistic to be used for the interim monitoring is

$$Z_j^{(E)} = \frac{\hat{\pi}_j - \pi_0}{\sqrt{\hat{\pi}_j(1-\hat{\pi}_j)/n_j}}\,. \tag{6.2}$$

The choice of variance should not make much of a difference to the type 1 error or power for

studies in which the sample size is large. In the present case however, it might matter. We shall therefore examine both the options. First, we select the **Under Null Hypothesis** radio button.

Click **Compute** button to generate output for Design Des2. With Des2 selected in the **Output Preview**, click  icon to save Des2 to the **Library**. In order to see the stopping probabilities, as well as other characteristics, select Des2 in the **Library**, and click  icon. The cumulative boundary stopping probabilities are shown in the **Stopping Boundaries** table. We see that for Des2 the maximum sample size is 91 subjects, with 90 expected under the null hypothesis $H_0$: $\pi = 0.15$ and 73 expected when the true value is $\pi$=0.25.

**Design: Discrete Endpoint: One-Sample Test - Single Arm Design - Single Proportion**

| Test Parameters | |
|---|---|
| Design ID: | Des2 |
| Design Type: | Superiority |
| Number of Looks: | 3 |
| Test Type: | 1-Sided |
| Specified α: | 0.05 |
| Power: | 0.801 |
| **Model Parameters** | |
| Prop. Response under Null ($\pi_0$): | 0.15 |
| Prop. Response under Alt. ($\pi_1$): | 0.25 |
| Variance of Std. Test Stat.: | Under Null Hypothesis |
| **Boundary Parameters** | |
| Spacing of Looks: | Equal |
| Efficacy Boundary: | LD (OF) |

**Stopping Boundaries: Look by Look**

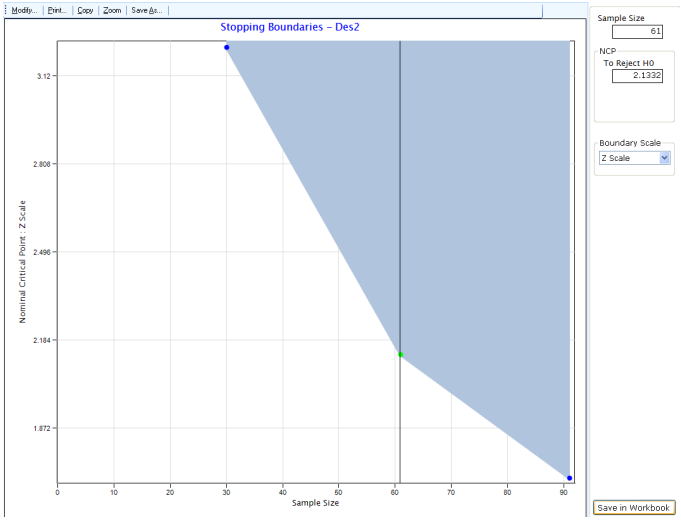
| Look # | Info. Fraction (n/n_max) | Sample Size (n) | Cumulative α Spent | Boundaries | Incr. Boundary Crossing Prob. | |
|---|---|---|---|---|---|---|
| | | | | | Under H0 | Under H1 |
| | | | | Efficacy Z | Efficacy | Efficacy |
| 1 | 0.33 | 30 | 6.412E-4 | 3.22 | 6.412E-4 | 0.082 |
| 2 | 0.67 | 61 | 0.017 | 2.133 | 0.016 | 0.439 |
| 3 | 1 | 91 | 0.05 | 1.696 | 0.033 | 0.28 |

**Sample Size Information:**

| | Maximum | Expected H1 | Expected H0 |
|---|---|---|---|
| Sample Size (n) | 91 | 72.823 | 90.48 |
| Information | 485.333 | 388.388 | 482.56 |

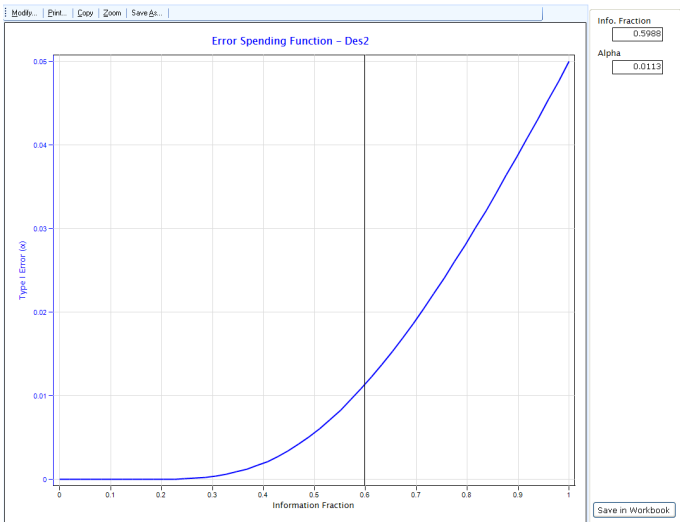
Close the Output window before continuing. The stopping boundary can be displayed by clicking on the  icon on the **Library** toolbar, and then clicking **Stopping Boundaries**.

The following chart will appear.

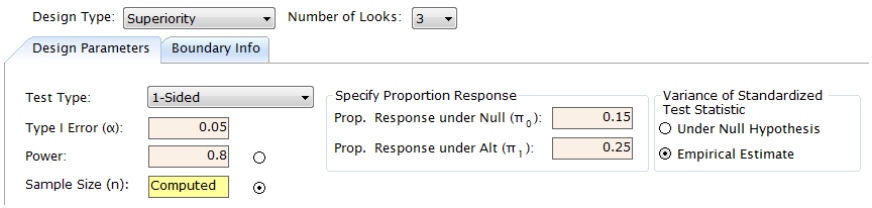



To examine the error spending function, click icon on the **Library** toolbar, and then click

**Error Spending**. The following chart will appear.




To examine the impact of using the empirical variance to standardized test statistic, select Des2 in the **Library**, and click icon on the **Library** toolbar. In the **Variance of Standardized Test Statistic** box, now select **Empirical Estimate**.




Next, click **Compute**. With Des3 selected in the **Output Preview**, click icon. In the **Library**, select the nodes Des2 and Des3, by holding the Ctrl key, and then click icon.

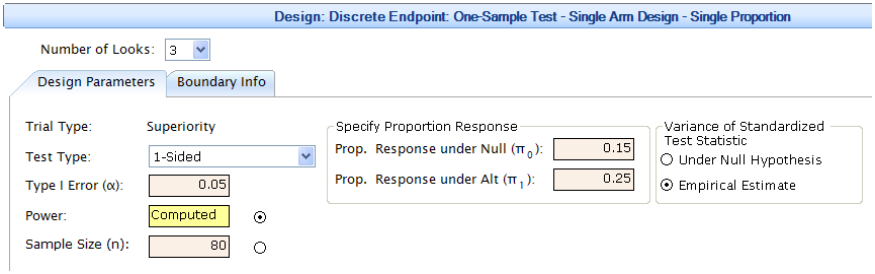
The upper pane will display the summary details of the two designs side-by-side:

| | Wbk1:Des2 | Wbk1:Des3 |
|---|---|---|
| Mnemonic | PN-1S-SP | PN-1S-SP |
| **Test Parameters** | | |
| Design Type | Superiority | Superiority |
| No. of Looks | 3 | 3 |
| Test Type | 1-Sided | 1-Sided |
| Specified α | 0.05 | 0.05 |
| Power | 0.801 | 0.802 |
| **Model Parameters** | | |
| Prop. Response under Null (π0) | 0.15 | 0.15 |
| Prop. Response under Alt. (π1) | 0.25 | 0.25 |
| Variance of Standardized Test Statistic | Under Null Hypothesis | Empirical Estimate |
| **Boundary Parameters** | | |
| Spacing of Looks | Equal | Equal |
| Efficacy Boundary | LD (OF) | LD (OF) |
| **Sample Size** | | |
| Maximum | 91 | 119 |
| Expected Under H0 | 90.48 | 118.326 |
| Expected Under H1 | 72.823 | 98.803 |

The maximum sample size needed for 80% power is 119, and the expected sample size is 99 under the alternative hypothesis $H_1$ with $\pi_1 = 0.25$, if we intend to standardize the test statistic with the empirical variance. The corresponding maximum and expected sample sizes if the null variance is to be used for the standardization are 91 and 73, respectively. Thus, for this configuration of design parameters, it would appear preferable to specify in advance that the test statistic will be standardized by the null variance. Evidently, this is the option with the smaller maximum and expected sample size. These results, however, are based on the large sample theory developed in Appendix **??**. Since the sample sizes in both Des2 and Des3 are fairly small, it would be advisable to verify that the power and type 1 error of both the plans are preserved by simulating these designs. We show how to simulate these plans in Section 6.1.2.

In some situations, the sample size is subject to external constraints. Then, the power can be computed for a specified maximum sample size. Suppose that in the above situation, using the observed estimates for the computation of the variance, the total sample size is constrained to be at most, 80 subjects. Select Des3 in the **Library** and click  on the **Library** toolbar. Change the selections in the ensuing dialog box so that the trial is now

designed to compute power for a maximum sample size of 80 subjects, as shown below.




Click **Compute** button to generate the output for Design Des4. With Des4 selected in the **Output Preview**, click icon. In the **Library**, select the nodes for Des2, Des3, and Des4 by holding the Ctrl key, and then click icon. The upper pane will display the summary details of the three designs side-by-side:

| | Wbk1:Des2 | Wbk1:Des3 | Wbk1:Des4 |
|---|---|---|---|
| Mnemonic | PN-1S-SP | PN-1S-SP | PN-1S-SP |
| **Test Parameters** | | | |
| Design Type | Superiority | Superiority | Superiority |
| No. of Looks | 3 | 3 | 3 |
| Test Type | 1-Sided | 1-Sided | 1-Sided |
| Specified α | 0.05 | 0.05 | 0.05 |
| Power | 0.801 | 0.802 | 0.655 |
| **Model Parameters** | | | |
| Prop. Response under Null (π0) | 0.15 | 0.15 | 0.15 |
| Prop. Response under Alt. (π1) | 0.25 | 0.25 | 0.25 |
| Variance of Standardized Test Statistic | Under Null Hypothesis | Empirical Estimate | Empirical Estimate |
| **Boundary Parameters** | | | |
| Spacing of Looks | Equal | Equal | Equal |
| Efficacy Boundary | LD (OF) | LD (OF) | LD (OF) |
| **Sample Size** | | | |
| Maximum | 91 | 119 | 80 |
| Expected Under H0 | 90.48 | 118.326 | 79.548 |
| Expected Under H1 | 72.823 | 98.803 | 70.711 |

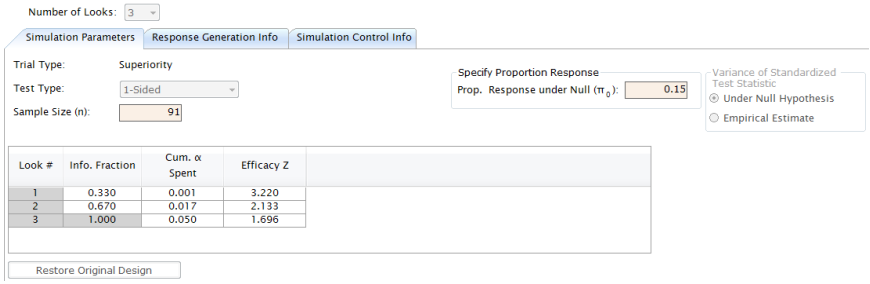
From this, we can see that Des4 has only 65.5 % power.

## *6.1.2 Trial Simulation*

In Section 6.1.1, we created group sequential designs with two different assumptions for the manner in which the test would be standardized at the interim monitoring stage. Under Des2, we assumed that the null variance, and hence the test statistic (6.1) would be used for the

interim monitoring. This plan required a maximum sample size of 91 subjects. Under Des3, we assumed that the empirical variance, and hence the test statistic (6.2) would be used for the interim monitoring. This plan required a maximum sample size of 119 subjects. Since the sample sizes for both plans are fairly small and the calculations involved the use of large sample theory, it would be wise to verify the operating characteristics of these two plans by simulation.

Select Des2 in the **Library**, and click the  icon from **Library** toolbar. Alternatively, right-click on Des2 node and select **Simulate**. A new Simulation worksheet will appear.




Click **Simulate** to start the simulation. Once the simulation run has completed, East will add an additional row to the **Output Preview** labeled Sim1. Select Sim1 row in the **Output Preview** and click  icon. Note that some of the simulation output details will be displayed in the upper pane. Click  icon to save it to the **Library**. Double-click on Sim1 node in the

**Library**. The simulation output details will be displayed.

**Simulation: Discrete Endpoint: One-Sample Test - Single Arm Design - Single Proportion**

| Simulation Parameters | |
|---|---|
| Simulation ID: | Sim1 |
| Design Type: | Superiority |
| Number of Looks: | 3 |
| Test Type: | 1-Sided |
| Prop. Response under Null ($\pi_0$): | 0.15 |
| Variance: | Under Null Hypothesis |
| Avg. Power at Termination: | 0.802 |
| **Response Generation Parameters** | |
| Proportion Response (π): | 0.25 |
| **Simulation Control Parameters** | |
| Starting Seed: | Fixed |
| Number of Simulations: | 10000 |

**Simulation Boundaries and Boundary Crossing Probabilities:**

| Look # | Sample Size (n) | Boundaries Efficacy | Early Stopping For | Total Simulations | |
|---|---|---|---|---|---|
| | | Upper | Efficacy | Count | % |
| 1 | 30 | 3.22 | 978 | 978 | 9.78 |
| 2 | 61 | 2.133 | 4824 | 4824 | 48.24 |
| 3 | 91 | 1.696 | 2217 | 4198 | 41.98 |
| Total | | | 8019 | 10000 | |
| % | | | 80.19 | | |

**Average Sample Size:**

| Look # | Average Sample Size (n) |
|---|---|
| 1 | 30 |
| 2 | 61 |
| 3 | 91 |
| Average | 70.562 |

**Overall Simulation Results**

| | |
|---|---|
| Starting Seed: | 68976527 |
| Total Number of Simulations: | 10000 |
| Elapsed Time: | 00:00:02 |

Upon running 10,000 simulations with $\pi = 0.25$ we obtain slightly over 80% power as shown above.

Next we run 10,000 simulations under $H_0$ by setting $\pi = 0.15$ in the choice of simulation parameters. Select Des2 in the **Library**, and click  icon from **Library** toolbar. Under the **Response Generation Info** tab, change the **Proportion Response** to $0.15$. Click **Simulate** to start the simulation. Once the simulation run has completed, East will add an additional row to the **Output Preview** labeled Sim2. Select Sim2 in the **Output Preview**. Click  icon to save it to the **Library**. Double-click on Sim2 in the **Library**. The simulation output details will

be displayed.

**Simulation: Discrete Endpoint: One-Sample Test - Single Arm Design - Single Proportion**

| Simulation Parameters | |
|---|---|
| Simulation ID: | Sim2 |
| Design Type: | Superiority |
| Number of Looks: | 3 |
| Test Type: | 1-Sided |
| Prop. Response under Null ($\pi_0$): | 0.15 |
| Variance: | Under Null Hypothesis |
| Avg. Power at Termination: | 0.057 |
| **Response Generation Parameters** | |
| Proportion Response (π): | 0.15 |
| **Simulation Control Parameters** | |
| Starting Seed: | Fixed |
| Number of Simulations: | 10000 |

**Simulation Boundaries and Boundary Crossing Probabilities:**

| Look # | Sample Size (n) | Boundaries Efficacy Upper | Early Stopping For Efficacy | Total Simulations Count | Total Simulations % |
|---|---|---|---|---|---|
| 1 | 30 | 3.22 | 22 | 22 | 0.22 |
| 2 | 61 | 2.133 | 304 | 304 | 3.04 |
| 3 | 91 | 1.696 | 244 | 9674 | 96.74 |
| Total | | | 570 | 10000 | |
| % | | | 5.7 | | |

**Average Sample Size:**

| Look # | Average Sample Size (n) |
|---|---|
| 1 | 30 |
| 2 | 61 |
| 3 | 91 |
| Average | 89.954 |

**Overall Simulation Results**

| | |
|---|---|
| Starting Seed: | 68976527 |
| Total Number of Simulations: | 10000 |
| Elapsed Time: | 00:00:01 |

We observe that 6% of these simulations reject the null hypothesis thereby confirming that these boundaries do indeed preserve the type 1 error (up to Monte Carlo accuracy).

Finally we repeat the same set of simulations for Des3. Select Des3 in the **Library**, and click  icon from **Library** toolbar. Upon running 10,000 simulations with $\pi = 0.25$, we obtain

82% power.

**Simulation: Discrete Endpoint: One-Sample Test - Single Arm Design - Single Proportion**

| Simulation Parameters | |
|---|---|
| Simulation ID: | Sim3 |
| Design Type: | Superiority |
| Number of Looks: | 3 |
| Test Type: | 1-Sided |
| Prop. Response under Null ($\pi_0$): | 0.15 |
| Variance: | Empirical Estimate |
| Avg. Power at Termination: | 0.825 |
| **Response Generation Parameters** | |
| Proportion Response ($\pi$): | 0.25 |
| **Simulation Control Parameters** | |
| Starting Seed: | Fixed |
| Number of Simulations: | 10000 |

**Simulation Boundaries and Boundary Crossing Probabilities:**

| Look # | Sample Size (n) | Boundaries Efficacy | Early Stopping For | Total Simulations | |
|---|---|---|---|---|---|
| | | Upper | Efficacy | Count | % |
| 1 | 40 | 3.185 | 247 | 247 | 2.47 |
| 2 | 79 | 2.147 | 4889 | 4889 | 48.89 |
| 3 | 119 | 1.694 | 3114 | 4864 | 48.64 |
| Total | | | 8250 | 10000 | |
| % | | | 82.5 | | |

**Average Sample Size:**

| Look # | Average Sample Size (n) |
|---|---|
| 1 | 40 |
| 2 | 79 |
| 3 | 119 |
| Average | 97.493 |

**Overall Simulation Results**

| | |
|---|---|
| Starting Seed: | 68976527 |
| Total Number of Simulations: | 10000 |
| Elapsed Time: | 00:00:01 |

However, when we run the simulations under $H_0$: $\pi = 0.15$, we obtain a type 1 error of about 3.23% instead of the specified 5% as shown below. While this ensures that the type 1 error is preserved, it also suggests that the use of the empirical variance rather than the null variance to standardize the test statistic might be problematic with small sample sizes.

**Simulation: Discrete Endpoint: One-Sample Test - Single Arm Design - Single Proportion**

| Simulation Parameters | |
|---|---|
| Simulation ID: | Sim4 |
| Design Type: | Superiority |
| Number of Looks: | 3 |
| Test Type: | 1-Sided |
| Prop. Response under Null ($\pi_0$): | 0.15 |
| Variance: | Empirical Estimate |
| Avg. Power at Termination: | 0.032 |
| **Response Generation Parameters** | |
| Proportion Response ($\pi$): | 0.15 |
| **Simulation Control Parameters** | |
| Starting Seed: | Fixed |
| Number of Simulations: | 10000 |

**Simulation Boundaries and Boundary Crossing Probabilities:**

| Look # | Sample Size (n) | Boundaries Efficacy | Early Stopping For | Total Simulations | |
|---|---|---|---|---|---|
| | | Upper | Efficacy | Count | % |
| 1 | 40 | 3.185 | 0 | 0 | 0 |
| 2 | 79 | 2.147 | 108 | 108 | 1.08 |
| 3 | 119 | 1.694 | 215 | 9892 | 98.92 |
| Total | | | 323 | 10000 | |
| % | | | 3.23 | | |

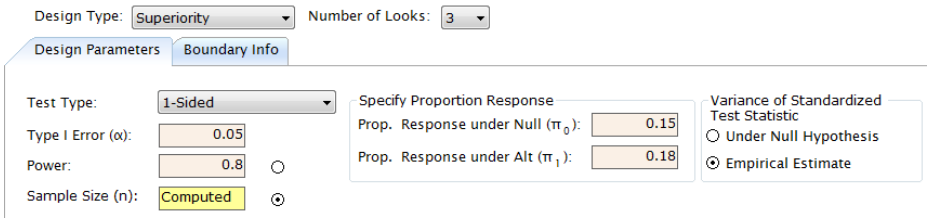
**Average Sample Size:**

| Look # | Average Sample Size (n) |
|---|---|
| 1 | 40 |
| 2 | 79 |
| 3 | 119 |
| Average | 118.568 |

**Overall Simulation Results**

| | |
|---|---|
| Starting Seed: | 68976527 |
| Total Number of Simulations: | 10000 |
| Elapsed Time: | 00:00:01 |

Let us now investigate if the problem disappears with larger studies. Select Des3 in the **Library** and click on the **Library** toolbar. Change the value of **Prop. Response under Alt ($\pi_1$)** from $0.25$ to $0.18$.




Click **Compute** to generate the output for Des5. In the **Output Preview**, we see that Des5 requires a sample size of 1035 subjects. To verify whether the use of the empirical variance will indeed produce the correct type-1 error for this large trial, select Des5 in the **Output Preview** and click icon. In the **Library**, select Des5 and click icon from **Library** toolbar . First, run 10,000 trials with $\pi = 0.15$. On the **Response Generation Info** tab, change **Proportion Response** from $0.18$ to $0.15$. Next click **Simulate**. Observe that the type-1 error obtained by simulating Des5 is about 4.5%, an improvement over the corresponding type 1

error obtained by simulating Des3.

**Simulation: Discrete Endpoint: One-Sample Test - Single Arm Design - Single Proportion**

| Simulation Parameters | |
|---|---|
| Simulation ID: | Sim5 |
| Design Type: | Superiority |
| Number of Looks: | 3 |
| Test Type: | 1-Sided |
| Prop. Response under Null ($\pi_0$): | 0.15 |
| Variance: | Empirical Estimate |
| Avg. Power at Termination: | 0.044 |
| **Response Generation Parameters** | |
| Proportion Response ($\pi$): | 0.15 |
| **Simulation Control Parameters** | |
| Starting Seed: | Clock |
| Number of Simulations: | 10000 |

**Simulation Boundaries and Boundary Crossing Probabilities:**

| Look # | Sample Size (n) | Boundaries Efficacy | Early Stopping For | Total Simulations | |
|---|---|---|---|---|---|
| | | Upper | Efficacy | Count | % |
| 1 | 345 | 3.2 | 6 | 6 | 0.06 |
| 2 | 690 | 2.141 | 128 | 128 | 1.28 |
| 3 | 1035 | 1.695 | 309 | 9866 | 98.66 |
| Total | | | 443 | 10000 | |
| % | | | 4.43 | | |

**Average Sample Size:**

| Look # | Average Sample Size (n) |
|---|---|
| 1 | 345 |
| 2 | 690 |
| 3 | 1035 |
| Average | 1030.17 |

**Overall Simulation Results**

| | |
|---|---|
| Starting Seed: | 1989154 |
| Total Number of Simulations: | 10000 |
| Elapsed Time: | 00:00:01 |

Next, verify that a sample size of 1035 suffices for producing 80% power by running 10,000

simulations with $\pi = 0.18$.

**Simulation: Discrete Endpoint: One-Sample Test - Single Arm Design - Single Proportion**

| Simulation Parameters | |
|---|---|
| Simulation ID: | Sim6 |
| Design Type: | Superiority |
| Number of Looks: | 3 |
| Test Type: | 1-Sided |
| Prop. Response under Null ($\pi_0$): | 0.15 |
| Variance: | Empirical Estimate |
| Avg. Power at Termination: | 0.814 |
| **Response Generation Parameters** | |
| Proportion Response (π): | 0.18 |
| **Simulation Control Parameters** | |
| Starting Seed: | Clock |
| Number of Simulations: | 10000 |

**Simulation Boundaries and Boundary Crossing Probabilities:**

| Look # | Sample Size (n) | Boundaries Efficacy | Early Stopping For | Total Simulations | |
|---|---|---|---|---|---|
| | | Upper | Efficacy | Count | % |
| 1 | 345 | 3.2 | 242 | 242 | 2.42 |
| 2 | 690 | 2.141 | 4601 | 4601 | 46.01 |
| 3 | 1035 | 1.695 | 3300 | 5157 | 51.57 |
| Total | | | 8143 | 10000 | |
| % | | | 81.43 | | |

**Average Sample Size:**

| Look # | Average Sample Size (n) |
|---|---|
| 1 | 345 |
| 2 | 690 |
| 3 | 1035 |
| Average | 859.568 |

**Overall Simulation Results**

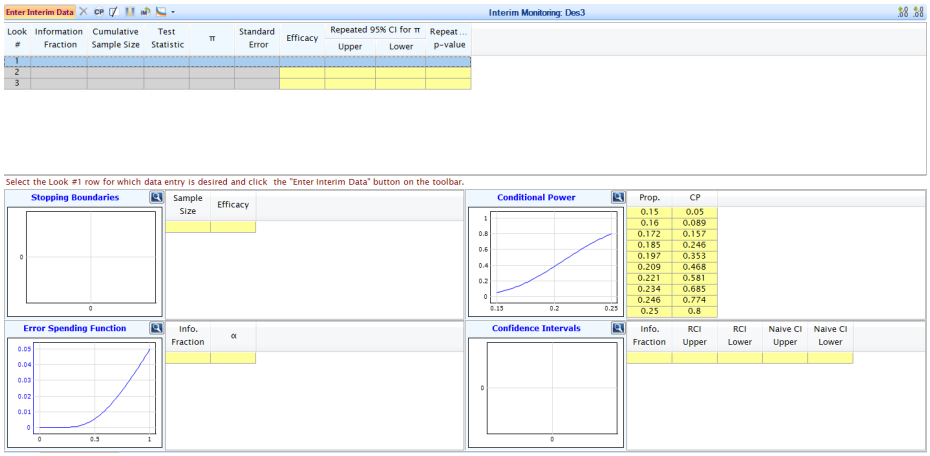
| | |
|---|---|
| Starting Seed: | 2072502 |
| Total Number of Simulations: | 10000 |
| Elapsed Time: | 00:00:02 |

This example has demonstrated the importance of simulating a design to verify that it does indeed possess the operating characteristics that are claimed for it. Since these operating characteristics were derived by large-sample theory, they might not hold for small sample sizes, in which case, the sample size or type-1 error might have to be adjusted appropriately.

### 6.1.3 Interim Monitoring

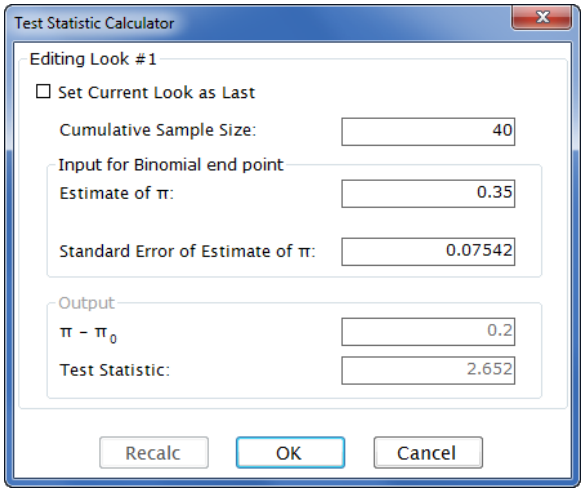
Consider interim monitoring of Des3, the design that has 80% power when the empirical estimate of variance is used to standardize the test statistic. Select Des3 in the **Library**, and click IM icon from the Library toolbar. Alternatively, right-click on Des3 and select **Create IM Dashboard**. The interim monitoring dashboard contains various controls for monitoring the trial, and is divided into two sections. The top section contains several columns for displaying output values based on the interim inputs. The bottom section contains four charts, each with a corresponding table to its right. These charts provide graphical and numerical

descriptions of the progress of the clinical trial and are useful tools for decision making by a data monitoring committee.



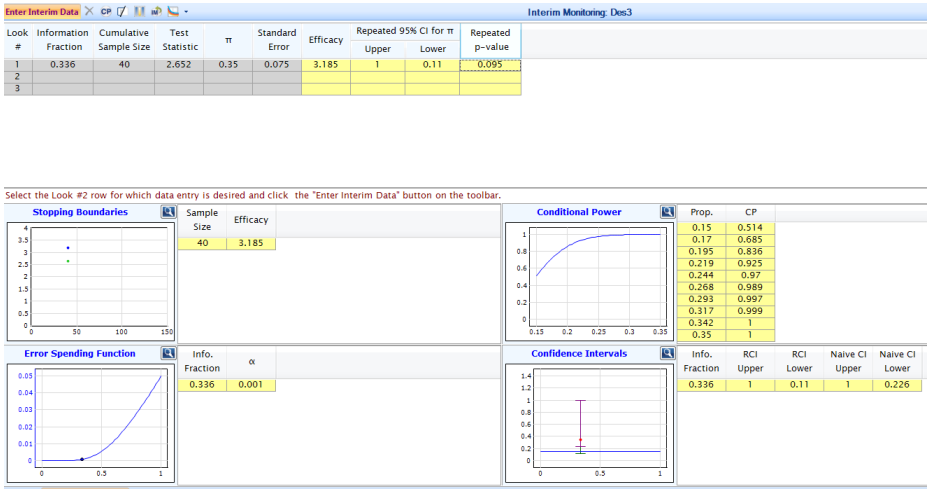

At the first interim look, when 40 subjects have enrolled, suppose that the observed response rate is 0.35. Click Enter Interim Data icon to invoke the **Test Statistic Calculator**. In the box next to **Cumulative Sample Size** enter $40$. Enter $0.35$ in the box next to **Estimate of** $\pi$. In the

box next to **Standard Error of Estimate of** $\pi$ enter $0.07542$. Next click **Recalc**.




Observe that upon pressing the **Recalc** button, the test statistic calculator automatically computes the value of the test statistic as 2.652.

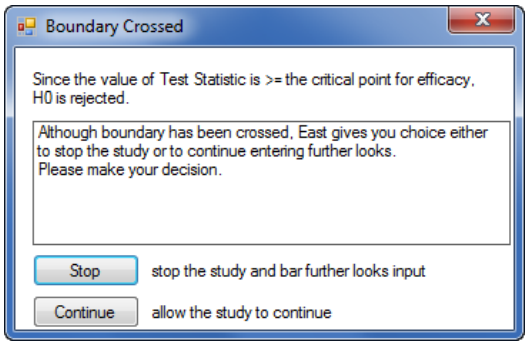
Clicking **OK** results in the following output.



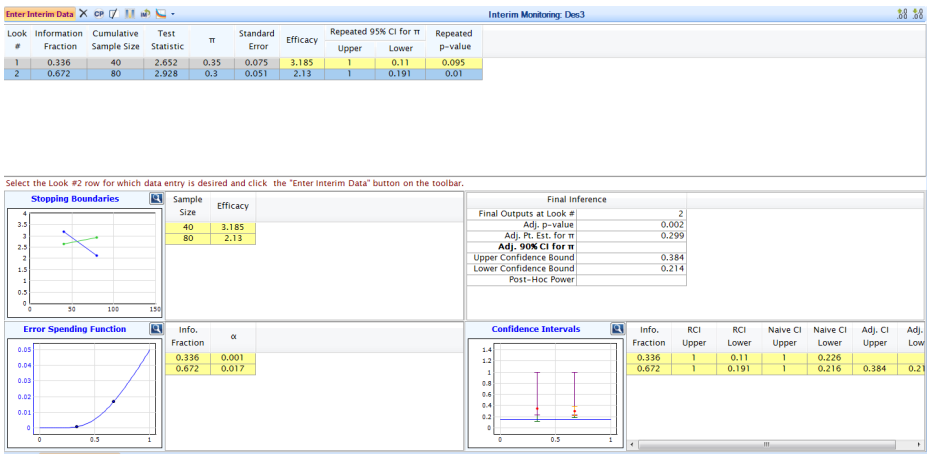

Since our test statistic, 2.652, is smaller than the stopping boundary, 3.185, the trial continues.

At the second interim monitoring time point, after 80 subjects have enrolled, suppose that the estimate of $\hat{\pi}$ based on all data up to that point is 0.30. Click on the second row in the table in the upper section. Then click Enter Interim Data icon. In the box next to **Cumulative Sample Size** enter $80$. Enter $0.30$ in the box next to **Estimate of** $\pi$. In the box next to **Standard Error of Estimate of** $\pi$ enter $0.05123$. Next click **Recalc**. Upon clicking **OK** we observe that the

stopping boundary is crossed and the following message is displayed.




We can conclude that $\pi > 0.15$ and terminate the trial. Clicking **Stop** yields the following output.




## 6.2 McNemar's Test

McNemar's Test is used in experimental situations where paired comparisons are observed. In a typical application, two binary response measurements are made on each subject – perhaps from two different treatments, or from two different time points. For example, in a

comparative clinical trial, subjects are matched on baseline demographics and disease characteristics and then randomized with one subject in the pair receiving the experimental treatment and the other subject receiving the control. Another example is the cross over clinical trial in which each subject receives both treatments. By random assignment, some subjects receive the experimental treatment followed by the control while others receive the control followed by the experimental treatment. Let $\pi_c$ and $\pi_t$ denote the response probabilities for the control and experimental treatments, respectively. The probability parameters for McNemar's test are displayed in Table 6.1.

Table 6.1: A 2 x 2 Table of Probabilities for McNemar's Test

| Control | Experimental | | Total |
|---|---|---|---|
| | No Response | Response | Probability |
| No Response | $\pi_{00}$ | $\pi_{01}$ | $1-\pi_c$ |
| Response | $\pi_{10}$ | $\pi_{11}$ | $\pi_c$ |
| Total Probability | $1-\pi_t$ | $\pi_t$ | 1 |

The null hypothesis

$$H_0\colon \pi_c = \pi_t$$

is tested against the alternative hypothesis

$$H_1\colon \pi_c \neq \pi_t$$

for the two sided testing problem or the alternative hypothesis

$$H_1\colon \pi_c > \pi_t$$

(or $H_1\colon \pi_c < \pi$) for the one-sided testing problem. Since $\pi_t = \pi_c$ if and only if $\pi_{01} = \pi_{10}$, the null hypothesis is also expressed as

$$H_0\colon \pi_{01} = \pi_{10}\ ,$$

and is tested against corresponding one and two sided alternatives. The power of this test depends on two quantities:

1. The difference between the two discordant probabilities (which is also the difference between the response rates of the two treatments)

$$\delta = \pi_{01} - \pi_{10} = \pi_t - \pi_c \ ;$$

2. The sum of the two discordant probabilities

$$\xi = \pi_{10} + \pi_{01} \ .$$

East accepts these two parameters as inputs at the design stage.

We next specify the test statistic to be used during the interim monitoring stage. Suppose we intend to execute McNemar's test a maximum of $K$ times in a group sequential setting. Let the **c**umulative data up to and including the $j$ th interim look consist of $N(j)$ matched pairs arranged in the form of the following $2 \times 2$ contingency table of counts:

Table 6.2: $2 \times 2$ Contingency Table of Counts of Matched Pairs at Look $j$

| Control | Experimental | | Total |
|---|---|---|---|
| | No Response | Response | Probability |
| No Response | $n_{00}(j)$ | $n_{01}(j)$ | $r_0(j)$ |
| Response | $n_{10}(j)$ | $n_{11}(j)$ | $r_1(j)$ |
| Total Probability | $c_0(j)$ | $c_1(j)$ | $N(j)$ |

For $a = 0, 1$ and $b = 0, 1$ define

$$\hat{\pi}_{ab}(j) = \frac{n_{ab}(j)}{N(j)} \tag{6.3}$$

Then the sequentially computed McNemar test statistic at look $j$ is

$$Z_j = \frac{\hat{\delta}_j}{\text{se}(\hat{\delta}_j)} \tag{6.4}$$

where

$$\hat{\delta}_j = \hat{\pi}_{01}(j) - \hat{\pi}_{10}(j) \tag{6.5}$$

and

$$\text{se}(\hat{\delta}_j) = \frac{\sqrt{[n_{00}(j) + n_{11}(j)][n_{01}(j) + n_{10}(j)] + 4n_{01}(j)n_{10}(j)}}{N(j)\sqrt{N(j)}}. \tag{6.6}$$

Note that the standard error (6.6) is equal to

$$\text{se}(\hat{\delta}_j) = \frac{\sqrt{\hat{\xi}_j - \hat{\delta}_j^2}}{\sqrt{N(j)}}. \tag{6.7}$$

The above statistic was defined in the non-sequential setting by Fleiss (1981, page 117). We now show how to use East to design and monitor a clinical trial based on McNemar's test.

### 6.2.1 Trial Design

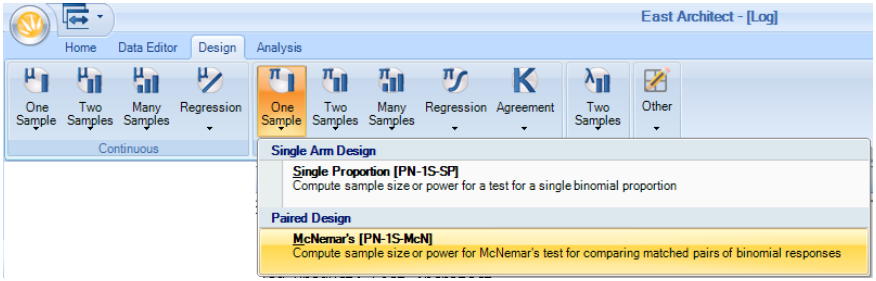
Consider a trial in which we wish to determine whether a transdermal delivery system (TDS) can be improved with a new adhesive. Subjects are to wear the old TDS (control) and new TDS (experimental) in the same area of the body for one week each. A response is said to occur if the TDS remains on for the entire one week observation period. From historical data, it is known that control has a response rate of 85% ($\pi_c = 0.85$). It is hoped that the new adhesive will increase this to 95% ($\pi_t = 0.95$). Furthermore, of the 15% of the subjects who did not respond on the control, it is hoped that 87% will respond on the experimental system. That is, $\pi_{01} = 0.87 \times 0.15 = 0.13$. Based on these data, we can fill in all the entries of Table 6.1 as displayed in Table 6.2.

Table 6.3: McNemar Probabilities for the TDS Trial

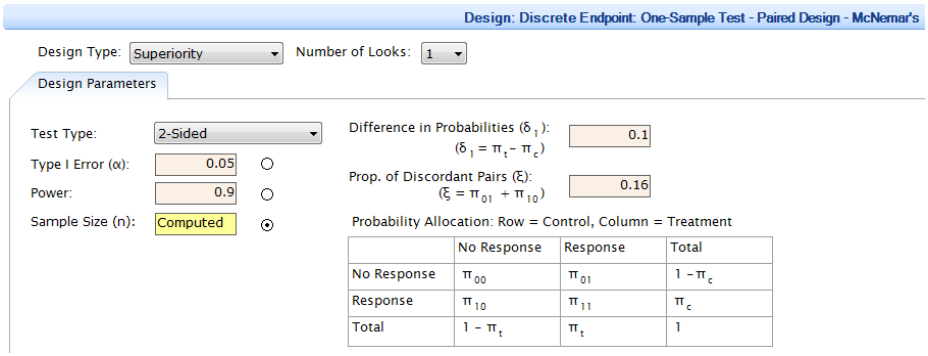
| | Experimental | | Total |
|---|---|---|---|
| Control | No Response | Response | Probability |
| No Response | 0.02 | 0.13 | 0.15 |
| Response | 0.03 | 0.82 | 0.85 |
| Total Probability | 0.05 | 0.95 | 1 |

Although it is expected that the new adhesive will increase the adherence rate, the comparison is posed as a two-sided testing problem, testing $H_0$: $\pi_c = \pi_t$ against $H_1$: $\pi_c \neq \pi_t$ at the 0.05

level. We wish to determine the sample size to have 90% power for the values displayed in Table 6.3. To design this trial, click **Design** tab, then **Single Sample** on the **Discrete** group, and then click **McNemar's Test for Matched Pairs**.




***Single-Look Design*** First, consider a study with no interim analyses, and 90% power for two sided test at $\alpha = 0.05$. Choose the design parameters as shown below. We first consider a single-look design, so leave the default value for **Number of Looks** to $1$. Enter $0.9$ for **Power**. As shown in Table 6.2, we must specify $\delta_1 = \pi_t - \pi_c = 0.1$ and $\xi = \pi_{01} + \pi_{10} = 0.16$.

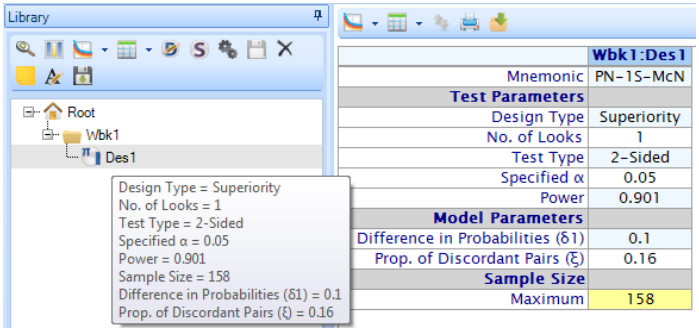



Click **Compute**. The design Des1 is shown as a row in the Output Preview located in the lower

pane of this window. A total of 158 subjects is required to have 90% power.

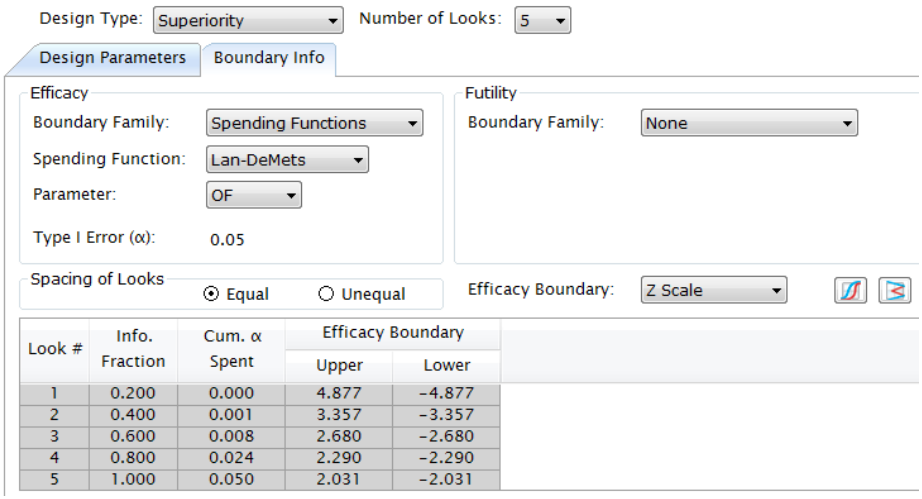
| | ID | Design Type | No. of Looks | Test Type | Specified α | Power | Sample Size | δ1 | ξ |
|---|---|---|---|---|---|---|---|---|---|
| | Des1 | Superiority | 1 | 2-Sided | 0.05 | 0.901 | 158 | 0.1 | 0.16 |

You can select this design by clicking anywhere on the row in the **Output Preview**. Click on icon to get the output summary displayed in the upper pane. In the **Output Preview** toolbar, click the icon to save this design Des1 to workbook Wbk1 in the **Library**. If you hover the cursor over Des1 in the Library, a tooltip will appear that summarizes the input parameters of the design.




***Five-Look Design*** Now consider the same design with a maximum of 5 looks, using the default Lan-DeMets (O'Brien-Fleming) spending function. Create a new design by selecting Des1 in the **Library**, and clicking icon on the **Library** toolbar. Change the **Number of Looks** from 1 to 5, to generate a study with four interim looks and a final analysis. A new tab **Boundary Info** will appear. Clicking on this tab will reveal the stopping boundary parameters. By default, the **Spacing of Looks** is set to **Equal**, which means that the interim analyses will be equally spaced in terms of the number of patients accrued between looks. The left side contains details for the **Efficacy** boundary, and the right side for the **Futility** boundary. By default, there is an efficacy boundary (to reject H0) selected, but no futility boundary (to reject

H1). The **Boundary Family** specified is of the `Spending Functions` type. The default **Spending function** is the Lan-DeMets (Lan & DeMets, 1983), with **Parameter** `OF` (O'Brien-Fleming), which generates boundaries that are very similar, though not identical, to the classical stopping boundaries of O'Brien and Fleming (1979). Technical details of these stopping boundaries are available in Appendix **??**.




| Look # | Info. Fraction | Cum. α Spent | Efficacy Boundary | |
|---|---|---|---|---|
| | | | Upper | Lower |
| 1 | 0.200 | 0.000 | 4.877 | -4.877 |
| 2 | 0.400 | 0.001 | 3.357 | -3.357 |
| 3 | 0.600 | 0.008 | 2.680 | -2.680 |
| 4 | 0.800 | 0.024 | 2.290 | -2.290 |
| 5 | 1.000 | 0.050 | 2.031 | -2.031 |

Click **Compute** to generate output for Des2. With Des2 selected in the **Output Preview**, click the  icon to save Des2 to the **Library**. In the **Library**, select the nodes for both Des1 and Des2, by holding the Ctrl key, and then click the  icon. The upper pane will display the

output summary of the two designs side-by-side:

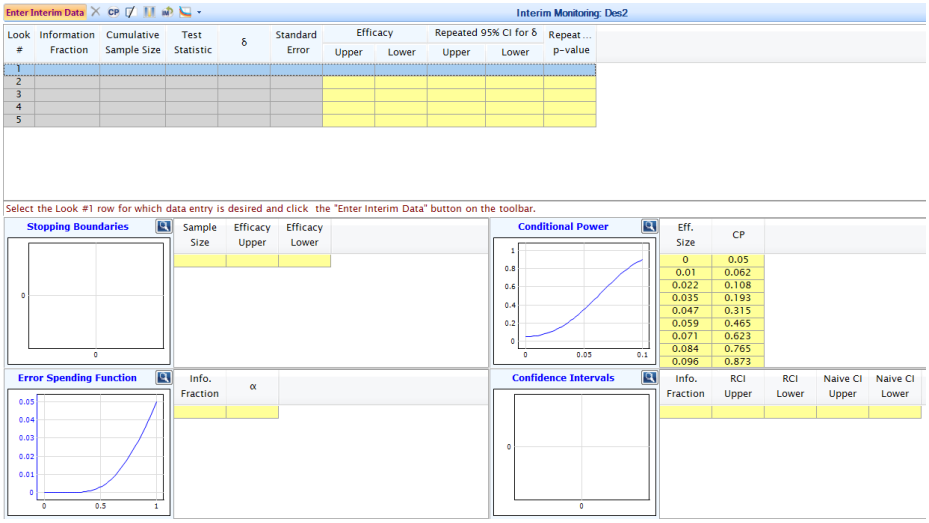
| | Wbk1:Des1 | Wbk1:Des2 |
|---|---|---|
| Mnemonic | PN-1S-McN | PN-1S-McN |
| **Test Parameters** | | |
| Design Type | Superiority | Superiority |
| No. of Looks | 1 | 5 |
| Test Type | 2-Sided | 2-Sided |
| Specified α | 0.05 | 0.05 |
| Power | 0.901 | 0.901 |
| **Model Parameters** | | |
| Difference in Probabilities (δ1) | 0.1 | 0.1 |
| Prop. of Discordant Pairs (ξ) | 0.16 | 0.16 |
| **Boundary Parameters** | | |
| Spacing of Looks | | Equal |
| Efficacy Boundary | | LD (OF) |
| **Sample Size** | | |
| Maximum | 158 | 162 |
| Expected Under H0 | | 160.935 |
| Expected Under H1 | | 119.965 |

There has been a slight inflation in the maximum sample size, from 158 to 162. However, the expected sample size is 120 subjects if the alternative hypothesis of $\delta_1 = 0.10$ and $\xi = 0.16$ holds. The stopping boundary, spending function, and Power vs. Sample Size charts can all be displayed by clicking on the appropriate icons from the **Library** toolbar.

### 6.2.2 Interim Monitoring

Consider interim monitoring of Des2. Select Des2 in the **Library**, and click IM icon from the Library toolbar. Alternatively, right-click on Des2 and select **Create IM Dashboard**. A new IM

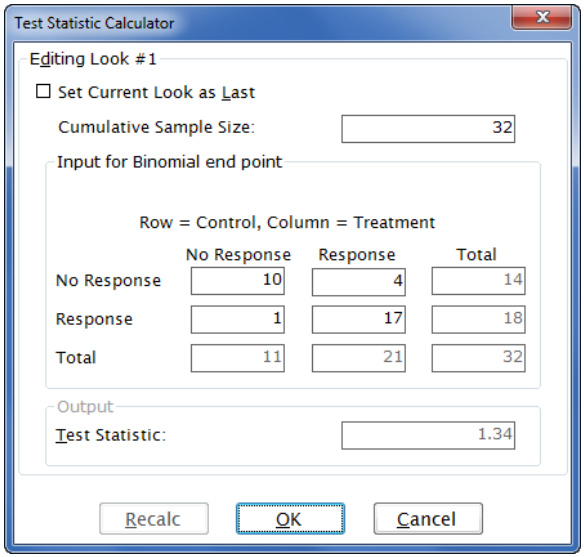
worksheet will appear.



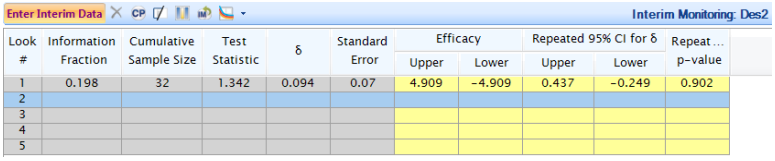

Suppose, that the results are to be analyzed after results are available for every 32 subjects. After the first 32 subjects were enrolled, one subject responded on the control arm and did not respond on the treatment arm; four subjects responded on the treatment arm but did not respond on the control arm; 10 subjects did not respond on either treatment; 17 subjects responded on both the arms. This information is sufficient to complete all the entries in Table 6.3 and hence to evaluate the test statistic value.

Click Enter Interim Data icon to invoke the **Test Statistic Calculator**. In the box next to **Cumulative Sample Size** enter 32. Enter the values in the table as shown below and click

**Recalc**.



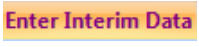

Clicking **OK** results in the following entry in the first look row.



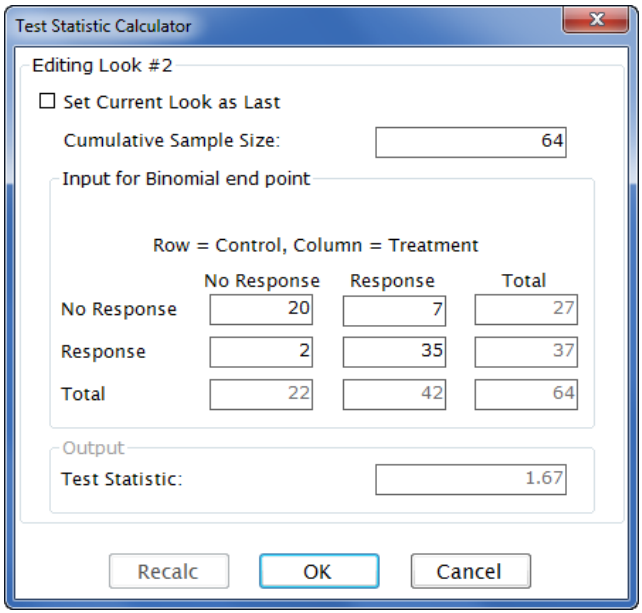

| Look # | Information Fraction | Cumulative Sample Size | Test Statistic | δ | Standard Error | Efficacy | | Repeated 95% CI for δ | | Repeat ... |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | Upper | Lower | Upper | Lower | p-value |
| 1 | 0.198 | 32 | 1.342 | 0.094 | 0.07 | 4.909 | -4.909 | 0.437 | -0.249 | 0.902 |
| 2 | | | | | | | | | | |
| 3 | | | | | | | | | | |
| 4 | | | | | | | | | | |
| 5 | | | | | | | | | | |

As you can see the value of the test statistic, 1.342, is within the stopping boundaries, (4.909,-4.909). Thus, the trial continues.

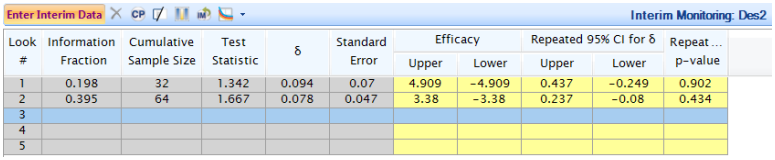
The second interim analysis was performed after data were available for 64 subjects. A total of two subjects responded on the control arm and failed to respond on the treatment arm; seven subjects responded on the treatment arm and failed to respond on the control arm; 20 subjects responded on neither arm; 35 subjects responded on both the arms.

Click on the second row in the table in the upper section. Then click  icon.

Enter the appropriate values in the table as shown below and click **Recalc**.




Then click **OK**. This results in the following screen.

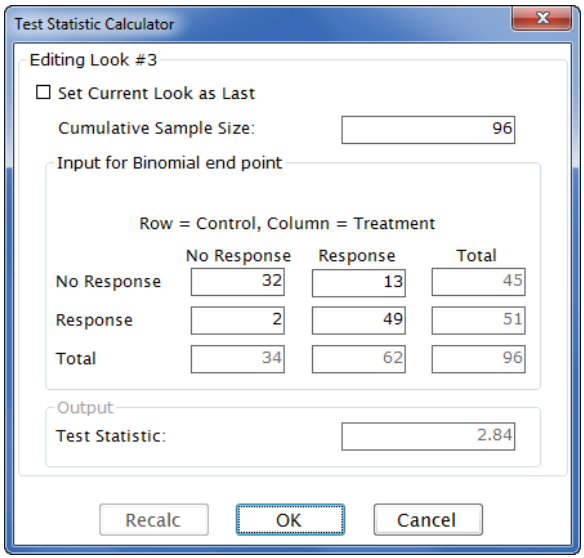



| Look # | Information Fraction | Cumulative Sample Size | Test Statistic | δ | Standard Error | Efficacy | | Repeated 95% CI for δ | | Repeat ... |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | Upper | Lower | Upper | Lower | p-value |
| 1 | 0.198 | 32 | 1.342 | 0.094 | 0.07 | 4.909 | -4.909 | 0.437 | -0.249 | 0.902 |
| 2 | 0.395 | 64 | 1.667 | 0.078 | 0.047 | 3.38 | -3.38 | 0.237 | -0.08 | 0.434 |
| 3 | | | | | | | | | | |
| 4 | | | | | | | | | | |
| 5 | | | | | | | | | | |

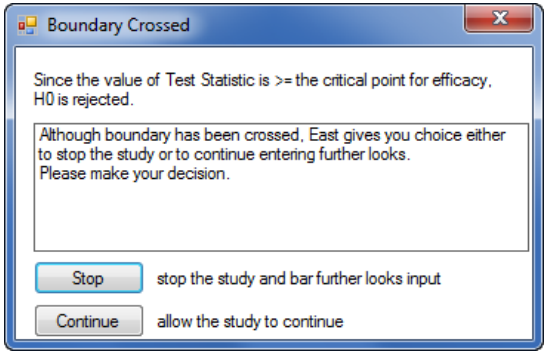
At the third interim analysis, after 96 subjects were enrolled, a total of two subjects responded on the control arm and failed to respond on the treatment arm; 13 subjects responded on the treatment arm and failed to respond on the control arm; 32 subjects did not respond on either arm; 49 subjects responded on both the arms.

Click on the third row in the table in the upper section. Then click Enter Interim Data icon.

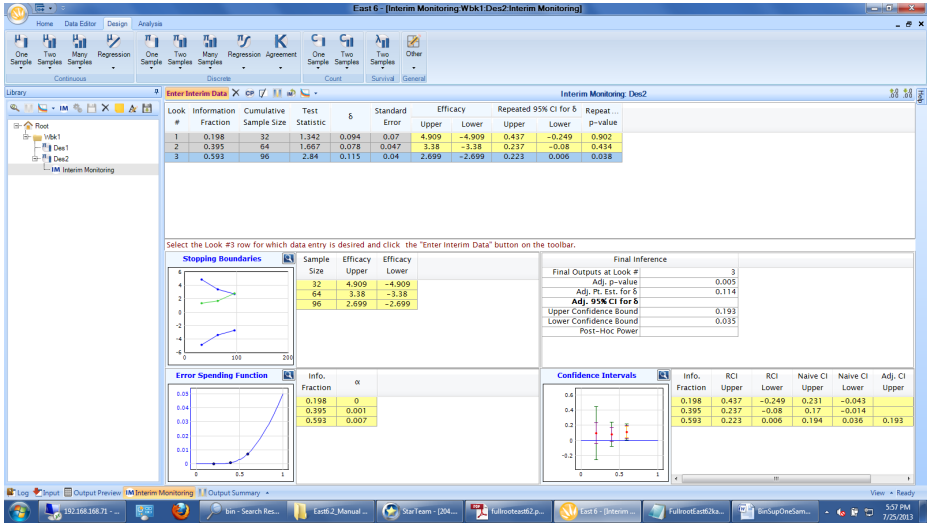
Enter the appropriate values in the table as shown below and click **Recalc**.




Then click **OK**. This results in the following message box.



Clicking on **Stop** yields the following Interim Monitoring output.



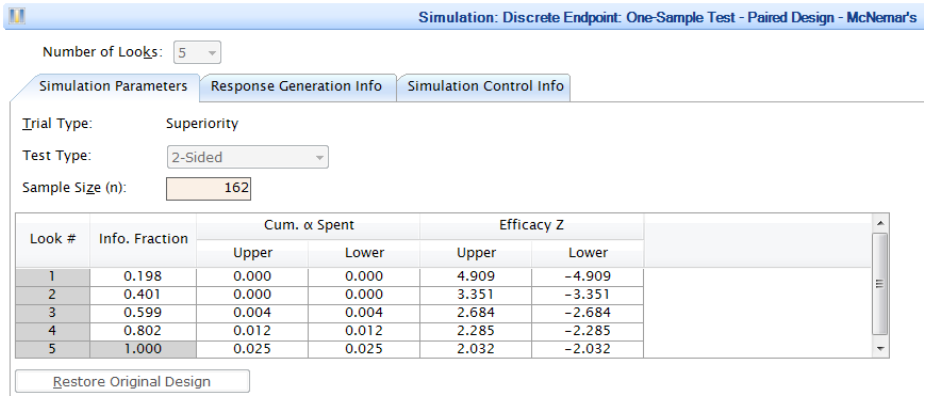

We reject the null hypothesis that $\delta = 0$, based on these data.

## 6.2.3 Simulation

Des2 can be simulated to examine the properties for different values of the parameters. First, we verify the results under the alternative hypothesis at which the power is to be controlled, namely $\delta_1$=0.10 and $\xi$=0.16.

Select Des2 in the **Library**, and click  icon from **Library** toolbar. Alternatively, right-click

on Des2 and select **Simulate**. A new Simulation worksheet will appear.




Click **Simulate** to start the simulation. Once the simulation run has completed, East will add an additional row to the **Output Preview** labeled Sim1. Select Sim1 in the **Output Preview**. If you click  icon, you will see some of the simulation output details displayed in the upper pane. Click  icon to save it to the **Library**. Double-click on Sim1 in the **Library**. The simulation output details will be displayed as shown below. The results confirm that the power

is at about 90%.

**Simulation: Discrete Endpoint: One-Sample Test - Paired Design - McNemar's**

| Simulation Parameters | |
|---|---|
| Simulation ID: | Sim1 |
| Design Type: | Superiority |
| Number of Looks: | 5 |
| Test Type: | 2-Sided |
| Avg. Power at Termination: | 0.904 |
| **Response Generation Parameters** | |
| Difference in Probabilities ($\delta_1$): | 0.1 |
| Prop. of Discordant Pairs ($\xi$): | 0.16 |
| **Simulation Control Parameters** | |
| Starting Seed: | Fixed |
| Number of Simulations: | 10000 |

**Simulation Boundaries and Boundary Crossing Probabilities:**

| Look # | Sample Size (n) | Boundaries Efficacy | | Early Stopping For | | Total Simulations | |
|---|---|---|---|---|---|---|---|
| | | Upper | Lower | Upper Efficacy | Lower Efficacy | Count | % |
| 1 | 32 | 4.909 | -4.909 | 0 | 0 | 0 | 0 |
| 2 | 65 | 3.351 | -3.351 | 175 | 0 | 175 | 1.75 |
| 3 | 97 | 2.684 | -2.684 | 3593 | 0 | 3593 | 35.93 |
| 4 | 130 | 2.285 | -2.285 | 3607 | 0 | 3607 | 36.07 |
| 5 | 162 | 2.032 | -2.032 | 1660 | 0 | 2625 | 26.25 |
| Total | | | | 9035 | 0 | 10000 | |
| % | | | | 90.35 | 0 | | |

**Average Sample Size:**

| Look # | Average Sample Size (n) |
|---|---|
| 1 | 32 |
| 2 | 65 |
| 3 | 97 |
| 4 | 130 |
| 5 | 162 |
| Average | 125.406 |

**Overall Simulation Results**

| | |
|---|---|
| Starting Seed: | 85535015 |
| Total Number of Simulations: | 10000 |
| Elapsed Time: | 00:00:01 |

To confirm the results under the null hypothesis, set $\delta_1 = 0$ in the **Response Generation Info** tab in the simulation worksheet and then click textbfSimulate. The results, which confirm that

the type-1 error rate is approximately 5%, are given below.

**Simulation: Discrete Endpoint: One-Sample Test - Paired Design - McNemar's**

| Simulation Parameters | |
|---|---|
| Simulation ID: | Sim2 |
| Design Type: | Superiority |
| Number of Looks: | 5 |
| Test Type: | 2-Sided |
| Avg. Power at Termination: | 0.044 |
| **Response Generation Parameters** | |
| Difference in Probabilities ($\delta_1$): | 0 |
| Prop. of Discordant Pairs ($\xi$): | 0.16 |
| **Simulation Control Parameters** | |
| Starting Seed: | Fixed |
| Number of Simulations: | 10000 |

**Simulation Boundaries and Boundary Crossing Probabilities:**

| Look # | Sample Size (n) | Boundaries: Efficacy Upper | Boundaries: Efficacy Lower | Early Stopping For: Upper Efficacy | Early Stopping For: Lower Efficacy | Total Simulations: Count | Total Simulations: % |
|---|---|---|---|---|---|---|---|
| 1 | 32 | 4.909 | -4.909 | 0 | 0 | 0 | 0 |
| 2 | 65 | 3.351 | -3.351 | 1 | 0 | 1 | 0.01 |
| 3 | 97 | 2.684 | -2.684 | 19 | 21 | 40 | 0.4 |
| 4 | 130 | 2.285 | -2.285 | 80 | 90 | 170 | 1.7 |
| 5 | 162 | 2.032 | -2.032 | 117 | 107 | 9789 | 97.89 |
| Total | | | | 217 | 218 | 10000 | |
| % | | | | 2.17 | 2.18 | | |

**Average Sample Size:**

| Look # | Average Sample Size (n) |
|---|---|
| 1 | 32 |
| 2 | 65 |
| 3 | 97 |
| 4 | 130 |
| 5 | 162 |
| Average | 161.186 |

**Overall Simulation Results**

| | |
|---|---|
| Starting Seed: | 85773537 |
| Total Number of Simulations: | 10000 |
| Elapsed Time: | 00:00:01 |

While it is often difficult to specify the absolute difference of the discordant probabilities, $\delta_1$, it is even more difficult to specify the sum of the discordant probabilities, $\xi$. Simulation can be used to examine the effects of misspecification of $\xi$. Run the simulations again, now with

$\delta_1$=0.10 and $\xi$=0.2. The results are given below.

**Simulation: Discrete Endpoint: One-Sample Test - Paired Design - McNemar's**

| Simulation Parameters | |
|---|---|
| Simulation ID: | Sim3 |
| Design Type: | Superiority |
| Number of Looks: | 5 |
| Test Type: | 2-Sided |
| Avg. Power at Termination: | 0.809 |
| **Response Generation Parameters** | |
| Difference in Probabilities ($\delta_1$): | 0.1 |
| Prop. of Discordant Pairs ($\xi$): | 0.2 |
| **Simulation Control Parameters** | |
| Starting Seed: | Fixed |
| Number of Simulations: | 10000 |

**Simulation Boundaries and Boundary Crossing Probabilities:**

| Look # | Sample Size (n) | Boundaries Efficacy Upper | Boundaries Efficacy Lower | Early Stopping For Upper Efficacy | Early Stopping For Lower Efficacy | Total Simulations Count | Total Simulations % |
|---|---|---|---|---|---|---|---|
| 1 | 32 | 4.909 | -4.909 | 0 | 0 | 0 | 0 |
| 2 | 65 | 3.351 | -3.351 | 203 | 0 | 203 | 2.03 |
| 3 | 97 | 2.684 | -2.684 | 2428 | 0 | 2428 | 24.28 |
| 4 | 130 | 2.285 | -2.285 | 3485 | 0 | 3485 | 34.85 |
| 5 | 162 | 2.032 | -2.032 | 1978 | 0 | 3884 | 38.84 |
| Total | | | | 8094 | 0 | 10000 | |
| % | | | | 80.94 | 0 | | |

**Average Sample Size:**

| Look # | Average Sample Size (n) |
|---|---|
| 1 | 32 |
| 2 | 65 |
| 3 | 97 |
| 4 | 130 |
| 5 | 162 |
| Average | 133.097 |

**Overall Simulation Results**

| | |
|---|---|
| Starting Seed: | 86003123 |
| Total Number of Simulations: | 10000 |
| Elapsed Time: | 00:00:01 |

Notice that this provides a power of approximately 81%. Larger values of $\xi$ would further decrease the power. However, values of $\xi$ >0.2 with $\delta_1$=0.1 would be inconsistent with the initial assumption of $\pi_c$ = 0.85 and $\pi_t$=0.95. Additional simulations for various values of $\delta$ and $\xi$ can provide information regarding the consequences of misspecification of the input parameters.

# 7 Dose Escalation

This chapter deals with the design, simulation, and interim monitoring of Phase 1 oncology trials. A brief overview of the designs is given below; more technical details are available in the Appendix.

One of the primary goals of Phase I trials in oncology is to find the maximum tolerated dose (MTD). Currently, the vast majority of such trials have employed traditional dose escalation methods such as the 3+3 design. The 3+3 design starts by allocating three patients typically to the lowest dose level, and then adaptively moves up and down in subsequent cohorts until either the MTD is obtained, or the trial is stopped for excessive toxicity. In addition to the 3+3, East also provides the Continual Reassessment Method (CRM), the modified Toxicity Probability Interval (mTPI) method, and the Bayesian logistic regression model (BLRM). Compared to the 3+3, these modern methods may offer a number of advantages, which can be explored systematically via simulation and interim monitoring.

The CRM (Goodman et al., 1995; O'Quigley et al., 1990) is a Bayesian model-based method that uses all available information from all doses to guide dose assignment. One first specifies a target toxicity, a one-parameter dose response curve and corresponding prior distribution. The posterior mean, and predictions for the probability of toxicity at each dose, is updated as the trial progresses. The next recommended dose is the one whose toxicity probability is closest to the target toxicity.

The mTPI method (Ji et al., 2010) is Bayesian like the CRM, but rule-based like the 3+3. In this way, the mTPI represents a useful compromise between the other methods. An independent beta distribution is assumed for the probability of toxicity at each dose. A set of decision intervals are specified, and subsequent dosing decisions (up, down, or stay) are determined by computing the normalized posterior probability in each interval at the current dose. The normalized probability for each interval is known as the unit probability mass (UPM).

# Chapter 7: *Dose Escalation*

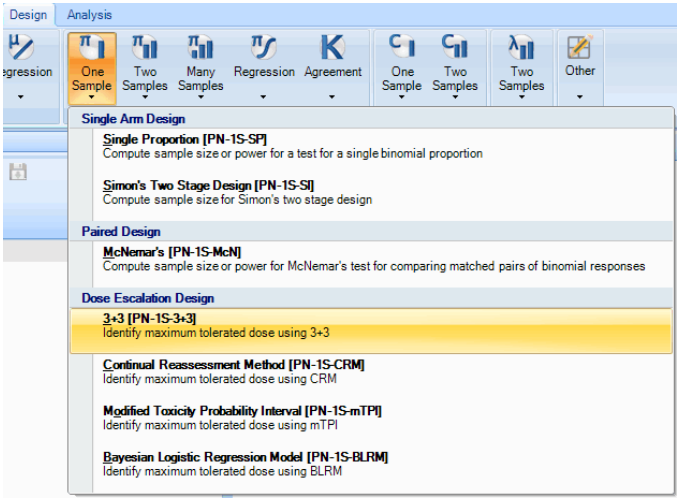
A more advanced version of the CRM is the BLRM (Neuenschwander et al., 2008), which assumes a two-parameter logistic dose response curve. In addition to a target toxicity, one specifies a set of decision intervals and associated losses for guiding dosing decisions. As data accumulate, the posterior expected loss (or *Bayes risk*), at each dose is calculated, and the next recommended dose is the one with the lowest expected loss.

## 7.1 3+3



### *7.1.1 Simulation*

Click **Discrete: One Sample** on the Design tab, and then click **Dose Escalation Design: 3+3**.



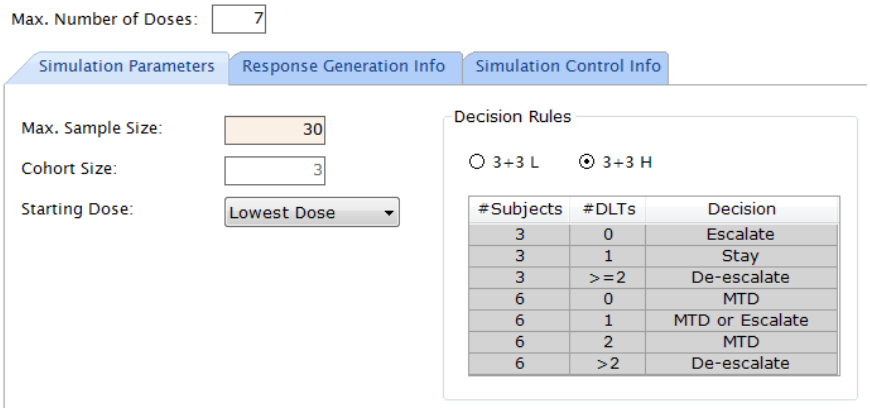

In the upper pane of this window is the Input dialog box, which is separated into three tabs: **Simulation Parameters**, **Response Generation Info**, and **Simulation Control Info**. First, you

may specify the **Max. Number of Doses** as 7.

In the **Simulation Parameters** tab, enter 30 as the **Max. Sample Size**. For the 3+3 design, the **Cohort Size** is fixed at 3. For the **Starting Dose**, select the `Lowest Dose`.

There are two flavors of 3+3 offered: L and H. The key difference between the 3+3 H method and 3+3 L method is: If we have observed 2 DLTs out of 6 patients at the current dose, the 3+3 H method will declare the current dose as MTD, while the 3+3 L method will recommend de-escalation.

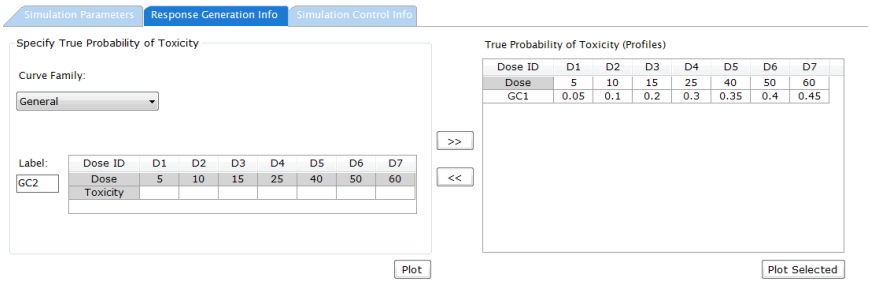
Select 3+3 H. The **Decision Rules** table gives a compact summary of the algorithm implemented here.




| #Subjects | #DLTs | Decision |
|---|---|---|
| 3 | 0 | Escalate |
| 3 | 1 | Stay |
| 3 | >=2 | De-escalate |
| 6 | 0 | MTD |
| 6 | 1 | MTD or Escalate |
| 6 | 2 | MTD |
| 6 | >2 | De-escalate |

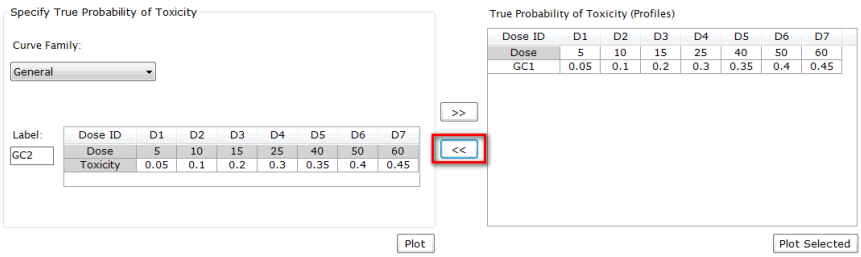
In the **Response Generation Info** tab, you can specify a set of true dose response curves from

which to simulate.



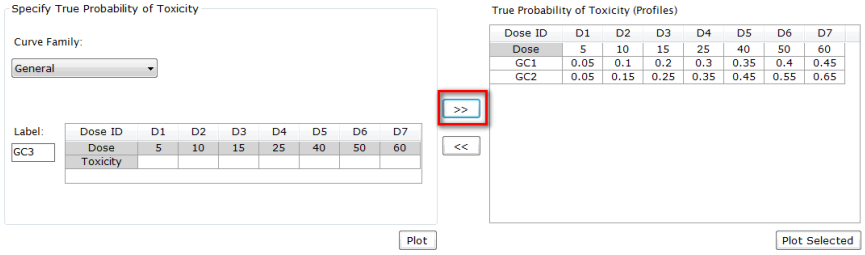

The grid on the right displays the set of dose response profiles from which East will simulate. In the row titled *Dose*, you can specify the dose levels (e.g., in mg). In the row titled *GC1*, you can edit the true probabilities of toxicity at each dose. You can also rename the profile by directly editing that cell. For now, leave all entries at their default values.

There are two ways to add profiles. The first way involves copying an existing profile on the right grid to the left grid. Select the row for GC1, and click the leftward pointing arrow to paste the GC1 profile onto the left grid.

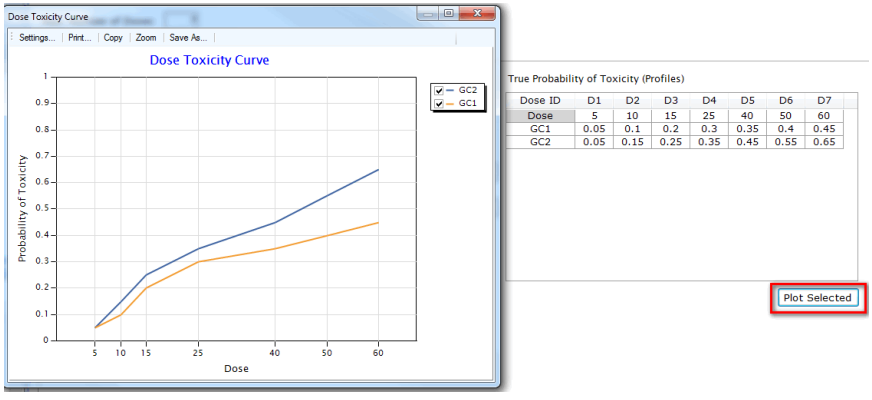



Edit the profile GC2 with the following probabilities (0.05, 0.15, 0.25, 0.35, 0.45, 0.55, 0.65), and

click the rightward pointing arrow to add this profile to the right grid.

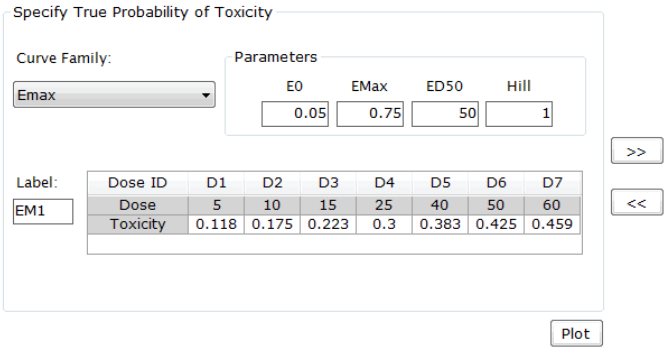



Select both rows GC1 and GC2 in the right grid, and click **Plot Selected**. The dose toxicity curves will be plotted on the same chart.



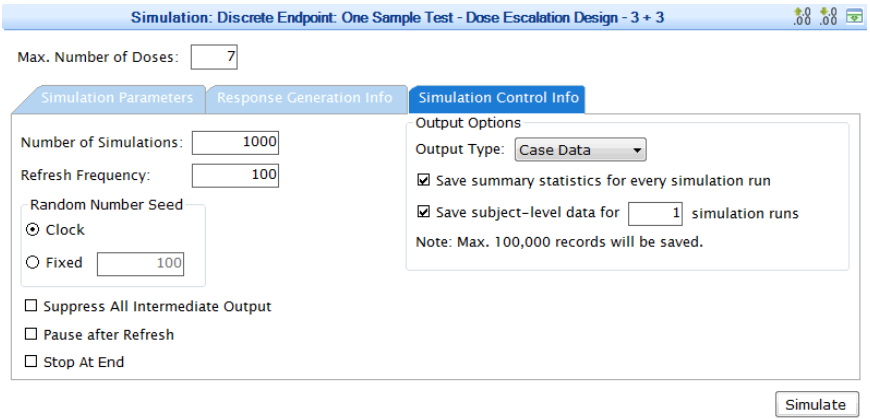

The second way to add a new profile is to generate from a parametric curve family. For example, click on the menu **Curve Family** and select **Emax**. You may construct a

four-parameter Emax curve by adjusting its parameters as below.




You can click **Plot** to generate the dose toxicity curve for this single profile in the left grid. For now, let us ignore the Emax curve, and continue with the two general curves.

In the **Simulation Control Info** tab, check the boxes corresponding to **Save summary statistics** and **Save subject-level data**. These options will provides access to several charts derived from these more detailed levels of simulated data. If you wish to display subject-level plots for more than one simulation, you can increase the number. For now, leave this at 1 to save computation time.



Click **Simulate**. East will simulate data generated from the two profiles you specified, and apply the 3+3 design to each simulation data set. Once completed, the two simulations will appear as two rows in the **Output Preview** pane below.

Output Preview

| | ID | Max. Sample Size | No. of Doses | Cohort Size | Starting Dose | Curve Family | Median Sample Size | Mean Sample Size | Median No. of DLTs | Mean No. of DLTs | Median Prop. of DLTs | Mean Prop. of DLTs | Median MTD | Mean MTD |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Sim1 | 30 | 7 | 3 | Lowest Dose | General | 18 | 17.124 | 3 | 3.177 | 0.185 | 0.192 | 15 | 21.325 |
| | Sim2 | 30 | 7 | 3 | Lowest Dose | General | 15 | 15.402 | 3 | 3.204 | 0.208 | 0.215 | 15 | 16.795 |

Select both rows in the **Output Preview** and click the  icon in the toolbar. The two simulations will be displayed side by side in the **Output Summary**.

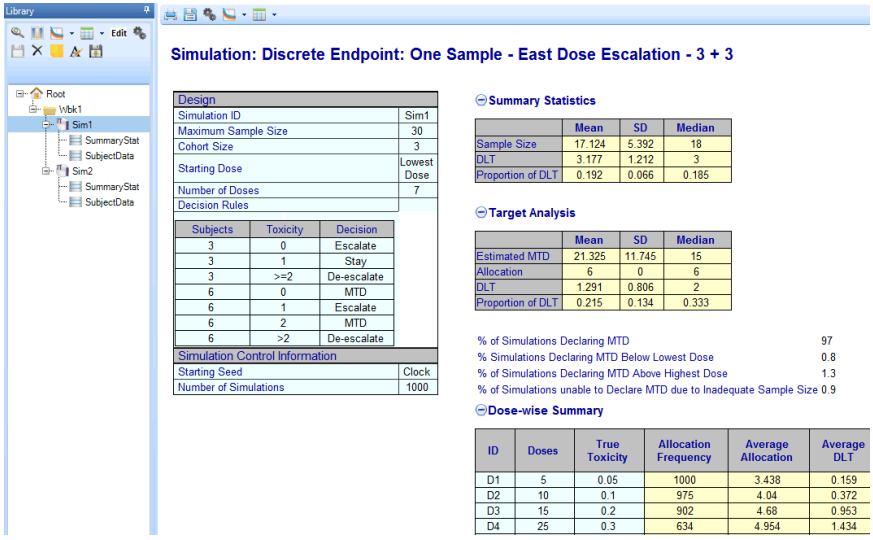
Output Summary

| | Sim1 | Sim2 |
|---|---|---|
| Mnemonic | PN-1S-3+3 | PN-1S-3+3 |
| **Design** | | |
| Max. Sample Size | 30 | 30 |
| Cohort Size | 3 | 3 |
| Starting Dose | Lowest Dose | Lowest Dose |
| **Response Generation** | | |
| Curve Family | General | General |
| **Summary Statistics** | | |
| Median Sample Size | 18 | 15 |
| Mean Sample Size | 17.124 | 15.402 |
| Median No. of DLTs | 3 | 3 |
| Mean No. of DLTs | 3.177 | 3.204 |
| Median Prop. of DLTs | 0.185 | 0.208 |
| Mean Prop. of DLTs | 0.192 | 0.215 |
| **MTD Analysis** | | |
| Median MTD | 15 | 15 |
| Mean MTD | 21.325 | 16.795 |
| Mode of MTD | 15 | 15 |
| Mode (%) | 31.6 | 34.8 |
| % of Overdosing | 0.8 | 0.9 |
| % of Underdosing | 1.3 | 0 |
| **Other Parameters** | | |
| No. of Doses | 7 | 7 |

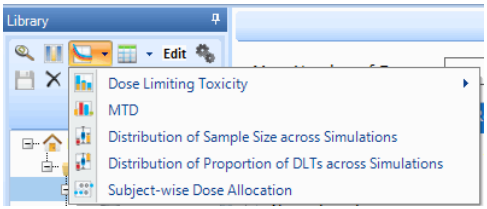
In the **Output Preview** toolbar, click the  icon to save both simulations to the **Library**.

# Chapter 7: *Dose Escalation*

Double-click Sim1 in the **Library** to display the simulation output details.

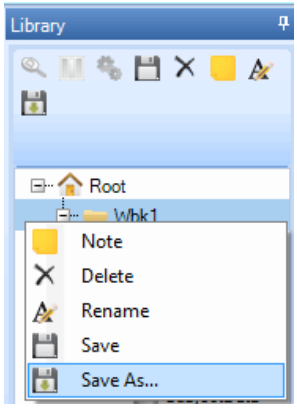



With Sim1 selected in the **Library**, click the Plots icon to access a wide range of available plots.



Below is an example of the MTD plot:


Close each plot after viewing, or save them by clicking **Save in Workbook**. To save your simulations and charts to disk, right-click Wbk1 in the **Library** and then **Save As...**.

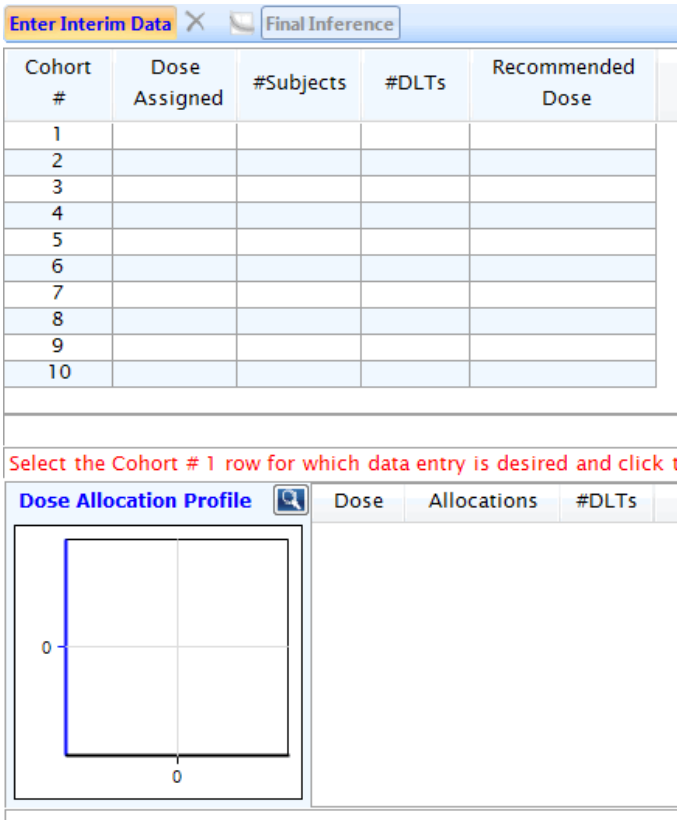



Once you have saved the workbook, you may like to clean up your library by selecting Wbk1 in the **Library** and clicking the Delete icon. The same action can be performed for Sim1 and Sim2

in the **Output Preview**.

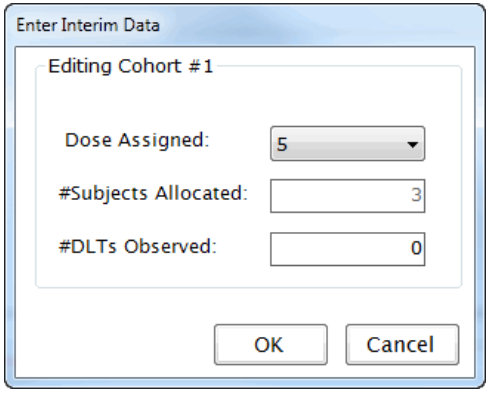
### *7.1.2 Interim Monitoring*

Right-click one of the Simulation nodes with 3+3 in the Library, and select **Interim Monitoring**. This will open an empty interim monitoring dashboard.




Click **Enter Interim Data** to open a window in which to enter data for the first cohort: in

particular, the **Dose Assigned** and the **DLTs Observed**. Click **OK** to continue.



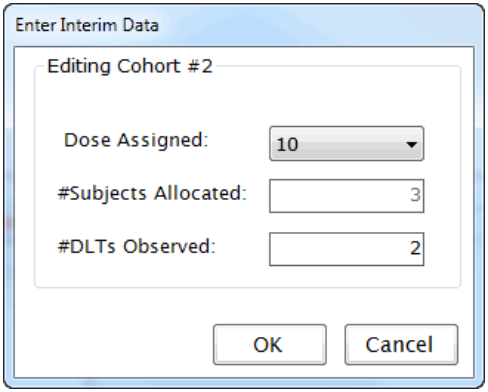

The dashboard will be updated accordingly, and the next **Recommended Dose** is 10.

Enter Interim Data | Final Inference

| Cohort # | Dose Assigned | #Subjects | #DLTs | Recommended Dose |
|---|---|---|---|---|
| 1 | 5 | 3 | 0 | 10 |
| 2 | | | | |
| 3 | | | | |
| 4 | | | | |
| 5 | | | | |
| 6 | | | | |
| 7 | | | | |
| 8 | | | | |
| 9 | | | | |
| 10 | | | | |

Click **Enter Interim Data** again, with 10 selected as **Dose Assigned**, enter 2 for **DLTs**

**Observed**, and click **OK**.




.

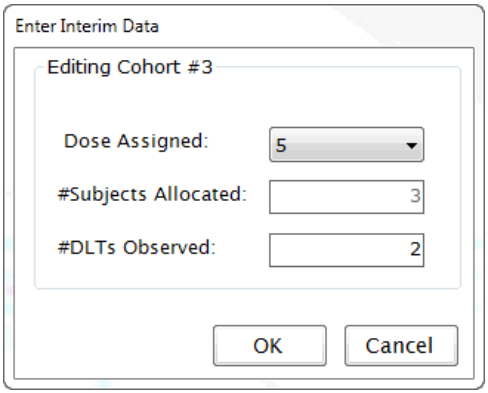
East now recommends de-escalation to 5.

Enter Interim Data | Final Inference

| Cohort # | Dose Assigned | #Subjects | #DLTs | Recommended Dose |
|---|---|---|---|---|
| 1 | 5 | 3 | 0 | 10 |
| 2 | 10 | 3 | 2 | 5 |
| 3 | | | | |
| 4 | | | | |
| 5 | | | | |
| 6 | | | | |
| 7 | | | | |
| 8 | | | | |
| 9 | | | | |
| 10 | | | | |

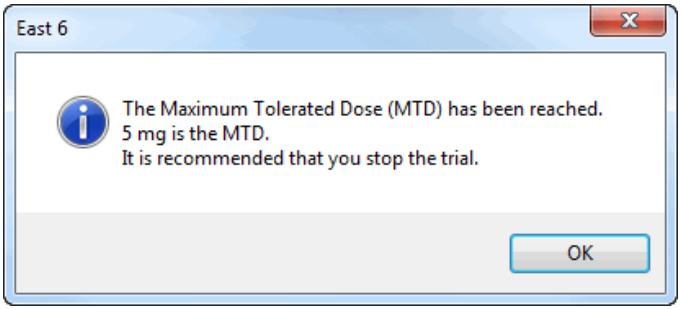
Click **Enter Interim Data**, with 5 selected as **Dose Assigned**, enter 2 for **DLTs Observed**, and

click **OK**.



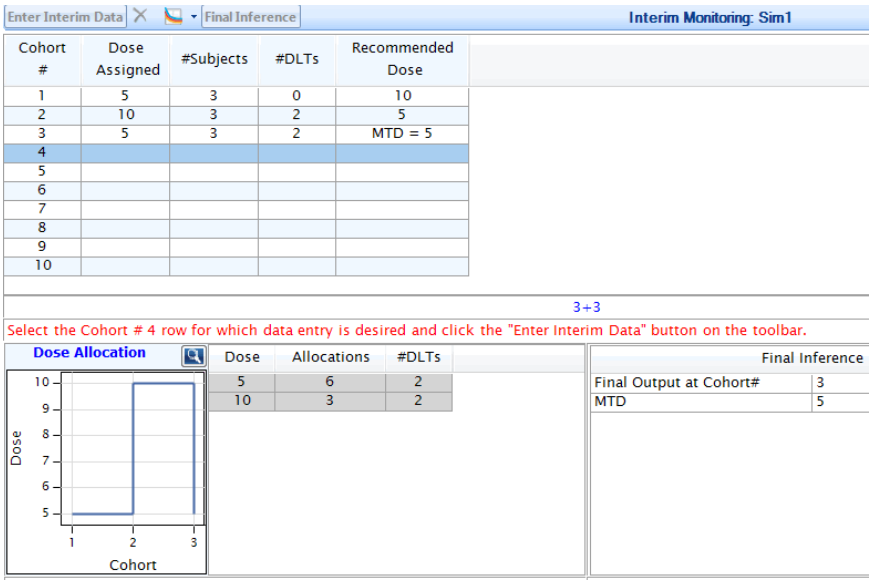

.

East recommends that you stop the trial.




.

Click **Final Inference** to generate a table for final inference.




.

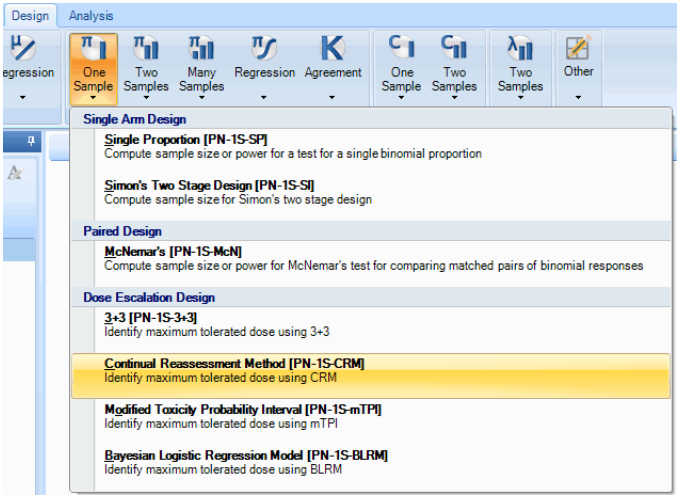
## 7.2 Continual Reassessment Method (CRM)

▪ 7.2.1 Simulation ▪ 7.2.2 Interim Monitoring

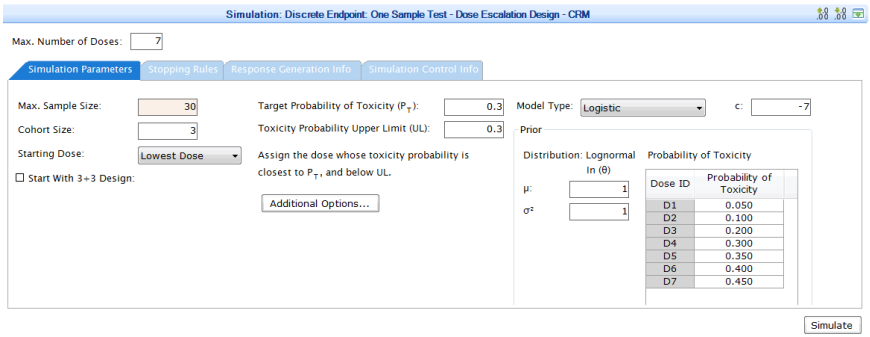
### *7.2.1 Simulation*

Click **Discrete: One Sample** on the Design tab, and then click **Dose Escalation Design:**

**Continual Reassessment Method**.



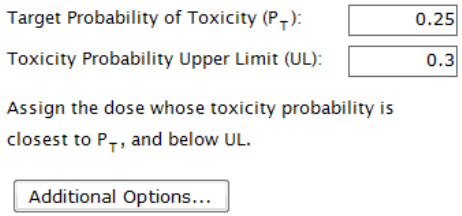

In the upper pane of this window is the Input dialog box, which is separated into four tabs: **Simulation Parameters**, **Stopping Rules**, **Response Generation Info**, and **Simulation Control Info**.



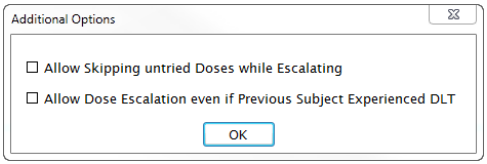

In the **Simulation Parameters** tab, enter 30 as the **Maximum Sample Size**, and 3 for **Cohort Size**. For the **Starting Dose**, select the **`Lowest Dose`**. If you were to check the box **Start with 3+3 Design**, then you would be simulating from the 3+3 design first, before switching to the

CRM, either upon reaching the MTD, or upon observing the first DLT. For this tutorial, however, leave the box unchecked.

Enter $0.25$ for the **Target Probability of Toxicity**, and $0.3$ for the **Target Probability Upper Limit**. This will ensure that the next dose assignment is that whose toxicity probability is closest to $0.25$, and below $0.3$.



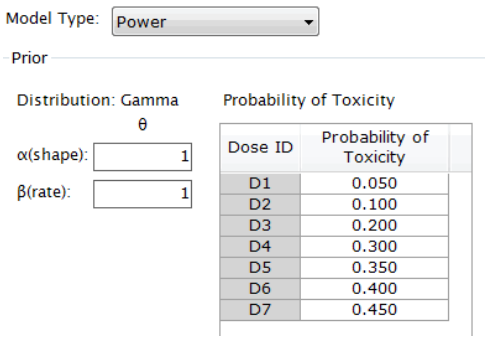

If you were to click **Additional Options...**, a new window will appear, which provides two options corresponding to the original CRM procedure: (1) Allow skipping of untried doses while escalating, and (2) Allow dose escalation even if previous subject experienced DLT.



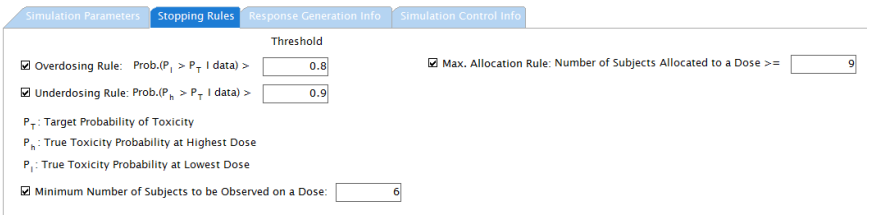

As was recommended in later variations of CRM, in the interests of promoting safety, leave these two options unchecked. This means that no doses will be skipped while escalating, and no dose escalation will occur when the most recent subject experienced a DLT.

For **Model Type**, select **`Power`**, with a Gamma($\alpha = 1, \beta = 1$) prior for $\theta$. Other model types available include the **`Logistic`** and the **`Hyperbolic Tangent`**. Finally, for the prior probabilities of toxicity of all doses (known as the *skeleton*), enter: 0.05, 0.1, 0.2, 0.3, 0.35, 0.4,

and 0.45.



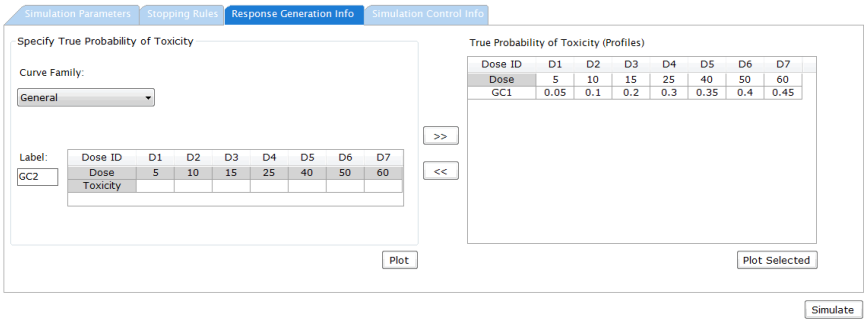

In the **Stopping Rules** tab, you may specify various rules for stopping the trial. Enter the following inputs as below.



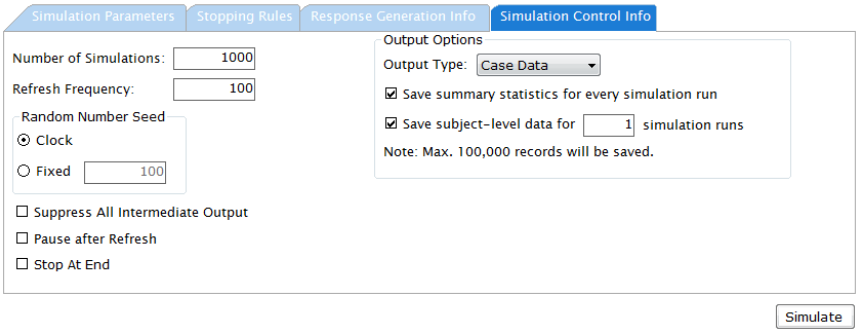

The **Overdosing Rule** states that if the posterior probability of overdosing (toxicity at the lowest dose is greater than target toxicity) exceeds $0.8$, then the trial will be stopped. The **Underdosing Rule** states that if the posterior probability of underdosing (toxicity at the highest dose is lower than target toxicity) exceeds $0.9$, then the trial will be stopped. A minimum of $6$ subjects will need to be observed on a dose before either of these two rules is activated. A further stopping rule is based on the **Max. Allocation Rule**: As soon as $9$ subjects are allocated to any single dose, the trial will be stopped.

In the **Response Generation Info** tab, you can specify a set of true dose response curves from which to simulate. Leave the default profile as shown below. If you wish to edit or add

additional profiles (dose response curves), see the corresponding section for the 3+3 design.




In the **Simulation Control Info** tab, check the boxes corresponding to **`Save summary statistics`** and **`Save subject-level data`**. These options will provides access to several charts derived from these more detailed levels of simulated data. If you wish to display subject-level plots for more than one simulation, you can increase the number. For now, leave this at 1 to save computation time.

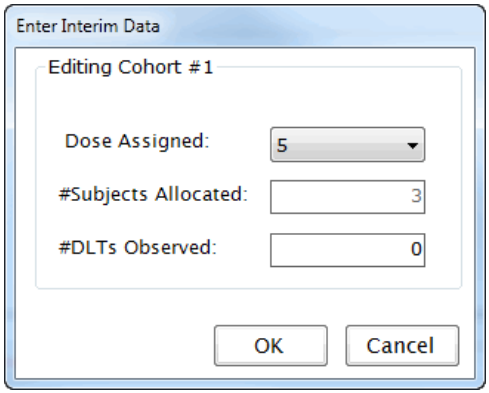



Click **Simulate** to simulate the CRM design. In the **Output Preview** toolbar, click the  icon to save the simulation to the **Library**. Double-click the simulation node in the **Library** to display the simulation output details. Click the Plots icon in the **Library** to access a wide range of available plots.

### 7.2.2 Interim Monitoring

Right-click the Simulation node with CRM in the Library, and select **Interim Monitoring**. This will open an empty interim monitoring dashboard.

Click **Enter Interim Data** to open a window in which to enter data for the first cohort: in particular, the **Dose Assigned** and the **DLTs Observed**. Click **OK** to continue.



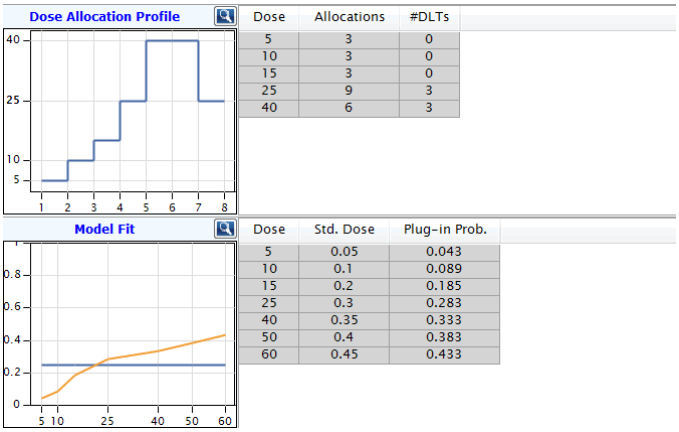

Continue in this manner by clicking **Enter Interim Data**, entering the following doses, and the corresponding number of DLTs: 0 DLTs at dose 10, 0 DLTs at dose 15, 1 DLT at dose 25, 1 DLT at dose 40, 2 DLTs at dose 40, 1 DLT and dose 25, and finally 1 DLT at dose 25.

Enter Interim Data | Final Inference

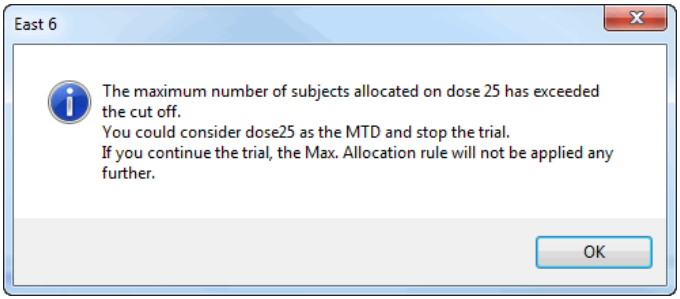
| Cohort # | Dose Assigned | #Subjects | #DLTs | Recommended Dose | Posterior Mean(Θ) |
|---|---|---|---|---|---|
| 1 | 5 | 3 | 0 | 10 | 1.493 |
| 2 | 10 | 3 | 0 | 15 | 1.773 |
| 3 | 15 | 3 | 0 | 25 | 2.074 |
| 4 | 25 | 3 | 1 | 40 | 1.509 |
| 5 | 40 | 3 | 1 | 40 | 1.375 |
| 6 | 40 | 3 | 2 | 25 | 1.105 |
| 7 | 25 | 3 | 1 | 25 | 1.072 |
| 8 | 25 | 3 | 1 | 25 | 1.049 |

## Chapter 7: *Dose Escalation*

After each cohort, East will update the Interim Monitoring Dashboard.




At this point, East recommends that you stop the trial.




.

Click **Final Inference** to generate a table for final inference.

| Final Inference | |
|---|---|
| Final Output at Cohort# | 8 |
| MTD | 25 (under max. allocation) |
| Fitted MTD | 21.662 |

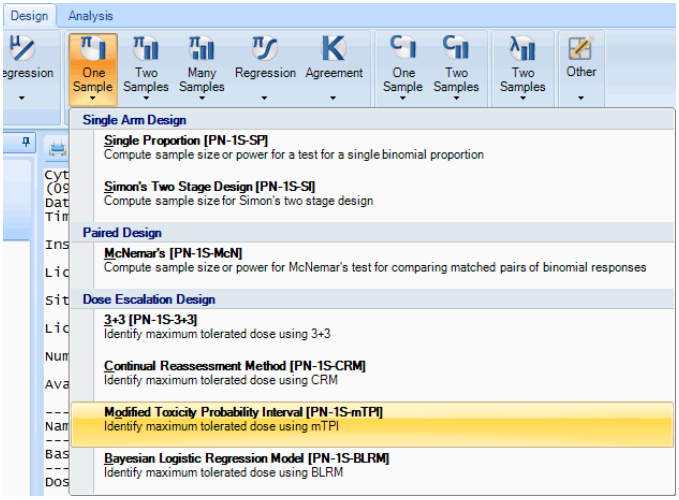
.

## 7.3 modified Toxicity Probability Interval (mTPI)



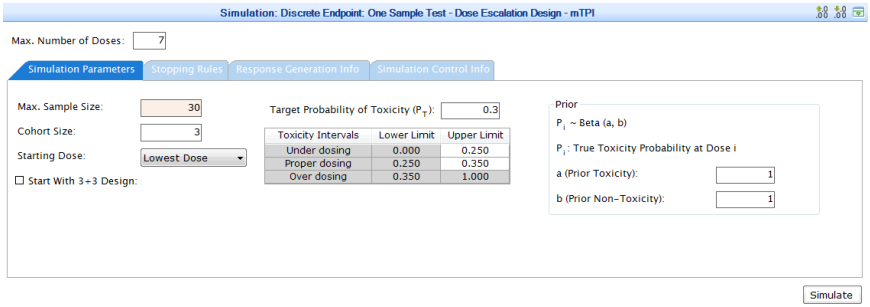
### *7.3.1 Simulation*

Click **Discrete: One Sample** on the Design tab, and then click **Dose Escalation Design: Modified Toxicity Probability Interval**.




In the upper pane of this window is the Input dialog box, which is separated into four tabs: **Simulation Parameters**, **Stopping Rules**, **Response Generation Info**, and **Simulation Control Info**.

# Chapter 7: *Dose Escalation*




In the **Simulation Parameters** tab, enter $30$ as the **Maximum Sample Size**, and $3$ for **Cohort Size**. For the **Starting Dose**, select the `Lowest Dose`. If you were to check the box **Start with 3+3 Design**, then you would be simulating from the 3+3 design first, before switching to the mTPI, either upon reaching the MTD, or upon observing the first DLT. For this tutorial, however, leave the box unchecked.

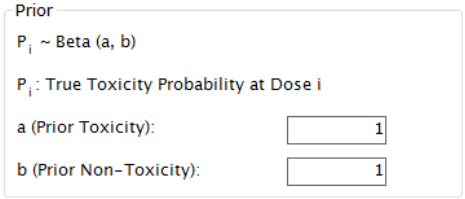
Enter $0.25$ for the **Target Probability of Toxicity**, $0.2$ for the upper limit of the **Under dosing** interval, and $0.3$ for the upper limit of **Proper dosing** interval.

Target Probability of Toxicity ($P_T$): 0.25

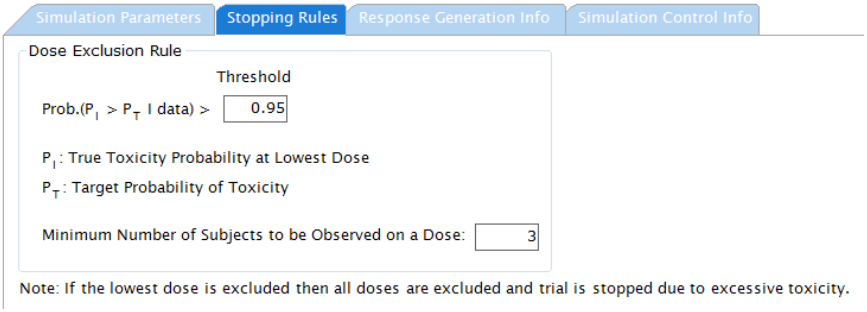
| Toxicity Intervals | Lower Limit | Upper Limit |
|---|---|---|
| Under dosing | 0.000 | 0.2 |
| Proper dosing | 0.200 | 0.3 |
| Over dosing | 0.300 | 1 |

These entries imply that toxicity probabilities within this interval [0.2 to 0.3] can be regarded as equivalent to the target toxicity (0.25) as far as dosing decisions are concerned. Finally, we will

assume a uniform Beta($a = 1$, $b = 1$) prior distribution for all doses.



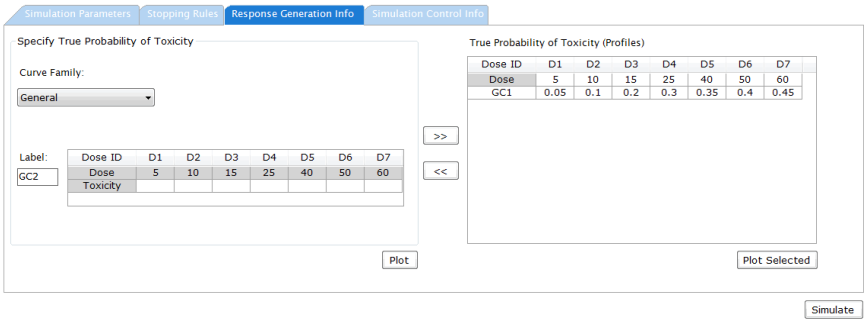

In the **Stopping Rules** tab, enter the following inputs as below.



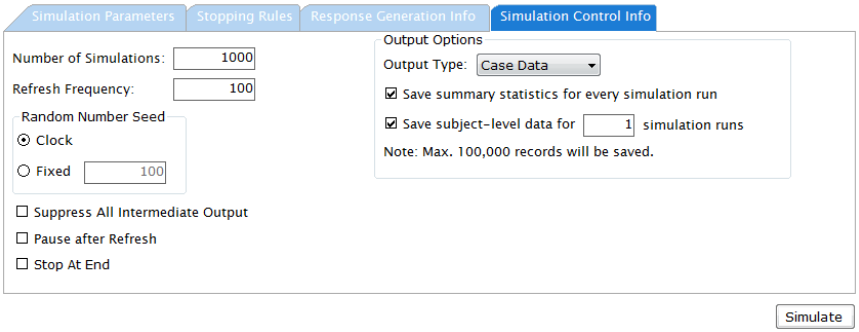

For the mTPI design, the stopping rule is based on dose exclusion rules. This states that if there is greater than a $0.95$ posterior probability that toxicity for a given dose is greater than the target toxicity, that dose and all higher doses will be excluded in subsequent cohorts. When this dose exclusion rule applies to the lowest dose, then all doses are excluded, and hence the trial will be stopped for excessive toxicity. Furthermore, the dose exclusion rule is not activated until at least $3$ subjects are observed on a dose.

In the **Response Generation Info** tab, you can specify a set of true dose response curves from which to simulate. Leave the default profile as shown below. If you wish to edit or add

additional profiles (dose response curves), see the corresponding section for the 3+3 design.



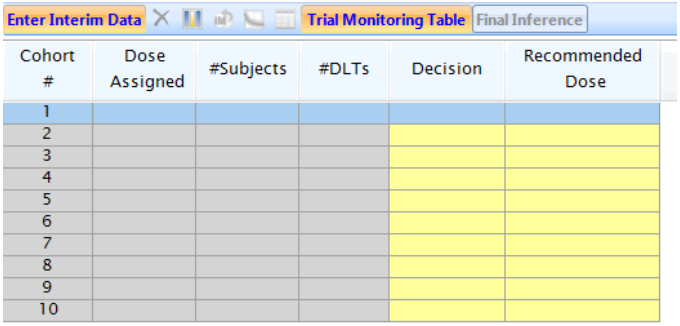

In the **Simulation Control Info** tab, check the boxes corresponding to **Save summary statistics** and **Save subject-level data**. These options will provides access to several charts derived from these more detailed levels of simulated data. If you wish to display subject-level plots for more than one simulation, you can increase the number. For now, leave this at 1 to save computation time.




Click **Simulate** to simulate the mTPI design. In the **Output Preview** toolbar, click the  icon to save the simulation to the **Library**. Double-click the simulation node in the **Library** to display the simulation output details. Click the Plots icon in the **Library** to access a wide range of available plots.

### 7.3.2 Interim Monitoring

Right-click one of the Simulation nodes with mTPI in the Library, and select **Interim Monitoring**. This will open an empty interim monitoring dashboard.




| Cohort # | Dose Assigned | #Subjects | #DLTs | Decision | Recommended Dose |
|---|---|---|---|---|---|
| 1 | | | | | |
| 2 | | | | | |
| 3 | | | | | |
| 4 | | | | | |
| 5 | | | | | |
| 6 | | | | | |
| 7 | | | | | |
| 8 | | | | | |
| 9 | | | | | |
| 10 | | | | | |

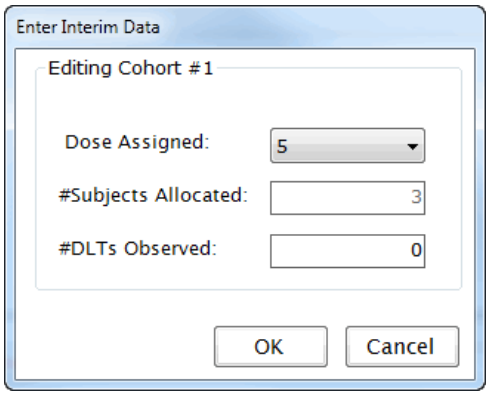
Click on **Trial Monitoring Table** to generate a table to guide dosing decisions for this trial. For example, if the cumulative number of patients treated at the current dose is 8, and the cumulative number of toxicities at this dose is 3, then the mTPI method recommends a Stay decision. Close this table before continuing.

Number of patients treated at current dose

| Number of Toxicities | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | E | E | E | E | E | E | E | E | E | E | E | E | E | E | E | E | E | E | E | E | E | E |
| 1 | D | S | S | S | S | E | E | E | E | E | E | E | E | E | E | E | E | E | E | E | E | E |
| 2 | | DU | D | S | S | S | S | S | S | S | E | E | E | E | E | E | E | E | E | E | E | E |
| 3 | | | DU | DU | D | S | S | S | S | S | S | S | S | S | E | E | E | E | E | E | E | E |
| 4 | | | | DU | DU | DU | D | D | S | S | S | S | S | S | S | S | S | S | S | S | E | E |
| 5 | | | | | DU | DU | DU | DU | DU | D | S | S | S | S | S | S | S | S | S | S | S | S |
| 6 | | | | | | DU | DU | DU | DU | DU | DU | D | S | S | S | S | S | S | S | S | S | S |
| 7 | | | | | | | DU | DU | DU | DU | DU | DU | DU | D | S | S | S | S | S | S | S | S |
| 8 | | | | | | | | DU | DU | DU | DU | DU | DU | DU | DU | DU | D | S | S | S | S | S |
| 9 | | | | | | | | | DU | DU | DU | DU | DU | DU | DU | DU | DU | DU | DU | S | S | S |
| 10 | | | | | | | | | | DU | DU | DU | DU | DU | DU | DU | DU | DU | DU | DU | DU | D |
| 11 | | | | | | | | | | | DU | DU | DU | DU | DU | DU | DU | DU | DU | DU | DU | DU |
| 12 | | | | | | | | | | | | DU | DU | DU | DU | DU | DU | DU | DU | DU | DU | DU |
| 13 | | | | | | | | | | | | | DU | DU | DU | DU | DU | DU | DU | DU | DU | DU |
| 14 | | | | | | | | | | | | | | DU | DU | DU | DU | DU | DU | DU | DU | DU |
| 15 | | | | | | | | | | | | | | | DU | DU | DU | DU | DU | DU | DU | DU |

Click **Enter Interim Data** to open a window in which to enter data for the first cohort: in

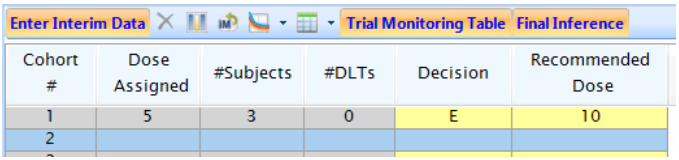
particular, the **Dose Assigned** and the **DLTs Observed**. Click **OK** to continue.




The dashboard will be updated accordingly. The decision for the next cohort is based on the highest Unit Probability Mass (UPM): the posterior probability for each toxicity interval divided by the length of the interval. The underdosing interval corresponds to an E (Escalate) decision, the proper dosing interval corresponds to an S (Stay) decision, and the overdosing interval corresponds to a D (De-escalate) decision. In this case, the UMP for underdosing is highest.

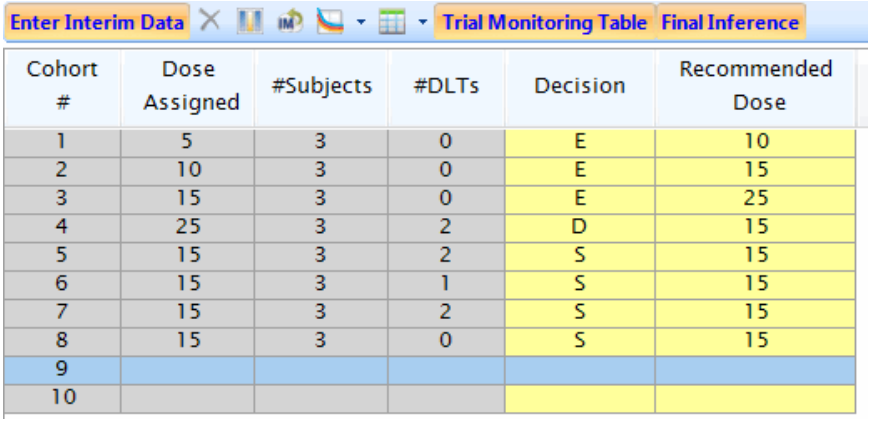
| Dose | UPM(E) | UPM(S) | UPM(D) |
|---|---|---|---|
| 5 | 2.734 | 1.379 | 0.275 |

Thus, the recommendation is to escalate to dose 10.




Continue in this manner by entering data for each subsequent cohort, and observe how the

interim monitoring dashboard updates. One example is given below.




| Cohort # | Dose Assigned | #Subjects | #DLTs | Decision | Recommended Dose |
|---|---|---|---|---|---|
| 1 | 5 | 3 | 0 | E | 10 |
| 2 | 10 | 3 | 0 | E | 15 |
| 3 | 15 | 3 | 0 | E | 25 |
| 4 | 25 | 3 | 2 | D | 15 |
| 5 | 15 | 3 | 2 | S | 15 |
| 6 | 15 | 3 | 1 | S | 15 |
| 7 | 15 | 3 | 2 | S | 15 |
| 8 | 15 | 3 | 0 | S | 15 |
| 9 | | | | | |
| 10 | | | | | |

Suppose we wished to end the study after 8 cohorts (24 patients). Click **Final Inference** to end the study and generate a table of final inference. Here, the MTD is 15, while the fitted MTD is 13.269, estimated from the interpolated isotonic estimates.

| Final Inference | |
|---|---|
| Final Output at Cohort# | 8 |
| MTD | 15 |
| Fitted MTD | 13.269 |

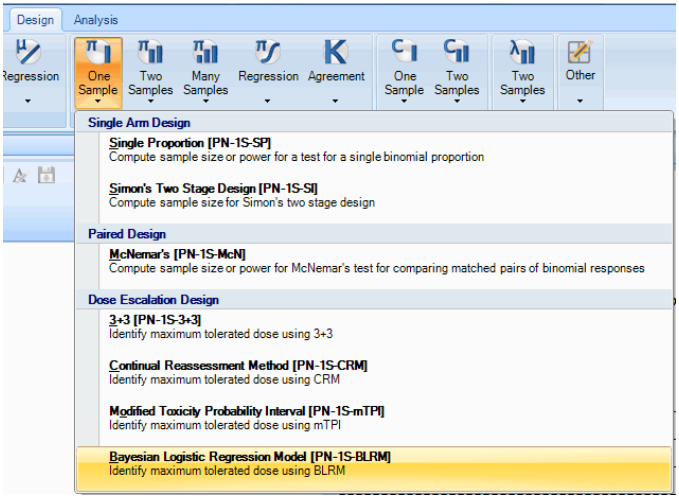
## 7.4 Bayesian logistic regression model (BLRM)

### *7.4.1 Simulation*

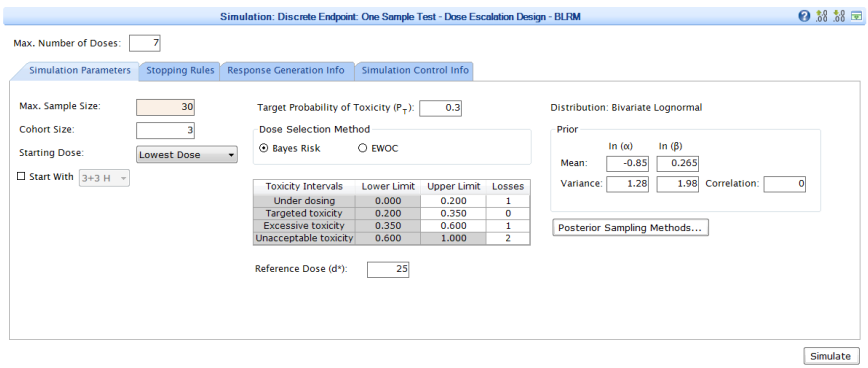
Click **Discrete: One Sample** on the Design tab, and then click **Dose Escalation Design:**

# Chapter 7: *Dose Escalation*

**Bayesian Logistic Regression Model**.



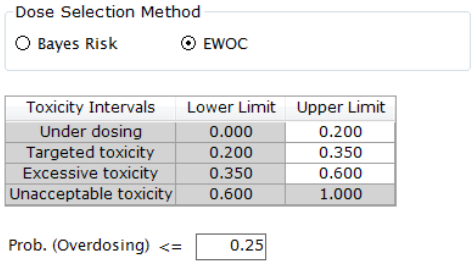

In the upper pane of this window is the Input dialog box, which is separated into four tabs: **Simulation Parameters**, **Stopping Rules**, **Response Generation Info**, and **Simulation Control Info**.



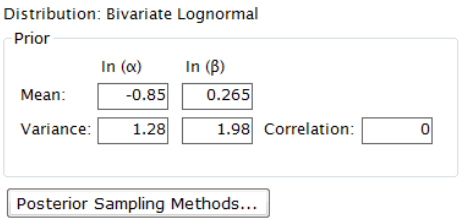

In the **Simulation Parameters** tab, enter 30 as the **Maximum Sample Size**, and 3 for **Cohort Size**. For the **Starting Dose**, select the `Lowest Dose`. If you were to check the box **Start with**

**3+3 Design**, then you would be simulating from the 3+3 design first, before switching to the BLRM, either upon reaching the MTD, or upon observing the first DLT. For this tutorial, however, leave the box unchecked.

The next step is to choose a **Dose Selection Method**: either by **`Bayes Risk`** or by **`EWOC`**. For the next cohort, the Bayes risk method selects the dose that minimizes the posterior expected loss, aka Bayes risk. In contrast, the escalation with overdose control (EWOC) method selects the dose that maximizes the posterior probability of targeted toxicity, for all doses where the posterior probability of overdosing (either excessive or unacceptable toxicity) is less than the user-specified threshold. In this example, we will use the EWOC method.

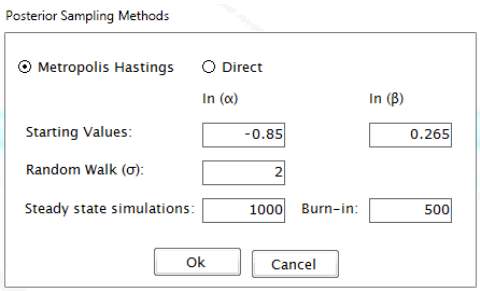



A bivariate normal distribution with corresponding means, variances, and correlation, can be specified for the $\ln(\alpha)$ and $\ln(\beta)$ parameters of the two-parameter logistic.

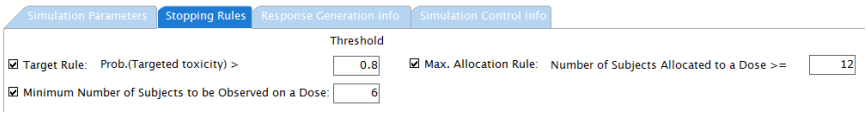



Click **Posterior Sampling Methods** to select from one of two methods: Metropolis Hastings,

or direct Monte Carlo. For this tutorial, click **OK** to select **`Metropolis Hastings`**.



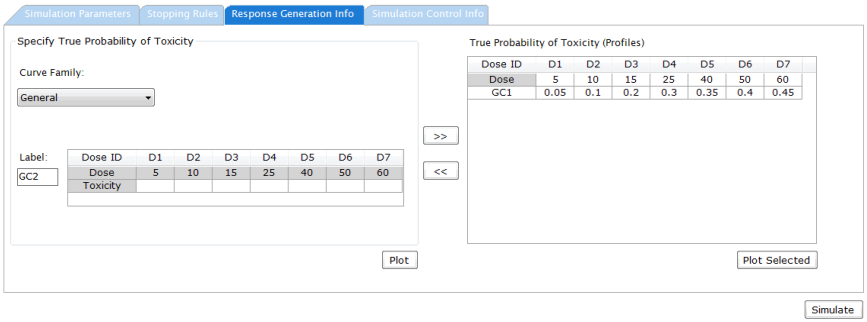

In the **Stopping Rules** tab, you can specify up to two rules for stopping the trial. Check the appropriate boxes and enter values as below.



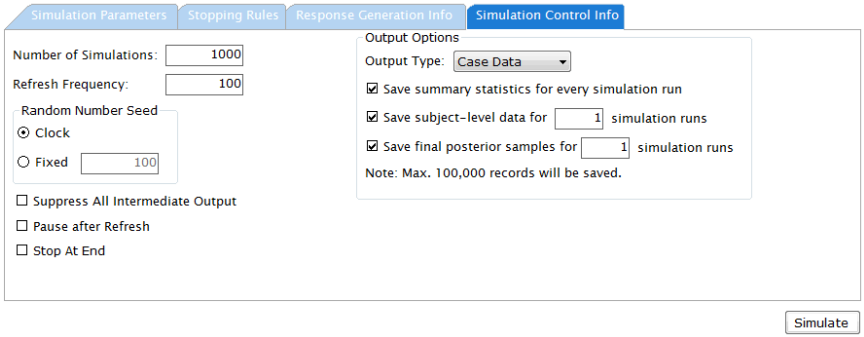

The **Target Rule** will stop the trial when the posterior probability of being in the Target toxicity interval exceeds $0.8$. A minimum of $6$ subjects should be observed before this rule is activated. The **Max. Allocation Rule** will stop the trial if at least $12$ subjects are allocated to any dose.

In the **Response Generation Info** tab, you can specify a set of true dose response curves from which to simulate. Leave the default profile as shown below. If you wish to edit or add

additional profiles (dose response curves), see the corresponding section for the 3+3 design.



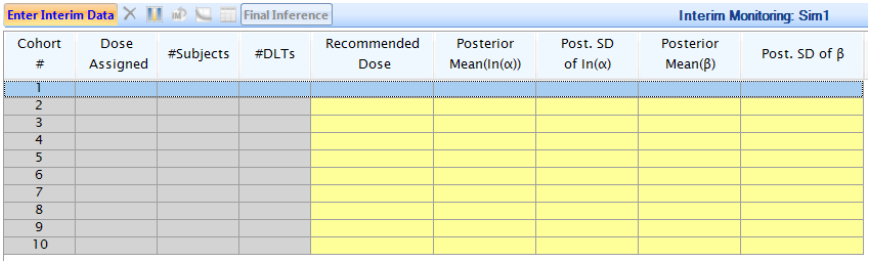

In the **Simulation Control Info** tab, check the boxes corresponding to **Save summary statistics**, **Save subject-level data**, and **Save final posterior samples**. These options will provides access to several charts derived from these more detailed levels of simulated data. If you wish to display subject-level plots, or posterior distribution plots, for more than one simulation, you can increase the number. For now, leave both of these at 1 to save computation time.



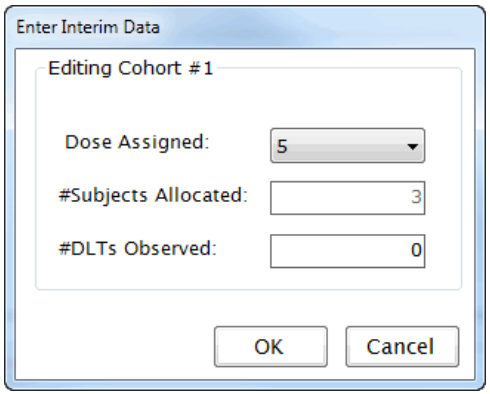

Click **Simulate** to simulate the BLRM design. In the **Output Preview** toolbar, click the  icon to save the simulation to the **Library**. Double-click the simulation node in the **Library** to display the simulation output details. Click the Plots icon in the **Library** to access a wide range of available plots.

### *7.4.2 Interim Monitoring*

Right-click the Simulation node with BLRM in the Library, and select **Interim Monitoring**. This will open an empty interim monitoring dashboard.




Click **Enter Interim Data** to open a window in which to enter data for the first cohort: in particular, the **Dose Assigned** and the **DLTs Observed**. Click **OK** to continue.




The dashboard will be updated accordingly. The decision for the next cohort is based on the dose with the highest posterior probability of targeted toxicity, with less than the

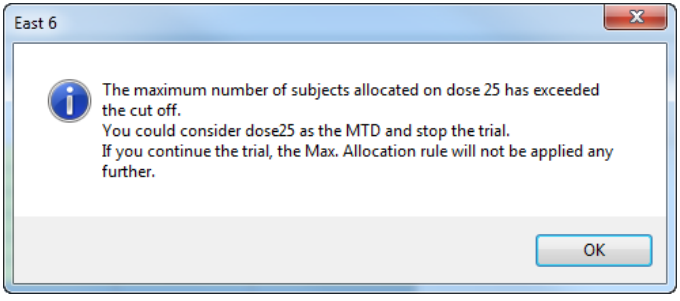
user-specified threshold (0.25) probability of overdosing. In this case, this is dose 15.

| Cohort # | Dose Assigned | #Subjects | #DLTs | Recommended Dose |
|---|---|---|---|---|
| 1 | 5 | 3 | 0 | 15 |
| 2 | | | | |
| 3 | | | | |
| 4 | | | | |
| 5 | | | | |
| 6 | | | | |
| 7 | | | | |
| 8 | | | | |
| 9 | | | | |
| 10 | | | | |

Continue in this manner by entering data for each subsequent cohort, and observe how the interim monitoring dashboard updates. One example is given below.

| Cohort # | Dose Assigned | #Subjects | #DLTs | Recommended Dose |
|---|---|---|---|---|
| 1 | 5 | 3 | 0 | 15 |
| 2 | 15 | 3 | 0 | 15 |
| 3 | 15 | 3 | 0 | 25 |
| 4 | 25 | 3 | 0 | 25 |
| 5 | 25 | 3 | 1 | 25 |
| 6 | 25 | 3 | 1 | 25 |
| 7 | | | | |
| 8 | | | | |
| 9 | | | | |
| 10 | | | | |

At this point, East will display the following message:

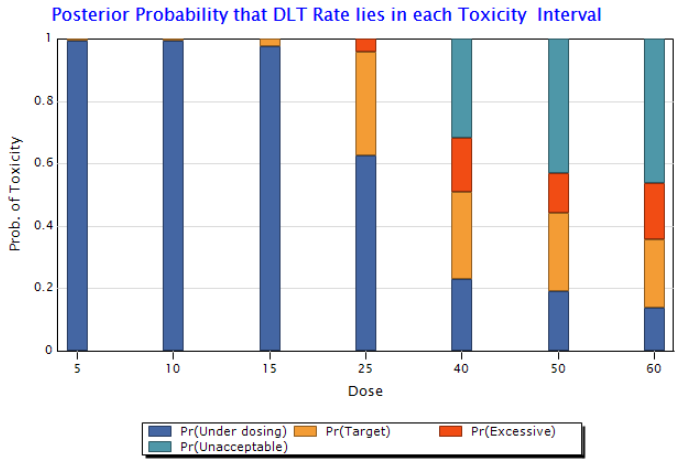



Click **OK** to continue, and then click **Final Inference** Interim Monitoring toolbar. The following

final inference table will be displayed.

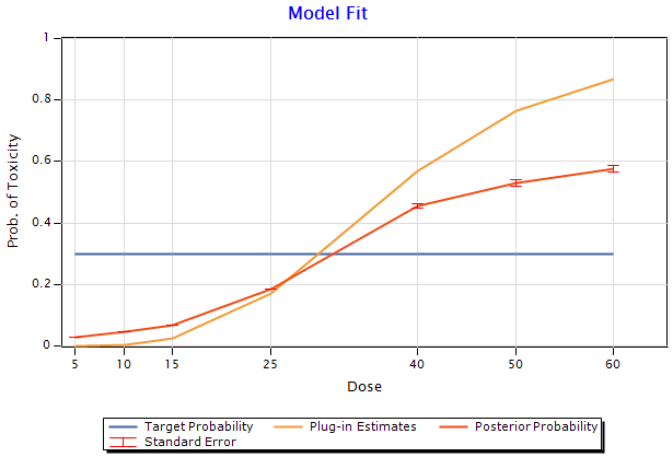
| Final Inference | |
|---|---|
| Final Output at Cohort# | 7 |
| MTD | 25 (under max. allocation) |
| Fitted MTD | 30.12 |

Dose 25 happens to have the highest posterior probability of being in the target toxicity interval, with little probability of being in the overdosing (excessive or unacceptable) intervals.




The **Model Fit** plot has also been updated. One curve is the two-parameter logistic function described by plug-in (posterior mean) estimates. The other curve interpolates between the posterior mean of the toxicity probability at each dose. In both cases, the target probability lies

quite close to the predicted toxicity probability at dose 25, with the fitted MTD at around 30.



# Volume 4 Exact Binomial Designs

# 8 Count Data One-Sample

This chapter deals with the design of tests involving count or Poisson response rates. Here, independent outcomes or events under examination can be counted in terms of whole numbers, and typically are considered rare. In other words, a basic assumption of the Poisson distribution is that the probability of an event occurring is proportional to the length of time under consideration. The longer the time interval, the more likely the event will occur. Therefore, in this context interest lies in the rate of occurrence of a particular event during a specified period. Section 8.1 focuses on designs in which an observed Poisson response rate is compared to a fixed response rate, possibly derived from historical data.

## 8.1 Single Poisson Rate

Data following a Poisson distribution are non-negative integers, and the probability that an outcome occurs exactly $k$ times can be calculated as:

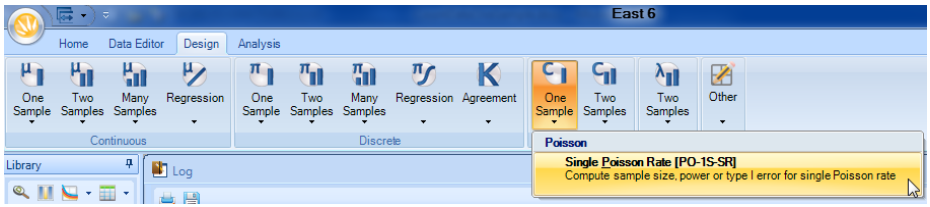
$P(k) = \frac{e^{-\lambda}\lambda^k}{k!}, k = 0, 1, 2, \ldots$ where $\lambda$ is the average rate of occurrence.

When comparing a new protocol or treatment to a well-established control, a preliminary single-sample study may result in valuable information prior to a full-scale investigation. In experimental situations it may be of interest to determine whether the response rate $\lambda$ differs from a fixed value $\lambda_0$. Specifically we wish to test the null hypothesis $H_0$: $\lambda = \lambda_0$ against the two sided alternative hypothesis $H_1$: $\lambda \neq \lambda_0$ or against one sided alternatives of the form $H_1$: $\lambda > \lambda_0$ or $H_1$: $\lambda < \lambda_0$. The sample size, or power, is determined for a specified value of $\lambda$ which is consistent with the alternative hypothesis, denoted $\lambda_1$.

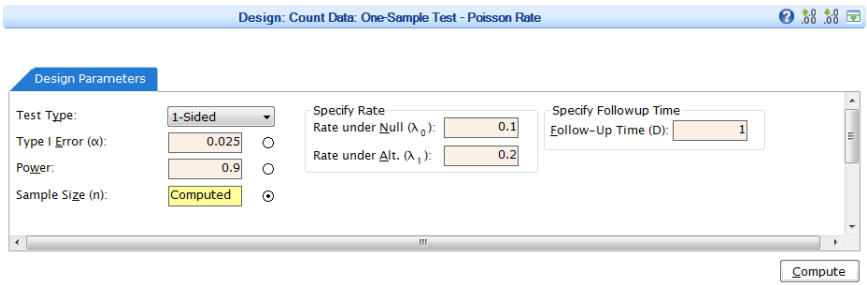
### 8.1.1 Trial Design

Consider the design of a single-arm clinical trial in which we wish to determine if the positive response rate of a new acute pain therapy is at least 30% per single treatment cycle. Thus, it is desired to test the null hypothesis $H_0$: $\lambda = 0.2$ against the one-sided alternative hypothesis $H_1$: $\lambda \geq 0.3$. The trial will be designed such that a one sided $\alpha = 0.05$ test achieves 80% power at $\lambda = \lambda_1 = 0.3$.

In the **Design** tab under the **Count** group choose **One Sample** and then **Single Poisson Rate**.



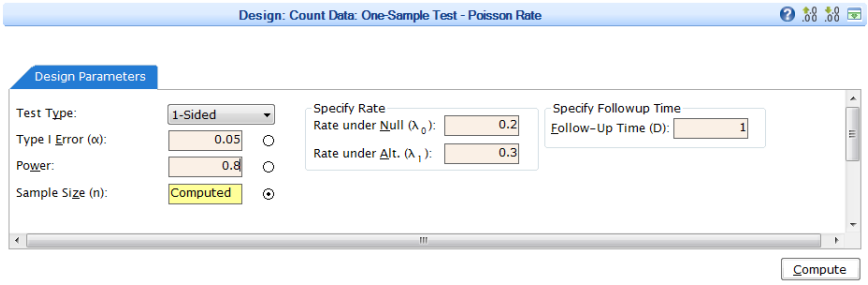

This will launch the following input window:



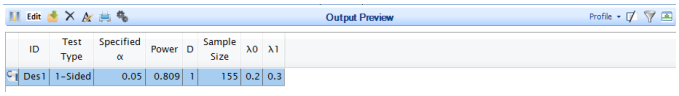

Enter the following design parameters:

Test Type: 1 sided
Type 1 Error $(\alpha)$: 0.05
Power: 0.8
Sample Size (n): Computed (select radio button)
Rate under Null $(\lambda_0)$: 0.2
Rate under Alt. $(\lambda_1)$: 0.3

Follow-up Time $(D)$: 1




Click **Compute**. The design is shown as a row in the **Output Preview** window:

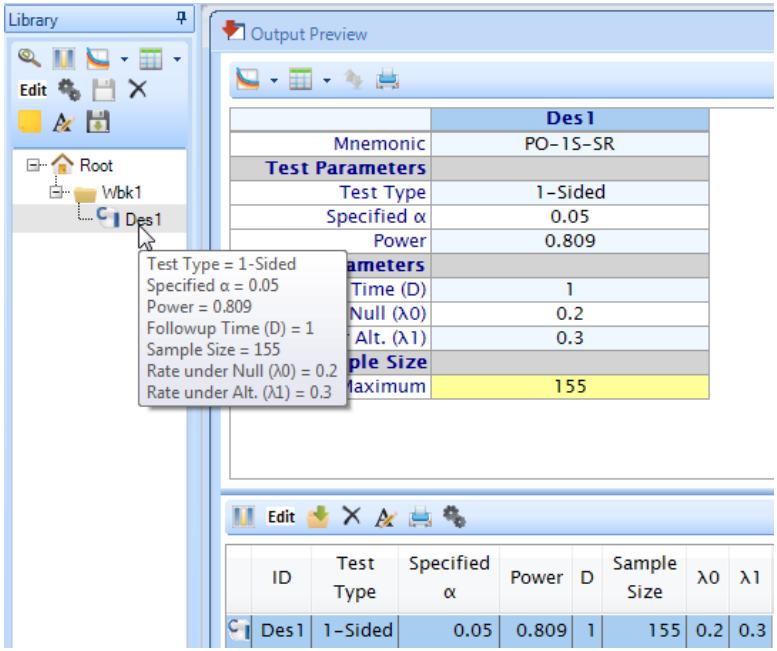



The sample size required in order to achieve the desired 80% power is 194 subjects. As is standard in East, this design has the default name **Des 1**. To see a summary of the output of this design, click anywhere in the row and then click the icon in the Output Preview toolbar. The design details will be displayed in the upper pane, labeled **Output Summary**.

| | Des1 |
|---|---|
| Mnemonic | PO-1S-SR |
| **Test Parameters** | |
| Test Type | 1-Sided |
| Specified α | 0.05 |
| Power | 0.809 |
| **Model Parameters** | |
| Follow-Up Time (D) | 1 |
| Rate under Null (λ0) | 0.2 |
| Rate under Alt. (λ1) | 0.3 |
| **Sample Size** | |
| Maximum | 155 |

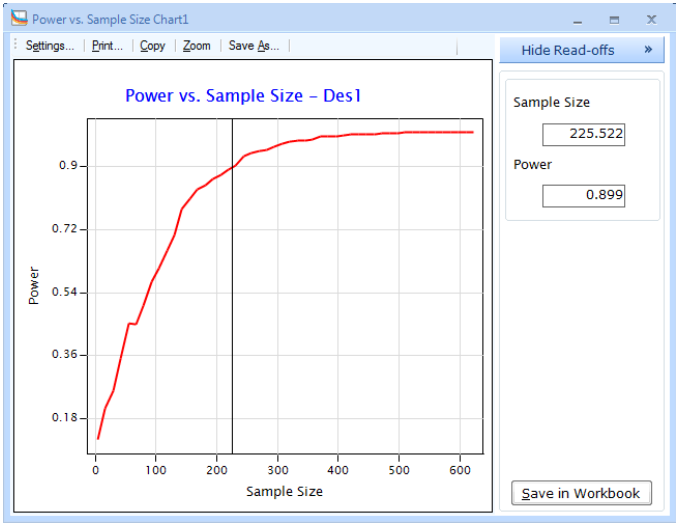
In the **Output Preview** toolbar, click icon to save this design **Des1** to workbook **Wbk1** in the **Library**. An alternative method to view design details is to hover the cursor over the node **Des1** in the **Library**. A tooltip will appear that summarizes the input parameters of the

design.

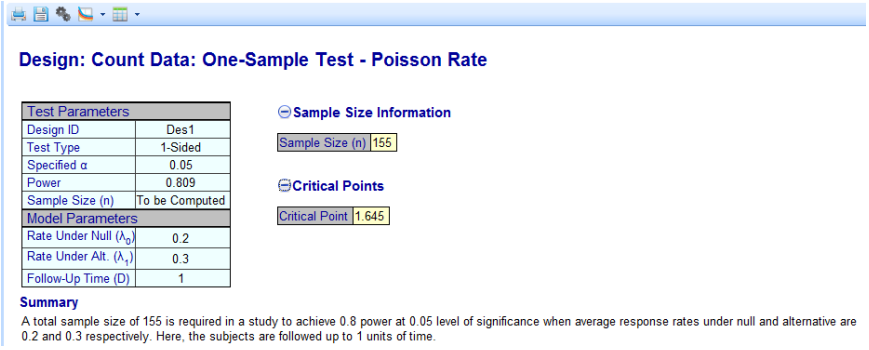



With the design **Des1** selected in the Library, click icon on the Library toolbar, and then click **Power vs. Sample Size**. The power curve for this design will be displayed. You can save this chart to the Library by clicking **Save inWorkbook**. Alternatively, you can export the chart in one of several image formats (e.g., Bitmap or JPEG) by clicking **Save As...** or **Export** into a

PowerPoint presentation.




Close the Power vs. Sample Size chart. To view a summary of all characteristics of this design, select **Des1** in the **Library**, and click  icon.



**Design: Count Data: One-Sample Test - Poisson Rate**

| Test Parameters | |
|---|---|
| Design ID | Des1 |
| Test Type | 1-Sided |
| Specified α | 0.05 |
| Power | 0.809 |
| Sample Size (n) | To be Computed |
| **Model Parameters** | |
| Rate Under Null ($\lambda_0$) | 0.2 |
| Rate Under Alt. ($\lambda_1$) | 0.3 |
| Follow-Up Time (D) | 1 |

**Sample Size Information**

| Sample Size (n) | 155 |
|---|---|

**Critical Points**

| Critical Point | 1.645 |
|---|---|

**Summary**

A total sample size of 155 is required in a study to achieve 0.8 power at 0.05 level of significance when average response rates under null and alternative are 0.2 and 0.3 respectively. Here, the subjects are followed up to 1 units of time.

In addition to the **Power vs. Sample** size chart and table, East also provides the efficacy boundary in the **Stopping Boundaries** chart and table.

Alternatively, East allows the computation of either the **Type-1 error ($\alpha$)** or **Power** for a given sample size. Using the **Design Input/Output** window as described above, simply enter the desired sample size and click **Compute** to calculate the resulting power of the test.

# 9 Count Data Two-Samples

Often in experiments based on count data, the aim is to compare independent samples from two populations in terms of the rate of occurrence of a particular outcome. In medical research, outcomes such as the number of times a patient responds to a therapy, develops a certain side effect, or requires specialized care, are of interest. Or perhaps a therapy is being evaluated to determine the number of times it must be applied until an acceptable response rate is observed. East supports the design of clinical trials in which this comparison is based on the ratio of rates, assuming a Poisson or Negative Binomial distribution. These two cases are presented in Sections 9.1 and 9.2, respectively.

## 9.1 Poisson - Ratio of Rates



Let $\lambda_c$ and $\lambda_t$ denote the Poisson rates for the control and treatment arms, respectively, and let $\rho_1 = \lambda_t/\lambda_c$. We want to test the null hypothesis that $\rho_1 = 1$ against one or two-sided alternatives. The sample size, or power, is determined to be consistent with the alternative hypothesis, that is $H_1 : \lambda_t \neq \lambda_c$, $H_1 : \lambda_t > \lambda_c$, or $H_1 : \lambda_t < \lambda_c$.

### 9.1.1 Trial Design

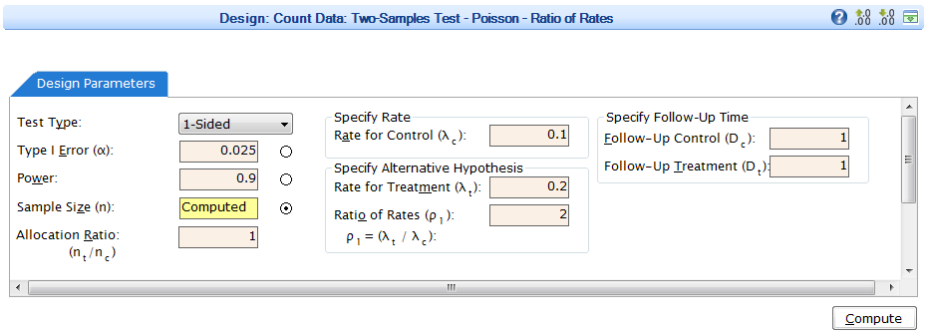
Suppose investigators are preparing design objectives for a prospective randomized trial of a standard treatment (control arm) vs. a new combination of medications (therapy arm) to present at a clinical trials workshop. The endpoint of interest is the number of abnormal ECGs (electrocardiogram) within seven days. The investigators were interested in comparing the therapy arm to the control arm with a two sided test conducted at the 0.025 level of

## Chapter 9: *Count Data Two-Samples*

significance. It can be assumed that the rate of abnormal ECGs in the control arm is 30%, thus $\lambda_t = \lambda_c = 0.3$ under $H_0$. The investigators wish to determine the sample size to attain power of 80% if there is a 25% decline in the event rate, that is $\lambda_t/\lambda_c = 0.75$. It is important to note that the power of the test depends on $\lambda_c$ and $\lambda_t$, not just the ratio, so different values of the pair ($\lambda_c$ , $\lambda_t$) with the same ratio will yield different solutions.

We will now design a study that compares the control arm to the combination therapy arm. In the **Design** tab under the **Count** group choose **Two Samples** and then **Poisson - Ratio of Rates**.

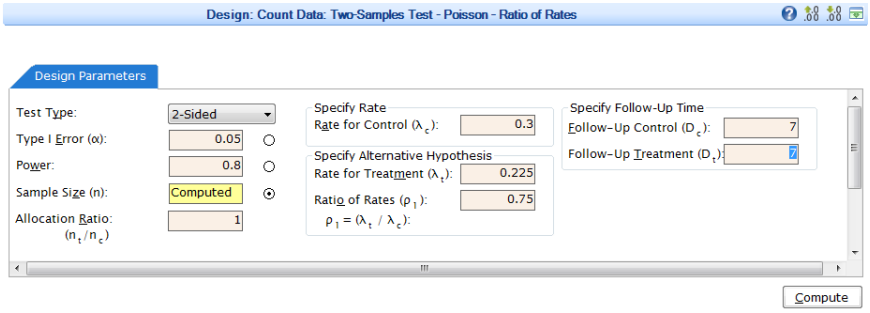

This will launch the following input window:



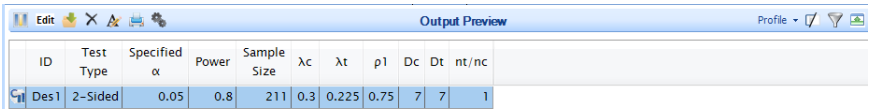

Enter the following design parameters:

Test Type: 2-sided
Type 1 Error $(\alpha)$: 0.05
Power: 0.8
Sample Size (n): Computed (select radio button)
Allocation Ratio $(n_t/n_c)$: 1

Rate for Control $(\lambda_c)$: 0.3
Rate for Treatment $(\lambda_t)$: 0.225 (will be automatically calculated)
Ratio of Rates $\rho_1 = (\lambda_t/\lambda_c)$: 0.75
Follow-up Control $(D_c)$: 7
Follow-up Treatment $(D_t)$: 7

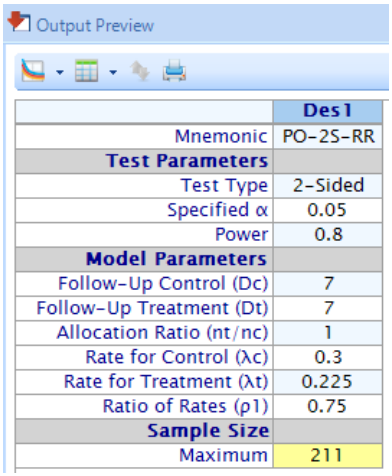



The Allocation Ratio $(n_t : n_c)$ describes the ratio of patients to each arm. For example, an allocation ratio of 3:1 indicates that 75% of the patients are randomized to the treatment arm as opposed to 25% to the control. Here we assume the same number of patients in both arms. Click **Compute**. The design is shown as a row in the **Output Preview** window:




The sample size required in order to achieve the desired 80% power is 211 subjects. As is standard in East, this design has the default name **Des 1**. To see a summary of the output of this design, click anywhere in the row and then click the  icon in the Output Preview

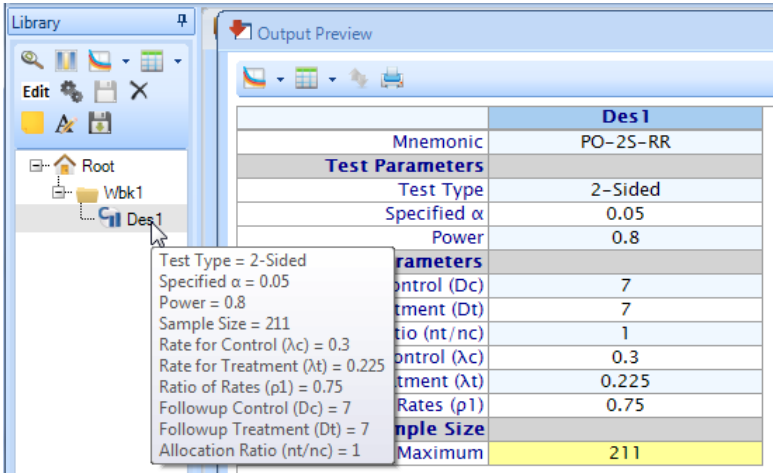
toolbar. The design details will be displayed in the upper pane, labeled **Output Summary**.


Output Preview

| | Des1 |
|---|---|
| Mnemonic | PO-2S-RR |
| **Test Parameters** | |
| Test Type | 2-Sided |
| Specified α | 0.05 |
| Power | 0.8 |
| **Model Parameters** | |
| Follow-Up Control (Dc) | 7 |
| Follow-Up Treatment (Dt) | 7 |
| Allocation Ratio (nt/nc) | 1 |
| Rate for Control (λc) | 0.3 |
| Rate for Treatment (λt) | 0.225 |
| Ratio of Rates (ρ1) | 0.75 |
| **Sample Size** | |
| Maximum | 211 |

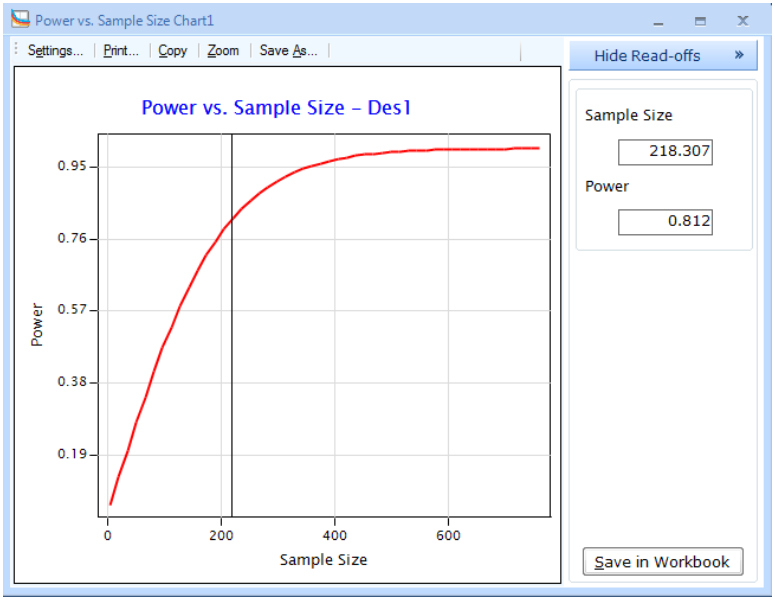
In the **Output Preview** toolbar, click icon to save this design **Des1** to workbook **Wbk1** in the **Library**. An alternative method to view design details is to hover the cursor over the node **Des1** in the **Library**. A tooltip will appear that summarizes the input parameters of the

design.




With the design **Des1** selected in the Library, click icon on the Library toolbar, and then click **Power vs. Sample Size**. The power curve for this design will be displayed. You can save this chart to the Library by clicking **Save inWorkbook**. Alternatively, you can export the chart in one of several image formats (e.g., Bitmap or JPEG) by clicking **Save As...** or **Export** into a

PowerPoint presentation.




Close the Power vs. Sample Size chart. To view all computed characteristics of this design,

select **Des1** in the **Library**, and click  icon.

**Design: Count Data: Two-Sample Test - Poisson - Ratio of Rates**

| Test Parameters | |
|---|---|
| Design ID | Des1 |
| Test Type | 2-Sided |
| Specified α | 0.05 |
| Power | 0.8 |
| Sample Size (n) | To be Computed |
| **Model Parameters** | |
| Rate Under Control ($\lambda_c$) | 0.3 |
| Rate Under Treatment ($\lambda_t$) | 0.225 |
| $\rho_1 = \lambda_t / \lambda_c$ | |
| Under H0 | 1 |
| Under H1 | 0.75 |
| Follow-Up Control ($D_c$) | 7 |
| Follow-Up Treatment ($D_t$) | 7 |
| Allocation Ratio ($n_t/n_c$) | 1 |

**Sample Size Information**

| | |
|---|---|
| Sample Size (n) | 211 |
| Treatment (n_t) | 105 |
| Control (n_c) | 106 |

**Critical Points**

| | |
|---|---|
| Lower Critical Point | -1.96 |
| Upper Critical Point | 1.96 |

**Summary**

A total sample size of 211 is required in a study to achieve 0.8 power at 0.05 level of significance when response rates for Control group and Treatment group are 0.3 and 0.225 respectively. Here, the subjects from the two groups are followed up to 7 and 7 units of time.

In addition to the **Power vs. Sample** size chart and table, East also provides the efficacy boundary in the **Stopping Boundaries** chart and table.

Alternatively, East allows the computation of either the **Type-1 error ($\alpha$)** or **Power** for a given sample size. Using the **Design Input Output** window as described above, simply enter the desired sample size and click **Compute** to calculate the resulting power of the test.

## 9.2 Negative Binomial Ratio of Rates

In experiments where the data follows a binomial distribution, the number of successful outcomes for a fixed number of trials is of importance when determining the sample size to adequately power a study. Suppose instead that it is of interest to observe a fixed number of successful outcomes (or failures), but the overall number of trials necessary to achieve this is unknown. In this case, the data is said to follow a Negative Binomial Distribution. There are two underlying parameters of interest. As with the Poisson distribution, $\lambda$ denotes the average rate of response for a given outcome. In addition, a **shape** parameter $\gamma$ specifies the desired

number of observed "successes". As with the Poisson distribution, the Negative Binomial distribution can be useful when designing a trial where one must wait for a particular event. Here, we are waiting for a specific number of successful outcomes to occur. A Poisson regression analysis assumes a common rate of events for all subjects within a stratum, as well as equal mean and variance (equidispersion). With over dispersed count data, estimates of standard error from these models can be invalid, leading to difficulties in planning a clinical trial. Increased variability resulting from over dispersed data requires a larger sample size in order to maintain power. To address this issue of allowing variability between patients, East provides valid sample size and power calculations for count data using a negative binomial model, resulting in a better evaluation of study design and increased likelihood of trial success.

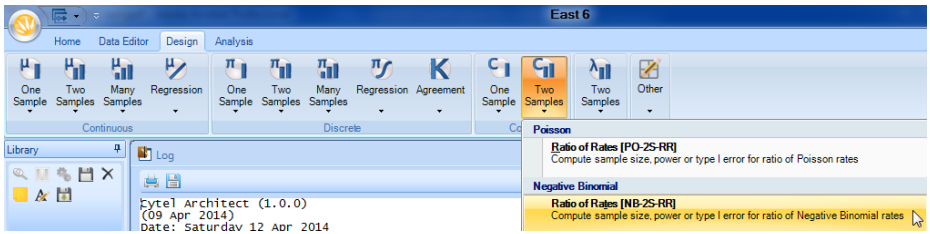
### 9.2.1 Trial Design

Suppose that a hypothetical manufacturer of robotic prostheses, those that require several components to fully function, has an order to produce a large quantity of artificial limbs. According to historical data, about 20% of the current limbs fail the rigorous quality control test and therefore cannot be shipped to patients. For each order, the manufacturer must produce more than requested; in fact they must continue to produce the limbs until the desired quantity passes quality control. Given that there is a high cost in producing these prosthetic limbs, it is of great interest reduce the number of those that fail the test.

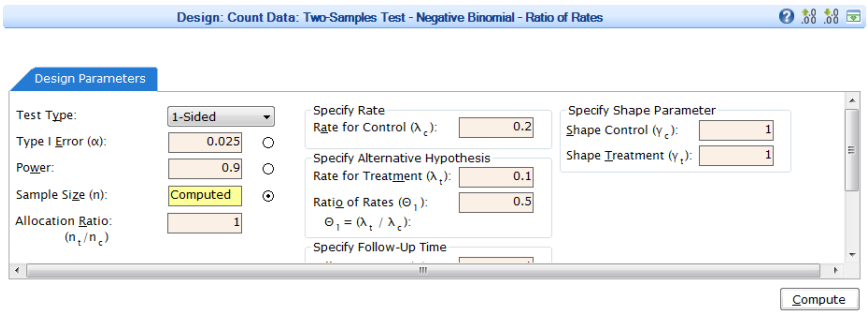
The company plans to introduce a new feature to the current model, the goal being the probability of failure is reduced to 10%. It is safe to assume that the enhancement will not cause a decline in the original success rate. In this scenario, we wish to test the null hypothesis $H_0$: $\lambda_c = \lambda_t = 0.2$ against the one sided alternative of the form $H_1$: $\lambda_c > \lambda_t$. Quality control investigators wish to conduct a one-sided test at the $\alpha = 0.05$ significance level to determine the sample size required obtain 90% power to observe a 50% decline in the event rate, i.e. $\lambda_t/\lambda_c = 0.5$. It is important to note that the power of the test depends on $\lambda_c$ and $\lambda_t$, not just the ratio, so different values of the pair ($\lambda_c$ , $\lambda_t$) with the same ratio will have different solutions. The same holds true for the shape parameter. Different values of ($\gamma_c$ , $\gamma_t$) will result in different sample sizes or power calculations. East allows user specific shape parameters for both the treatment and control groups, however for this example assume that the desired

number of successful outcomes for both groups is 10.

The following illustrates the design of a two-arm study comparing the control arm, which the current model of the prosthesis, to the treatment arm, which is the enhanced model. In the **Design** tab under the **Count** group choose **Two Samples** and then **Negative Binomial - Ratio of Rates**.



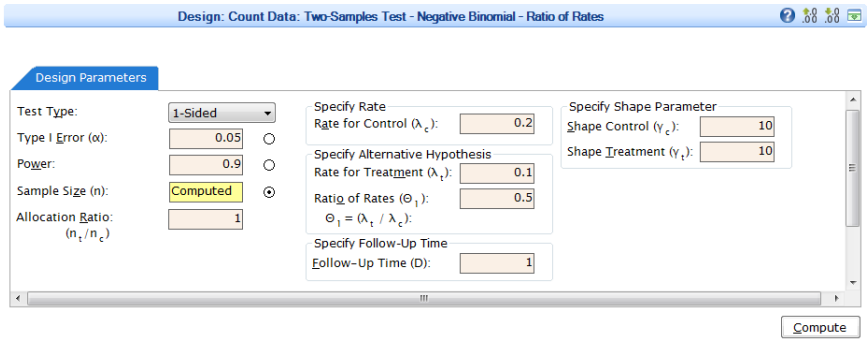

This will launch the following input window:



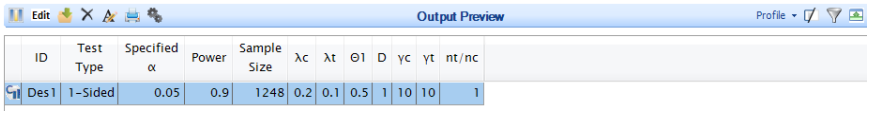

Enter the following design parameters:

Test Type: 1 sided
Type 1 Error $(\alpha)$: 0.05
Power: 0.9
Sample Size (n): Computed (select radio button)
Allocation Ratio $(n_t/n_c)$: 1
Rate for Control $(\lambda_c)$: 0.2
Rate for Treatment $(\lambda_t)$: 0.1

Ratio of Rates $\rho = (\lambda_t/\lambda_c)$: 0.5
Follow-up Time $(D)$: 1
Shape Control $(\gamma_c)$: 10
Shape Treatment $(\gamma_t)$: 10

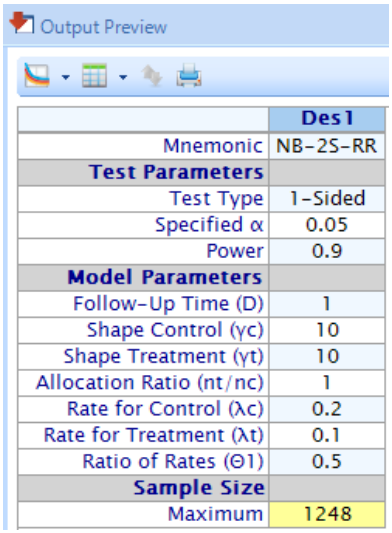



The Allocation Ratio $(n_t : n_c)$ describes the ratio of patients to each arm. For example, an allocation ratio of 3:1 indicates that 75% of the patients are randomized to the treatment arm as opposed to 25% to the control. Here we assume the same number of patients in both arms. Click **Compute**. The design is shown as a row in the **Output Preview** window:




The sample size required in order to achieve the desired 90% power is 1248 subjects. As is standard in East, this design has the default name **Des 1**. To see a summary of the output of this design, click anywhere in the row and then click the  icon in the Output Preview

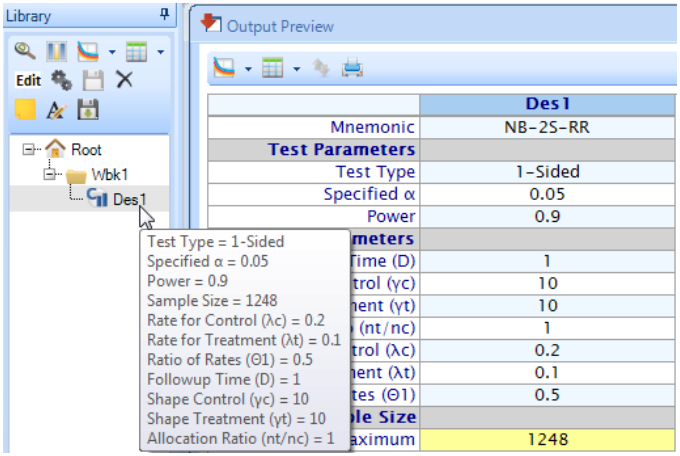
toolbar. The design details will be displayed in the upper pane, labeled **Output Summary**.




| | Des1 |
|---|---|
| Mnemonic | NB-2S-RR |
| **Test Parameters** | |
| Test Type | 1-Sided |
| Specified α | 0.05 |
| Power | 0.9 |
| **Model Parameters** | |
| Follow-Up Time (D) | 1 |
| Shape Control (γc) | 10 |
| Shape Treatment (γt) | 10 |
| Allocation Ratio (nt/nc) | 1 |
| Rate for Control (λc) | 0.2 |
| Rate for Treatment (λt) | 0.1 |
| Ratio of Rates (Θ1) | 0.5 |
| **Sample Size** | |
| Maximum | 1248 |

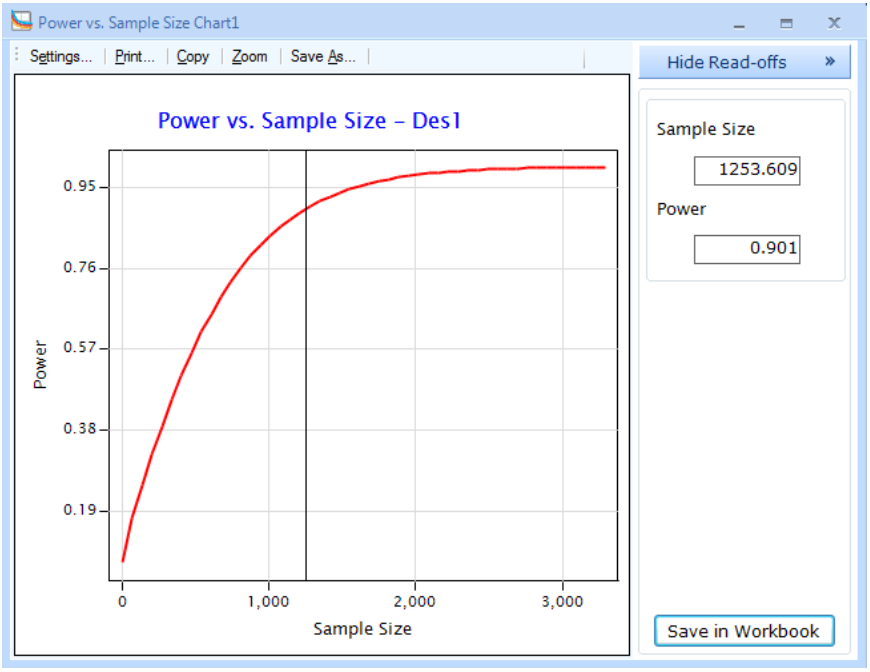
In the **Output Preview** toolbar, click icon to save this design **Des1** to workbook **Wbk1** in the **Library**. An alternative method to view design details is to hover the cursor over the node **Des1** in the **Library**. A tooltip will appear that summarizes the input parameters of the

design.



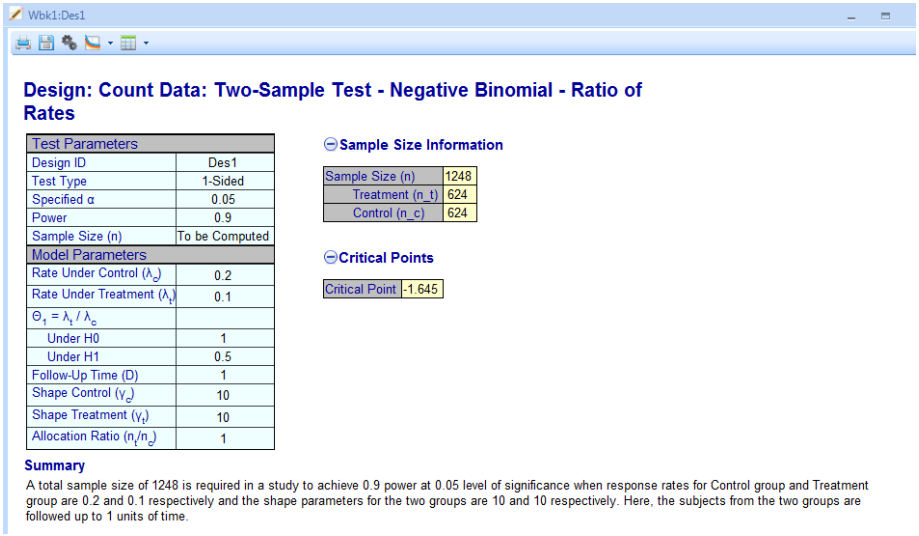

With the design **Des1** selected in the Library, click icon on the Library toolbar, and then click **Power vs. Sample Size**. The power curve for this design will be displayed. You can save this chart to the Library by clicking **Save inWorkbook**. Alternatively, you can export the chart in one of several image formats (e.g., Bitmap or JPEG) by clicking **Save As...** or **Export** into a

PowerPoint presentation.




Close the Power vs. Sample Size chart. To view all computed characteristics of this design,

select **Des1** in the **Library**, and click  icon.



Wbk1:Des1

**Design: Count Data: Two-Sample Test - Negative Binomial - Ratio of Rates**

| Test Parameters | |
|---|---|
| Design ID | Des1 |
| Test Type | 1-Sided |
| Specified α | 0.05 |
| Power | 0.9 |
| Sample Size (n) | To be Computed |
| **Model Parameters** | |
| Rate Under Control ($\lambda_c$) | 0.2 |
| Rate Under Treatment ($\lambda_t$) | 0.1 |
| $\Theta_1 = \lambda_t / \lambda_c$ | |
| Under H0 | 1 |
| Under H1 | 0.5 |
| Follow-Up Time (D) | 1 |
| Shape Control ($\gamma_c$) | 10 |
| Shape Treatment ($\gamma_t$) | 10 |
| Allocation Ratio ($n_t/n_c$) | 1 |

**Sample Size Information**

| | |
|---|---|
| Sample Size (n) | 1248 |
| Treatment (n_t) | 624 |
| Control (n_c) | 624 |

**Critical Points**

| | |
|---|---|
| Critical Point | -1.645 |

**Summary**

A total sample size of 1248 is required in a study to achieve 0.9 power at 0.05 level of significance when response rates for Control group and Treatment group are 0.2 and 0.1 respectively and the shape parameters for the two groups are 10 and 10 respectively. Here, the subjects from the two groups are followed up to 1 units of time.

In addition to the **Power vs. Sample** size chart and table, East also provides the efficacy boundary in the **Stopping Boundaries** chart and table.

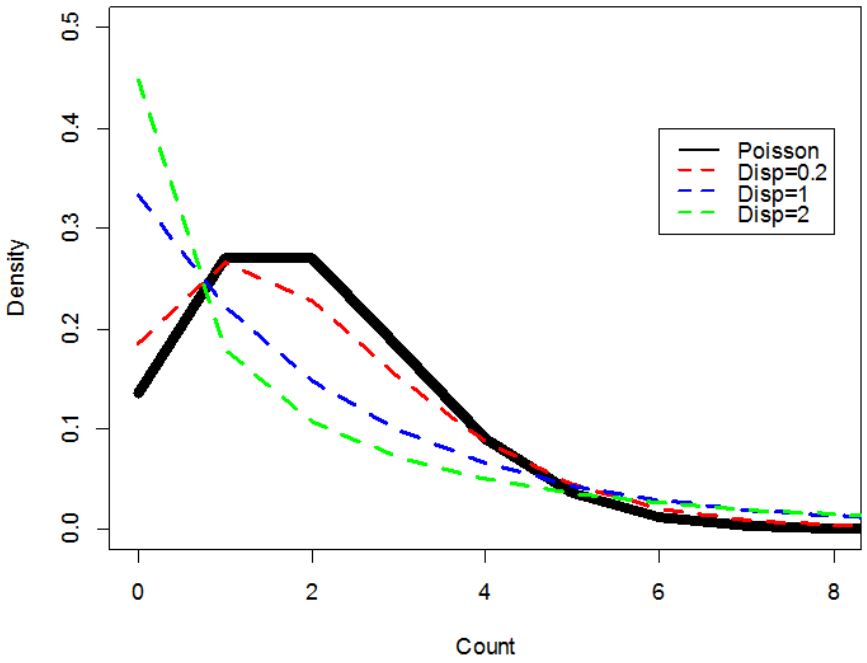
For a specific desired sample size, East allows the computation of either the **Type-1 error ($\alpha$)** or **Power** for a test. Using the **Design Input Output** window and methods as described above, simply enter the desired sample size and click **Compute** to calculate the resulting power of the test.

In addition to this example, consider the following illustration of the benefit of using the negative binomial model in clinical trials. In real life settings, the variance of count data observed between patients is typically higher than the observed mean. The negative binomial model accommodates between subject heterogeneity according to a Gamma distribution. For example:

Poisson: $Y \sim Poisson(\lambda)$

Negative Binomial: $Y_i \sim Poisson(\lambda k_i)$ where $k_i \sim Gamma(k)$

In the case of no overdispersion ($k = 0$) the negative binomial model reduces to the Poisson model. In the figure below, the Poisson and negative binomial models are displayed under various values of the dispersion parameter.




Assuming the above parameterization, the variance of the negative binomial model is $\lambda + k\lambda^2$. The inflation in variance is thus linear by the factor $1 + k * \lambda$ and dependent on the mean. Depending on the distributional assumption used and its impact on the variance, sample size and power can vary widely.

In multiple sclerosis (MS) patients, magnetic resonance imaging (MRI) is used as a marker of efficacy by means of serial counts of lesions appearing on the brain. Exacerbations rates as a primary endpoint are frequently used in MS as well as in chronic obstructive pulmonary disease (COPD) and asthma (*Keene et al. 2007*). Poisson regression could be considered, however this model would not address variability between patients, resulting in over

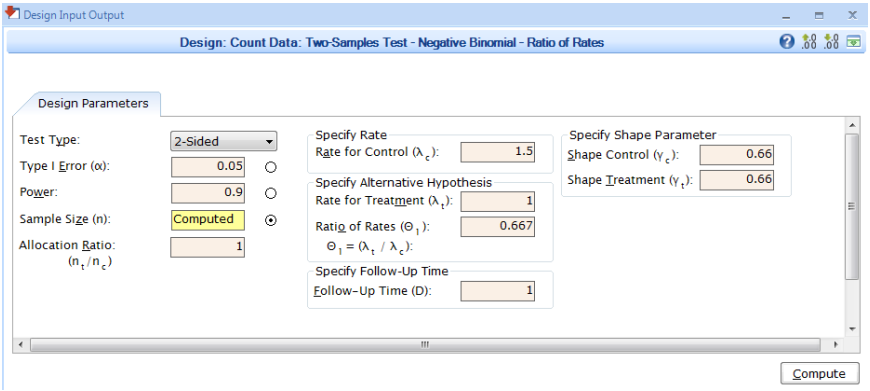
dispersion. The negative binomial model offers an alternative approach.

TRISTAN (*Keene et al. 2007*) was a double-blind, randomized study for COPD comparing the effects of the salmeterol/fluticasone propionate combination product (SFC) to salmeterol alone, fluticasone proprionate alone and placebo. Although the primary end-point was pre-bronchodilator FEV1, the number of exacerbations was an important secondary endpoint.

Suppose we are to design a new trial to be observed over a period of 1 to 2 years. The primary objective is the reduction of the rate of exacerbations, defined as a worsening of COPD symptoms that require treatment with antibiotics, cortisone or both, with the combination product SFC versus placebo. Based on the TRISTAN results, we aim to reduce the incidence of events by $33\%$. Suppose the exacerbation rate is 1.5 per year, and can expect to detect a rate of 1.0 in the combination group. Assume a 2-sided test with a $5\%$ significance level and $90\%$ power. Using a Poisson model, a total of 214 patients are needed to be enrolled in the study.

For the TRISTAN data, the estimate of the overdispersion parameter was 0.46 ($95\%$ CI: 0.34-0.60). Using a negative binomial model with overdispersion of $0.33, 0.66, 1$ and $2$, the

increase in sample size ranged from 299 to 726, respectively.

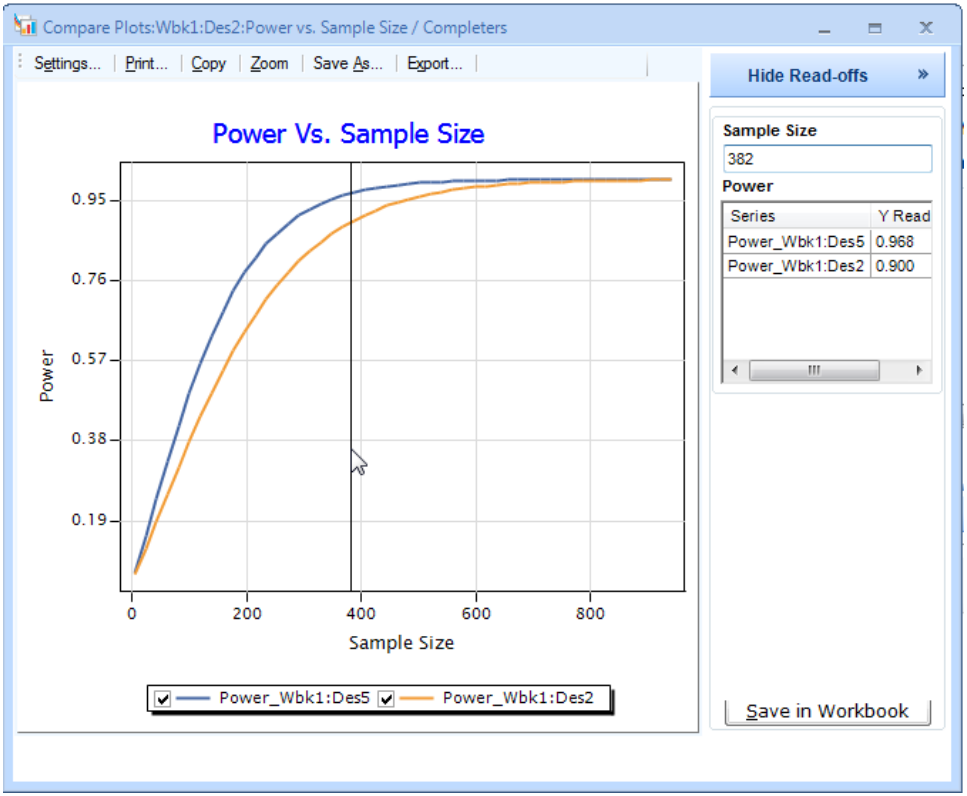



| | Des1 | Des2 | Des3 | Des4 |
|---|---|---|---|---|
| Mnemonic | NB-2S-RR | NB-2S-RR | NB-2S-RR | NB-2S-RR |
| **Test Parameters** | | | | |
| Test Type | 2-Sided | 2-Sided | 2-Sided | 2-Sided |
| Specified α | 0.05 | 0.05 | 0.05 | 0.05 |
| Power | 0.902 | 0.9 | 0.901 | 0.901 |
| **Model Parameters** | | | | |
| Follow-Up Time (D) | 1 | 1 | 1 | 1 |
| Shape Control (γc) | 0.33 | 0.66 | 1 | 2 |
| Shape Treatment (γt) | 0.33 | 0.66 | 1 | 2 |
| Allocation Ratio (nt/nc) | 1 | 1 | 1 | 1 |
| Rate for Control (λc) | 1.5 | 1.5 | 1.5 | 1.5 |
| Rate for Treatment (λt) | 1 | 1 | 1 | 1 |
| Ratio of Rates (Θ1) | 0.667 | 0.667 | 0.667 | 0.667 |
| **Sample Size** | | | | |
| Maximum | 299 | 382 | 470 | 726 |

Exacerbation rates are calculated as number of exacerbations divided by the length of time in treatment in years. EAST can be used to illustrate the impact of a one versus two year study by changing the follow-up duration.

Using a shape parameter of 0.66 for 382 patients, power is increased from $90\%$ to $97\%$ when follow-up time is doubled (see below). Alternatively, 277 patients observed for two years

would results in 90% power, which is the same as with 382 patients observed one year.




Output Preview

| | ID | Test Type | Specified α | Power | Sample Size | λc | λt | Θ1 | D | γc | γt | nt/nc |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Des2 | 2-Sided | 0.05 | 0.9 | 382 | 1.5 | 1 | 0.667 | 1 | 0.66 | 0.66 | 1 |
| | Des6 | 2-Sided | 0.05 | 0.902 | 277 | 1.5 | 1 | 0.667 | 2 | 0.66 | 0.66 | 1 |

Negative binomial models are increasing in popularity for medical research, and as the industry standard for trial design, East continues to evolve by incorporating sample size

methods for count data. These models allow the count to vary around the mean for groups of patients instead of the population means. Additionally, increased variability does lead to a larger test population; consequently the balance between power, sample size and duration of observation needs to be evaluated.

Reference: *Oliver N. Keene, Mark R. K. Jones, Peter W. Lane, Julie Anderson (2007). Analysis of exacerbation rates in asthma and chronic obstructive pulmonary disease: example from the TRISTAN study. Pharmaceutical statistics, 6, 89-97*

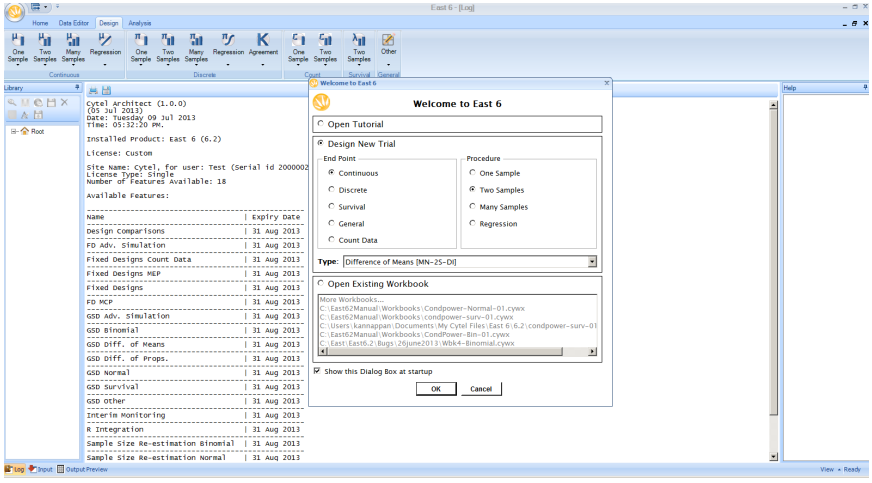
# Volume 5 Poisson and Negative Binomial Endpoints

# 10 Tutorial: Survival Endpoint

This tutorial introduces you to East 6, using examples for designing a clinical trial to compare survival in two groups. It is suggested that you go through the tutorial while you are at the computer, with East 6 running in it.

## 10.1 A Quick Feel of the Software

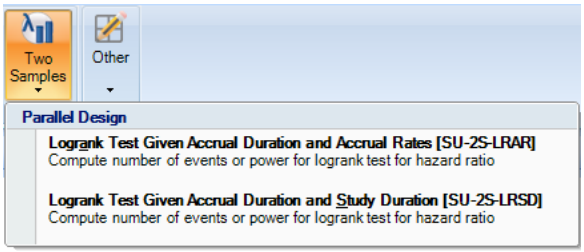
When you open East 6, the screen will look as shown below.



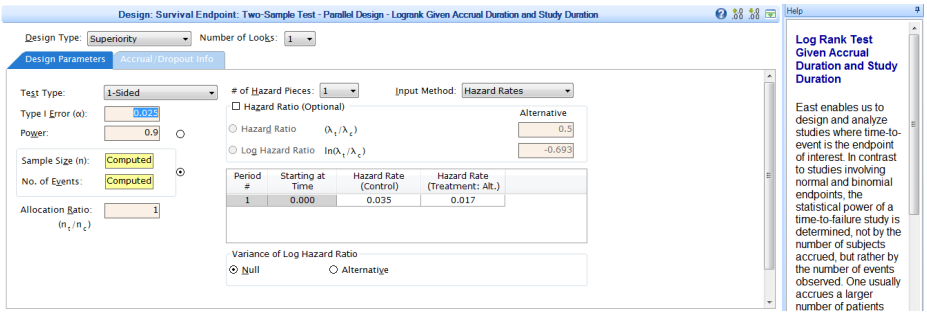

This is the **Welcome** screen of East 6 which enables us to open the tutorial file, select any design and open any existing workbook. Close this screen by clicking the Cancel button.

In the tabs bar at the top of the ribbon, Design tab is already selected. Each tab has its own ribbon. All the commands buttons under Design tab are displayed in its ribbon, with

suggestive icons. These commands have been grouped under the categories of Continuous, Discrete, Count, Survival and General. For this tutorial, let us explore the command **Two Samples** under **Survival** category. In East, we use the terms 'time to event' and 'survival' interchangeably. Click on **Two Samples**. You will see a list of action items, which are dialog box launchers.



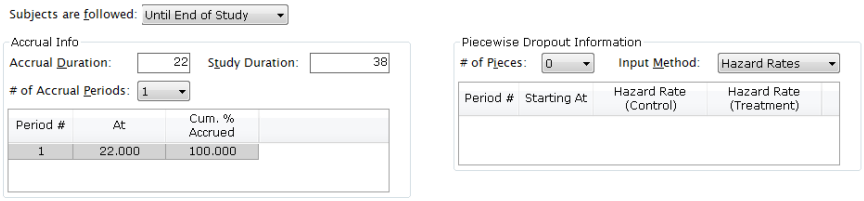

Click on **Logrank Test Given Accrual Duration and Study Duration**. You will get the following dialog box in the work area.




This dialog box is for computing Sample Size (n) and Number of Events. All the default input specifications under the tab Design Parameters are on display: Design Type=Superiority, Number of Looks=1, Test Type=1-Sided, Type-1 Error ($\alpha$)=0.025, Power (1-$\beta$)=0.9, Allocation Ratio ($n_t/n_c$)=1, # of Hazard Pieces=1, Input Method=Hazard Rates, Hazard Ratio ($\lambda_t/\lambda_c$)=0.5, Log Hazard Ratio ln($\lambda_t/\lambda_c$)=-0.693, Hazard Rate (Control)=0.0347, Hazard Rate (Treatment)=0.0173, and Variance of Log-Hazard Ratio=Null. There are two radio buttons in this dialog box, one at the side of Power (1-$\beta$) box and the second at the side of the combined

boxes for Sample Size (n) and Number of Events. By default, the latter radio button is selected indicating that the items against this radio button are to be computed using all other inputs. Similarly, if the first radio button is selected, then Power will be computed using all other inputs.

Now click on the tab Accrual/Dropout and you will see the following dialog box.




The default specifications in this dialog box are: Subjects are followed=Until End of Study, Accrual Duration=22, Study Duration=38, # of Accrual Periods=1, and no Dropouts. Now accept all the default specifications that are displayed for this single look design and be ready to compute the Sample Size (n) and the Number of Events for the design. Click Compute.

At the end of the computation, you will see the results appearing at the bottom of the screen, in the Output Preview pane, as shown below.

| ID | Design Type | No. of Looks | Test Type | Specified α | Power | nt/nc | Sample Size | Expected SS (H0) | Expected SS (H1) | Maximum Events | Exp. Events (H0) | Exp. Events (H1) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Des1 | Superiority | 1 | 1-Sided | 0.025 | 0.902 | 1 | 182 | 182 | 182 | 88 | 88 | 88 |

| Comm. Accr. (Dur.) | Exp. Accrual Duration (H0) | Exp. Accrual Duration (H1) | Hazard Ratio (Alt.) | Study Duration | Exp. Study Duration (H0) | Exp. Study Duration (H1) | Var (Log HR) | No. of Accrual Periods |
|---|---|---|---|---|---|---|---|---|
| 22 | 22 | 22 | 0.5 | 38 | 30.758 | 37.959 | Null | 1 |

This single row of output preview contains relevant details of all the inputs and the computed results for events and accruals. The maximum value for events is 88 and the committed accrual is 182 subjects. Since this is a fixed-look design, the expected events are same as the maximum required. Click anywhere in this row, and then click on the  icon to get a

detailed display in the upper pane of the screen as shown below.

| | Des1 |
|---|---|
| Mnemonic | SU-2S-LRSD |
| **Test Parameters** | |
| Design Type | Superiority |
| No. of Looks | 1 |
| Test Type | 1-Sided |
| Specified α | 0.025 |
| Power | 0.902 |
| **Model Parameters** | |
| Hazard Ratio (Alt.) | 0.5 |
| Var (Log HR) | Null |
| Allocation Ratio (nt/nc) | 1 |
| **Accrual & Dropout Parameters** | |
| Subjects are Followed | Until End of Study |
| No. of Accrual Periods | 1 |
| No. of Dropout Pieces | 0 |
| **Sample Size** | |
| Maximum | 182 |
| Expected Under H0 | 182 |
| Expected Under H1 | 182 |
| **Events** | |
| Maximum | 88 |
| Expected Under H0 | 88 |
| Expected Under H1 | 88 |
| **Study Duration** | |
| Maximum | 38 |
| Expected Under H0 | 30.758 |
| Expected Under H1 | 37.959 |
| **Accrual Duration** | |
| Maximum | 22 |
| Expected Under H0 | 22 |
| Expected Under H1 | 22 |

The contents of this output, displayed in the upper pane, are the same as what is contained in the output preview row for Design1 shown in the lower pane, but the upper pane display is easier to read and comprehend. The title of the upper pane display is **Output Summary**. This

is because, you can choose more than one design in the Output Preview pane and the display in the upper pane will show the details of all the selected designs in juxtaposed columns.

The discussion so far gives you a quick feel of the software for computing the required events and sample size for a single look survival design. We have not discussed about all the icons in the output preview pane or the library pane or the hidden Help pane in the screen. We will describe them taking an example for a group sequential design in the next section.

## 10.2 Group Sequential Design for a Survival Superiority Trial



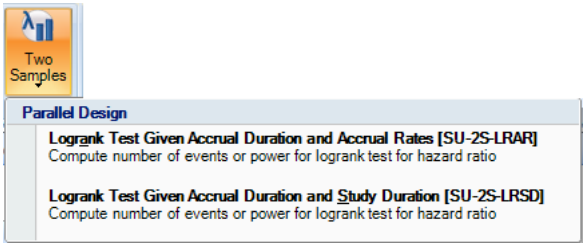
### 10.2.1 Background Information on the study

The randomized aldactone evaluation study (RALES) was a double-blind multicenter clinical trial of aldeosterone-recepter blocker vs. placebo published in New England Journal of Medicine (vol 341, 10, pages 709-717, 1999). This trial was open to patients with severe heart failure due to systolic left ventricular dysfunction. The Primary endpoint was all-causes mortality. The anticipated accrual rate was 960 patients/year. The mortality rate for the placebo group was 38%. The investigators wanted 90% power to detect a 17% reduction in the mortality hazard rate for the Aldactone group (from 0.38 to 0.3154) with $\alpha$ = 0.05, 2-sided test. Six DMC meetings were planned. The dropout rate in both the groups is expected to be 5% each year. The patient accrual period is planned to be 1.7 years and the total study duration to be 6 years.

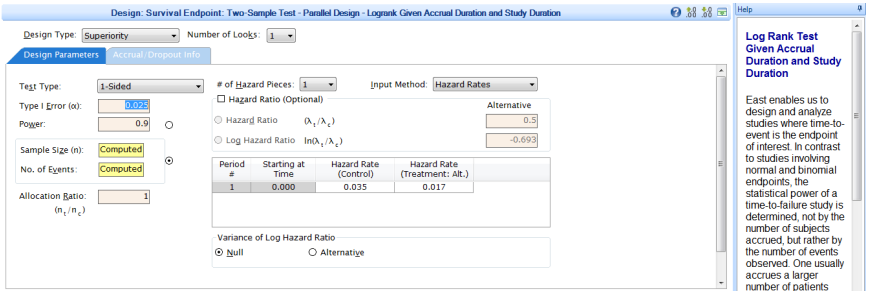
### 10.2.2 Creating the design in East

For our purpose, let us create our own design from the basic details of this study. Now start

afresh East. On the Design tab, click on **Two Samples** under **Survival** category. You will see a list of action items, which are dialog box launchers.

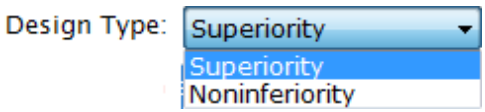



Click on the second option **Logrank Test Given Accrual Duration and Study Duration**. You will get the following dialog box in the work area.



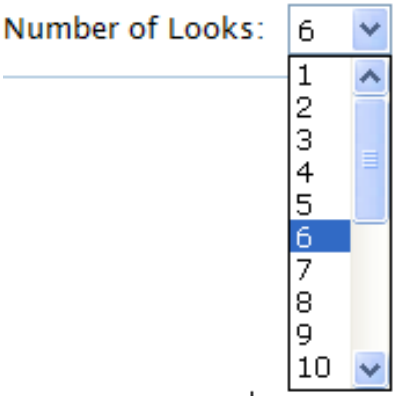

All the specifications you see in this dialog box are default values, which you will have to modify for the study under consideration.

Now, let the Design Type be Superiority.



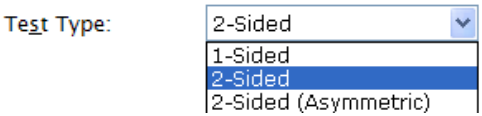

Next, enter 6 in the Number of Looks box. You can see the range of choices for the number of

looks is from 1 to 20.




Immediately after this selection, you will see a new tab **Boundary Info** added to the input dialog box. We will look into this tab, after you complete the filling of current tab **Design Parameters**.

Next, choose 2-Sided in the Test Type box.

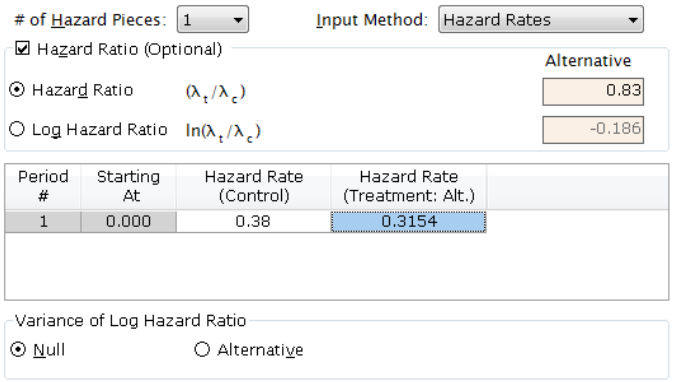



Next, enter 0.05 in the Type-1 Error ($\alpha$) box, and 0.9 in the Power box.

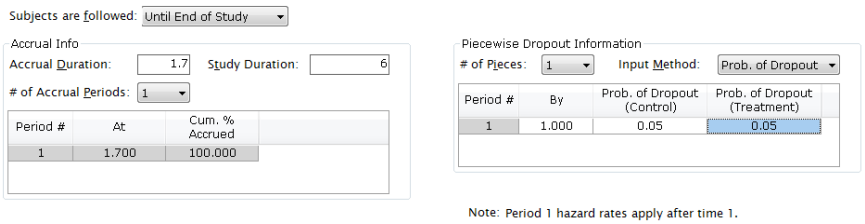

Next enter the specifications for survival parameters. Keep **# of Hazard Pieces** as 1. Click on the check box against Hazard Ratio and choose Hazard Rates as the Input Method. Enter 0.83 as the Hazard Ratio and 0.38 as the Hazard Rate (Control). East computes and displays the Hazard Rate (Treatment) as 0.3154. Keep the default choice of Null for Variance of Log-Hazard

## Chapter 10: *Tutorial: Survival Endpoint*

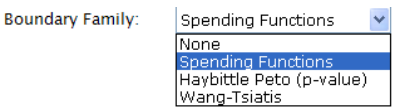
Ratio. Now the dialog box will look as shown below.



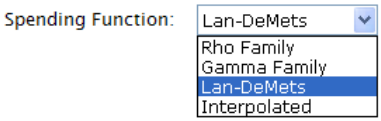

Next click the tab **Accrual/Dropout Info**. Keep the specification 'Until End of Study' for **Subjects are followed**. Enter 1.7 as **Accrual Duration** and 6 as **Study Duration**. Keep **# of Accrual Periods** as 1. Change the **# of Pieces** for dropouts to 1. Choose 'Prob. of Dropout' as the **Input Method** for entering information on dropouts. Enter 0.05 as probability of dropout at end of 1 year for both the groups. Now the dialog box will appear as shown below.




Now click on the **Boundary Info** tab. In the dialog box of this tab, you can specify stopping boundaries for efficacy or futility or both. For this trial, let us consider only Efficacy boundaries only. Choose 'Spending Functions' as the Efficacy Boundary Family.



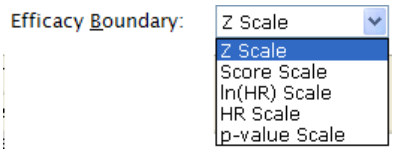
Choose 'Lan-DeMets' in the Spending Function box.




Choose 'OF' in the Parameter box.


Next, click the radio button near 'Equal' for Spacing of Looks.

Choose 'Z Scale' in the Efficacy Boundary Scale box.

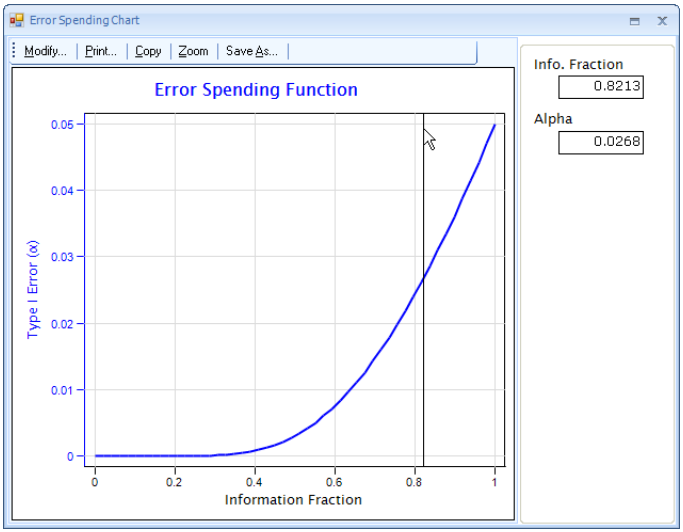


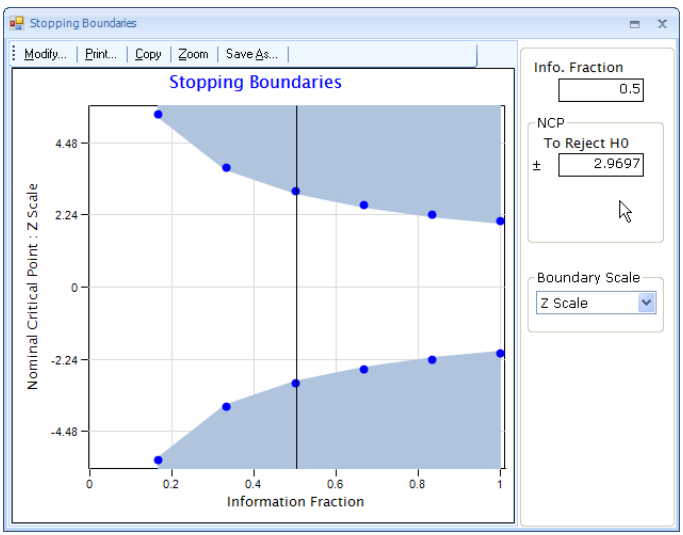

In the table below of look-wise details, the columns - Info Fraction, Cumulative Alpha Spent, and the upper and lower efficacy boundaries are computed and displayed as shown here. Scroll a little bit to see the sixth look details.

| Look # | Info. Fraction | Cum. α Spent | Efficacy Boundary | |
|---|---|---|---|---|
| | | | Upper | Lower |
| 1 | 0.167 | 0.0000 | 5.3666 | -5.3666 |
| 2 | 0.333 | 0.0002 | 3.7103 | -3.7103 |
| 3 | 0.500 | 0.0031 | 2.9697 | -2.9697 |
| 4 | 0.667 | 0.0121 | 2.5387 | -2.5387 |
| 5 | 0.833 | 0.0282 | 2.2522 | -2.2522 |

The two icons  and  represent buttons for Error Spending Function chart and Stopping Boundaries chart respectively. Click these two buttons one by one to see the

following charts.



### 10.2.3 Design Outputs

Now you have completed specifying all the inputs required for a group sequential trial design and you are ready to compute the required events and sample size or accruals for the trial. Click on the **Compute** button. After the computation is over, East will show in the Output Preview pane the following results:

| | ID | Design Type | No. of Looks | Test Type | Specified α | Power | nt/nc | Spacing of Looks | Efficacy Boundary | Sample Size | Expected SS (H0) | Expected SS (H1) | Maximum Events | Exp. Events (H0) | Exp. Events (H1) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Des1 | Superiority | 6 | 2-Sided | 0.05 | 0.9 | 1 | Equal | LD (OF) | 1646 | 1645.996 | 1645.986 | 1243 | 1233.984 | 903.595 |

| Comm. Accr. (Dur.) | Exp. Accrual Duration (H0) | Exp. Accrual Duration (H1) | Hazard Ratio (Alt.) | Study Duration | Exp. Study Duration (H0) | Exp. Study Duration (H1) | Var (Log HR) | No. of Accrual Periods |
|---|---|---|---|---|---|---|---|---|
| 1.7 | 1.7 | 1.7 | 0.83 | 6 | 5.354 | 3.725 | Null | 1 |

This single row of output preview contains relevant details of all the inputs and the computed results for events and accruals. The maximum required Events is computed as 1243 and the Committed Accrual to be 1646 subjects. The expected Events under H0 and H1 are estimated to be 1234 and 904 respectively. The expected Study Duration under H0 and H1 are 5.354 and 3.725 respectively.

Click anywhere in this Output Preview row and then click on  icon to get a summary in

the upper pane of the screen as shown below.

| | Des1 |
|---|---|
| Mnemonic | SU-2S-LRSD |
| **Test Parameters** | |
| Design Type | Superiority |
| No. of Looks | 6 |
| Test Type | 2-Sided |
| Specified α | 0.05 |
| Power | 0.9 |
| **Model Parameters** | |
| Hazard Ratio (Alt.) | 0.83 |
| Var (Log HR) | Null |
| Allocation Ratio (nt/nc) | 1 |
| **Boundary Parameters** | |
| Spacing of Looks | Equal |
| Efficacy Boundary | LD (OF) |
| **Accrual & Dropout Parameters** | |
| Subjects are Followed | Until End of Study |
| No. of Accrual Periods | 1 |
| No. of Dropout Pieces | 1 |
| **Sample Size** | |
| Maximum | 1646 |
| Expected Under H0 | 1645.996 |
| Expected Under H1 | 1645.986 |
| **Events** | |
| Maximum | 1243 |
| Expected Under H0 | 1233.984 |
| Expected Under H1 | 903.595 |
| **Study Duration** | |
| Maximum | 6 |
| Expected Under H0 | 5.354 |
| Expected Under H1 | 3.725 |
| **Accrual Duration** | |
| Maximum | 1.7 |
| Expected Under H0 | 1.7 |
| Expected Under H1 | 1.7 |

### *10.2.4 East icons explained*

In the 'Output Preview' pane, you see the following icons in the upper row.

The functions of the above icons are as indicated below. The tooltips also will indicate their functions.

 Output Summary(The output summary of selected design(s) will appear in the upper pane)

 Edit Design (The input dialog box of a selected design will appear in the upper pane)

 Save in Workbook (Save one or more selected designs in a workbook)

 Delete (Delete one or more selected designs)

 Rename (Rename Design names)

 Print (Print selected designs)

 Display Precision (Local Settings)

 Filter (Filter and select designs according to specified conditions)

 Show/Hide Columns (Show/Hide Columns of the designs in the Output Preview panel)

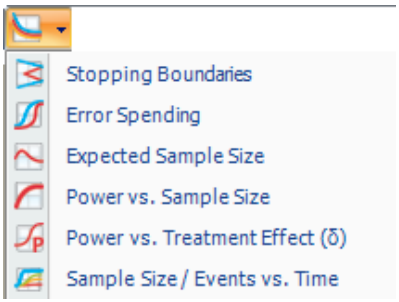
The following icons can be seen at the right end of Output Preview pane and Output Summary or Input/Output window respectively. Their functions are:

 Maximize Output Preview Pane

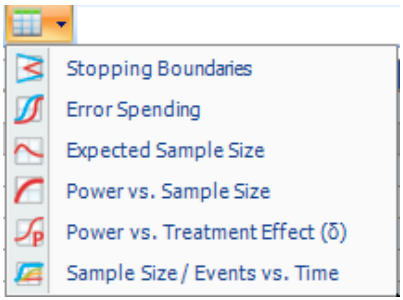
 Minimize Output Preview Pane

You may also notice a row of icons at the top of Output Summary window as shown below.

The first icon is for Plot (Plots of a selected design will appear in a pop-up window).

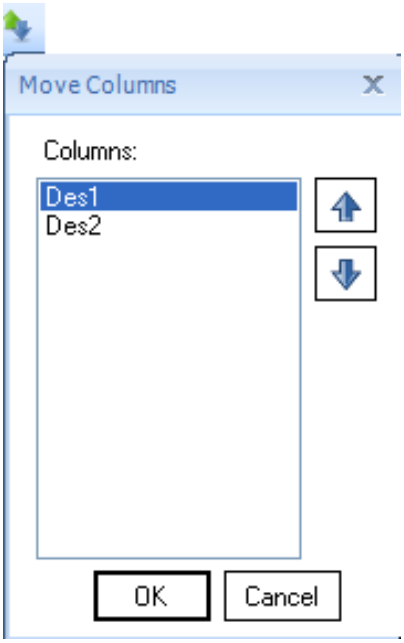



The second icon is for Show Tables (The data for different plots can be displayed in tabular form in pop-up windows).




If you have multiple designs in the output summary window, the third icon becomes active

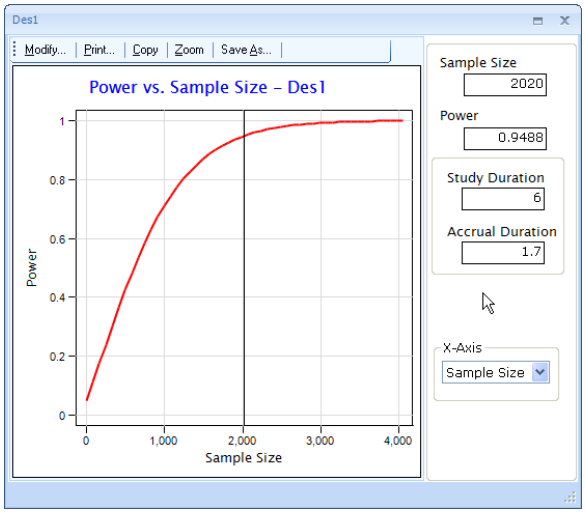
and can be used to move the order of those columns in the Output Summary.




The fourth icon is to print the Output Summary window.

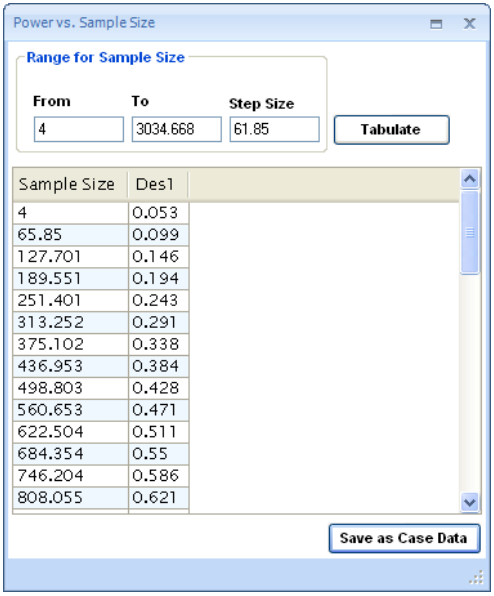
As an example, if you click Power vs. Sample Size under Plot icon, you will get the following

chart.




If you want to see the data underlying the above chart, click Show Table icon and click Power

vs. Sample Size. You will see the following table in a pop-up window.



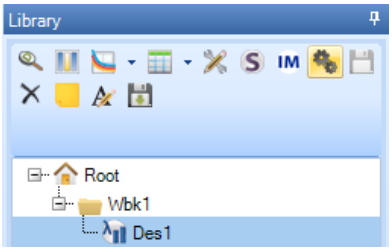

You can customize the format of the above table and also save it as case data in a workbook. You may experiment with all the above icon / buttons to understand their functions.

### 10.2.5 Saving created Designs in the library and hard disk

In the Output Preview pane, select one or more design rows and click the  icon,

The selected design(s) will then get added as a node(s) in the current workbook, as shown

below.




The above action only adds the design to the workbook node in the library and it is not saved in the hard disk. For saving in the hard disk, you may either save the entire workbook or only the design by right-clicking on the desired item and choosing save or save as options.

Here in the library also, you see rows of icons.


Some of these icons you have already seen. The functions of other icons are:

 Details (Details of a selected design will appear on the upper pane in the work area)

 Output Settings (Output Settings can be changed here)

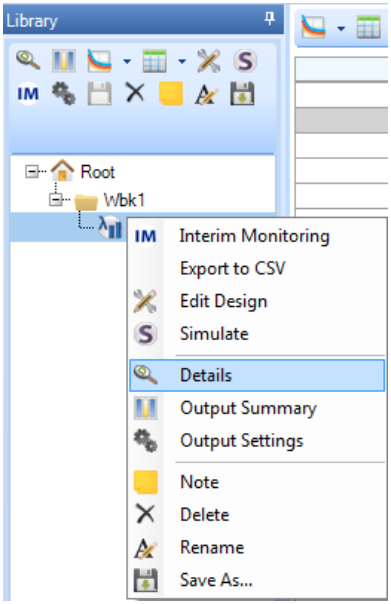
 Simulate (Start the simulation process for any selected design node)

 Interim Monitoring (Start the Interim Monitoring process for any selected design)

### *10.2.6 Displaying Detailed Output*

Select the design from the Library and click the  icon or Right-click on the Des1 node in

the library and click Details.



# Chapter 10: *Tutorial: Survival Endpoint*

You will see the detailed output of the design displayed in the Work area.

**Design: Survival Endpoint: Two-Sample Test - Parallel Design - Logrank Given Accrual Duration and Study Duration**

| Test Parameters | |
|---|---|
| Design ID | Des2 |
| Design Type | Superiority |
| Number of Looks | 6 |
| Test Type | 2-Sided |
| Specified α | 0.05 |
| Power | 0.9 |
| **Model Parameters** | |
| HR = $\lambda_t/\lambda_c$ | |
| Under H0 | 1 |
| Under H1 | 0.83 |
| δ = ln(HR) | -0.186 |
| Var (Log HR) | Null |
| Allocation Ratio ($n_t/n_c$) | 1 |
| **Boundary Parameters** | |
| Spacing of Looks | Equal |
| Efficacy Boundary | LD (OF) |
| **Accrual/Dropout Parameters** | |
| Accrual Duration | 1.7 |
| Max Study Duration | 6 |
| Dropout | Yes |

**Sample Size Information**

| | Control Arm | Treatment Arm | Total |
|---|---|---|---|
| Sample Size (n) | | | |
| Maximum | 822 | 824 | 1646 |
| Expected H1 | 822.993 | 822.993 | 1645.986 |
| Expected H0 | 822.998 | 822.998 | 1645.996 |
| Events (s) | | | |
| Maximum | 643 | 600 | 1243 |
| Expected H1 | 510.541 | 457.198 | 903.595 |
| Expected H0 | 618.84 | 618.84 | 1233.984 |
| Dropouts (d) | | | |
| Maximum | 87 | 98 | 185 |
| Expected H1 | 64.004 | 70.207 | 134.211 |
| Expected H0 | 83.39 | 83.39 | 166.781 |
| Maximum Information (I):310.75 | | | |

**Accrual and Study Duration**

| | Accrual Duration | Study Duration |
|---|---|---|
| Maximum | 1.7 | 5.99 |
| Expected H1 | 1.7 | 3.725 |
| Expected H0 | 1.7 | 5.354 |

**Stopping Boundaries: Look by Look**

| Look # | Info. Fraction (s/s_max) | Events (s) | Cumulative α Spent | Boundaries Efficacy Z | |
|---|---|---|---|---|---|
| | | | | Upper | Lower |
| 1 | 0.167 | 207 | 7.926E-8 | 5.369 | -5.369 |
| 2 | 0.333 | 414 | 2.057E-4 | 3.712 | -3.712 |
| 3 | 0.5 | 622 | 0.003 | 2.968 | -2.968 |
| 4 | 0.667 | 829 | 0.012 | 2.538 | -2.538 |
| 5 | 0.833 | 1036 | 0.028 | 2.252 | -2.252 |
| 6 | 1 | 1243 | 0.05 | 2.045 | -2.045 |

⊖Events, Sample Size, Dropouts, Pipeline and Analysis Times: Look by Look (Under H0)

| Look # | Info. Fraction (s/s_max) | Sample Size (n) | Events (s) | Dropouts (d) | Pipeline (n-s-d) | Analysis Time | Boundary Crossing Probability (Incremental) | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | Efficacy | |
| | | | | | | | Upper | Lower |
| 1 | 0.167 | 1112 | 207 | 28 | 877 | 1.148 | 3.963E-8 | 3.963E-8 |
| 2 | 0.333 | 1628 | 414 | 56 | 1158 | 1.681 | 1.028E-4 | 1.028E-4 |
| 3 | 0.5 | 1646 | 622 | 84 | 940 | 2.201 | 0.001 | 0.001 |
| 4 | 0.667 | 1646 | 829 | 112 | 705 | 2.867 | 0.005 | 0.005 |
| 5 | 0.833 | 1646 | 1036 | 140 | 470 | 3.807 | 0.008 | 0.008 |
| 6 | 1 | 1646 | 1243 | 168 | 235 | 5.413 | 0.011 | 0.011 |

⊖Events, Sample Size, Dropouts, Pipeline and Analysis Times: Look by Look (Under H1)

| Look # | Info. Fraction (s/s_max) | Sample Size (n) | Events (s) | Dropouts (d) | Pipeline (n-s-d) | Analysis Time | Boundary Crossing Probability (Incremental) | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | Efficacy | |
| | | | | | | | Upper | Lower |
| 1 | 0.167 | 1160 | 207 | 31 | 922 | 1.198 | 9.783E-12 | 2.808E-5 |
| 2 | 0.333 | 1646 | 414 | 62 | 1170 | 1.753 | 1.026E-8 | 0.035 |
| 3 | 0.5 | 1646 | 622 | 92 | 932 | 2.326 | 5.886E-8 | 0.226 |
| 4 | 0.667 | 1646 | 829 | 123 | 694 | 3.067 | 7.835E-8 | 0.303 |
| 5 | 0.833 | 1646 | 1036 | 154 | 456 | 4.125 | 5.829E-8 | 0.218 |
| 6 | 1 | 1646 | 1243 | 185 | 218 | 5.99 | 3.309E-8 | 0.119 |

⊕Survival Information : Hazard Rates

⊕Accrual Information

⊕Dropout Information : %Prob. of Dropout

Note: Period 1 hazard rates apply after time 1.

Variable Follow-Up Design: All subjects are followed until failure, drop out or end of study.

## 10.2.7 Comparing Multiple Designs

Click on Des1 row and then click Edit icon  . You will get the input dialog box in the upper pane. Change the Power value to 0.8 and then click Compute.

You will see now Des2 is created and a row added to Output Preview pane as shown below.

| | ID | Design Type | No. of Looks | Test Type | Specified α | Power | nt/nc | Spacing of Looks | Efficacy Boundary | Sample Size | Expected SS (H0) | Expected SS (H1) | Maximum Events | Exp. Events (H0) | Exp. Events (H1) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Des1 | Superiority | 6 | 2-Sided | 0.05 | 0.9 | 1 | Equal | LD (OF) | 1646 | 1645.996 | 1645.986 | 1243 | 1233.984 | 903.595 |
| | Des2 | Superiority | 6 | 2-Sided | 0.05 | 0.8 | 1 | Equal | LD (OF) | 1233 | 1232.997 | 1232.995 | 931 | 924.247 | 736.316 |

Click on Des1 row and then keeping Ctrl key pressed, click on Des2 row. Now both the rows will be selected. Next, click the Output Summary icon  .

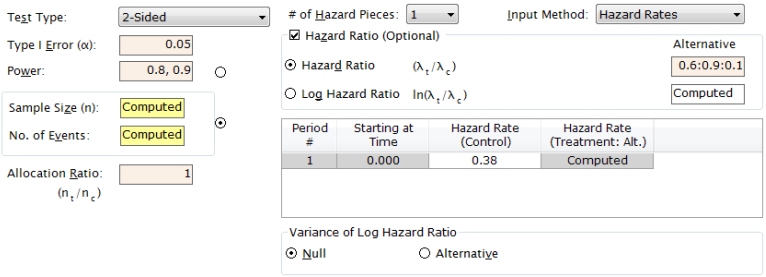
Now you will see the output details of these two designs displayed in the upper pane

Compare Designs in juxtaposed columns, as shown below.

| | Des1 | Des2 |
|---|---|---|
| Mnemonic | SU-2S-LRSD | SU-2S-LRSD |
| **Test Parameters** | | |
| Design Type | Superiority | Superiority |
| No. of Looks | 6 | 6 |
| Test Type | 2-Sided | 2-Sided |
| Specified α | 0.05 | 0.05 |
| Power | 0.9 | 0.8 |
| **Model Parameters** | | |
| Hazard Ratio (Alt.) | 0.83 | 0.83 |
| Var (Log HR) | Null | Null |
| Allocation Ratio (nt/nc) | 1 | 1 |
| **Boundary Parameters** | | |
| Spacing of Looks | Equal | Equal |
| Efficacy Boundary | LD (OF) | LD (OF) |
| **Accrual & Dropout Parameters** | | |
| Subjects are Followed | Until End of Study | Until End of Study |
| No. of Accrual Periods | 1 | 1 |
| No. of Dropout Pieces | 1 | 1 |
| **Sample Size** | | |
| Maximum | 1646 | 1233 |
| Expected Under H0 | 1645.996 | 1232.997 |
| Expected Under H1 | 1645.986 | 1232.995 |
| **Events** | | |
| Maximum | 1243 | 931 |
| Expected Under H0 | 1233.984 | 924.247 |
| Expected Under H1 | 903.595 | 736.316 |
| **Study Duration** | | |
| Maximum | 6 | 6 |
| Expected Under H0 | 5.354 | 5.352 |
| Expected Under H1 | 3.725 | 4.203 |
| **Accrual Duration** | | |
| Maximum | 1.7 | 1.7 |
| Expected Under H0 | 1.7 | 1.7 |
| Expected Under H1 | 1.7 | 1.7 |

In a similar way, East allows the user to easily create multiple designs by specifying a range of values for certain parameters in the design window. For example, in a survival trial the **Logrank Test given Accrual Duration and Study Duration** design allows the input of multiple key parameters at once to simultaneously create a number of different designs. For example, suppose in a multi-look study the user wants to generate designs for all combinations of the following parameter values: **Power** = 0.8 and 0.9, and **Hazard Ratio - Alternative** = 0.6, 0.7, 0.8 and 0.9. The number of combinations is 2 x 4 = 8. East creates all permutations using only a

single specification under the **Design Parameters** tab in the design window. As shown below, the values for **Power** are entered as a list of comma separated values, while the alternative hazard ratios are entered as a colon separated range of values, 0.6 to 0.9 in steps of 0.1.




East computes all 8 designs and displays them in the **Output Preview** window:

| | ID | Design Type | No. of Looks | Test Type | Specified α | Power | nt/nc | Spacing of Looks | Efficacy Boundary | Sample Size | Expected SS (H0) | Expected SS (H1) | Maximum Events | Exp. Events (H0) | Exp. Events (H1) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Des3 | Superiority | 6 | 2-Sided | 0.05 | 0.801 | 1 | Equal | LD (OF) | 177 | 176.998 | 176.999 | 124 | 123.101 | 98.054 |
| | Des4 | Superiority | 6 | 2-Sided | 0.05 | 0.8 | 1 | Equal | LD (OF) | 349 | 348.998 | 348.999 | 254 | 252.158 | 200.903 |
| | Des5 | Superiority | 6 | 2-Sided | 0.05 | 0.8 | 1 | Equal | LD (OF) | 866 | 865.997 | 865.997 | 649 | 644.291 | 513.311 |
| | Des6 | Superiority | 6 | 2-Sided | 0.05 | 0.8 | 1 | Equal | LD (OF) | 3789 | 3788.999 | 3788.985 | 2910 | 2888.9 | 2301.893 |
| | Des7 | Superiority | 6 | 2-Sided | 0.05 | 0.901 | 1 | Equal | LD (OF) | 237 | 236.997 | 236.998 | 166 | 164.797 | 120.548 |
| | Des8 | Superiority | 6 | 2-Sided | 0.05 | 0.901 | 1 | Equal | LD (OF) | 468 | 467.997 | 467.996 | 340 | 337.535 | 246.997 |
| | Des9 | Superiority | 6 | 2-Sided | 0.05 | 0.9 | 1 | Equal | LD (OF) | 1157 | 1156.996 | 1156.99 | 867 | 860.714 | 630.127 |
| | Des10 | Superiority | 6 | 2-Sided | 0.05 | 0.9 | 1 | Equal | LD (OF) | 5063 | 5062.998 | 5062.957 | 3888 | 3859.809 | 2826.252 |

East provides the capability to analyze multiple designs in ways that make comparisons between the designs visually simple and efficient. To illustrate this, a selection of a few of the above designs can be viewed simultaneously in both the **Output Summary** section as well as in the various tables and plots. The following is a subsection of the designs computed from the above example with differing values for number of looks, power and hazard ratio. Designs

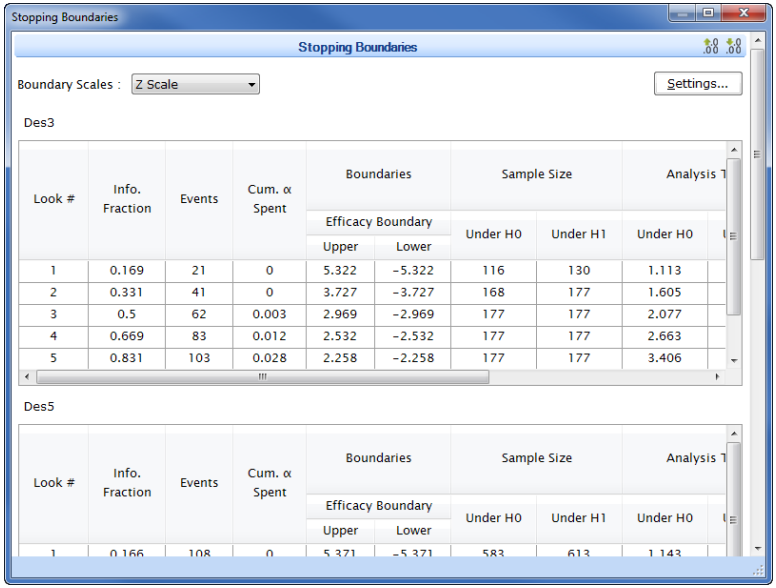
are displayed side by side, allowing details to be easily compared:

| | Des3 | Des5 | Des7 | Des10 |
|---|---|---|---|---|
| Mnemonic | SU-2S-LRSD | SU-2S-LRSD | SU-2S-LRSD | SU-2S-LRSD |
| **Test Parameters** | | | | |
| Design Type | Superiority | Superiority | Superiority | Superiority |
| No. of Looks | 6 | 6 | 6 | 6 |
| Test Type | 2-Sided | 2-Sided | 2-Sided | 2-Sided |
| Specified α | 0.05 | 0.05 | 0.05 | 0.05 |
| Power | 0.801 | 0.8 | 0.901 | 0.9 |
| **Model Parameters** | | | | |
| Hazard Ratio (Alt.) | 0.6 | 0.8 | 0.6 | 0.9 |
| Var (Log HR) | Null | Null | Null | Null |
| Allocation Ratio (nt/nc) | 1 | 1 | 1 | 1 |
| **Boundary Parameters** | | | | |
| Spacing of Looks | Equal | Equal | Equal | Equal |
| Efficacy Boundary | LD (OF) | LD (OF) | LD (OF) | LD (OF) |
| **Accrual & Dropout Parameters** | | | | |
| Subjects are Followed | Until End of Study | Until End of Study | Until End of Study | Until End of Study |
| No. of Accrual Periods | 1 | 1 | 1 | 1 |
| No. of Dropout Pieces | 1 | 1 | 1 | 1 |
| **Sample Size** | | | | |
| Maximum | 177 | 866 | 237 | 5063 |
| Expected Under H0 | 176.998 | 865.997 | 236.997 | 5062.998 |
| Expected Under H1 | 176.999 | 865.997 | 236.998 | 5062.957 |
| **Events** | | | | |
| Maximum | 124 | 649 | 166 | 3888 |
| Expected Under H0 | 123.101 | 644.291 | 164.797 | 3859.809 |
| Expected Under H1 | 98.054 | 513.311 | 120.548 | 2826.252 |
| **Study Duration** | | | | |
| Maximum | 6 | 6 | 6 | 6 |
| Expected Under H0 | 4.534 | 5.253 | 4.532 | 5.597 |
| Expected Under H1 | 4.257 | 4.214 | 3.786 | 3.703 |
| **Accrual Duration** | | | | |
| Maximum | 1.7 | 1.7 | 1.7 | 1.7 |
| Expected Under H0 | 1.7 | 1.7 | 1.7 | 1.7 |
| Expected Under H1 | 1.7 | 1.7 | 1.7 | 1.7 |

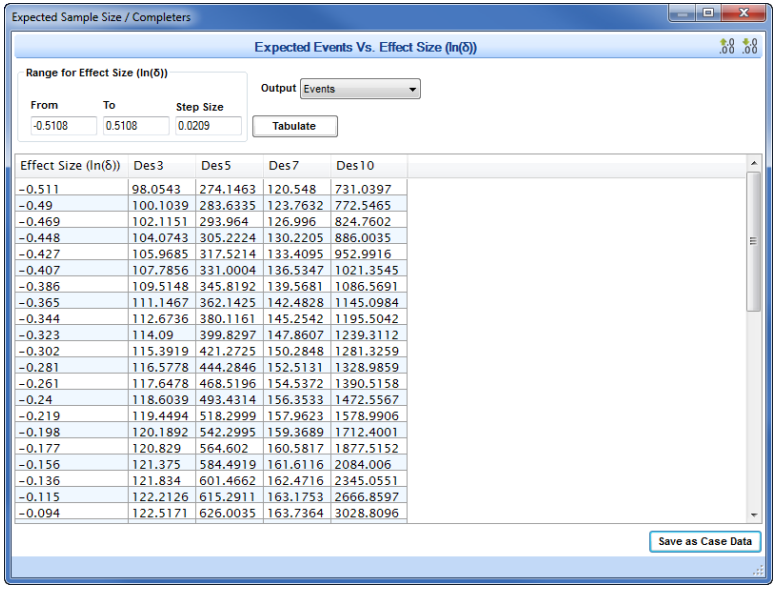
In addition East allows multiple designs to be viewed simultaneously either graphically or in tabular format: Notice that all the four designs in the Output Summary window are selected. Following figures compare these four designs in different formats.

# Chapter 10: *Tutorial: Survival Endpoint*

## Stopping Boundaries (table)



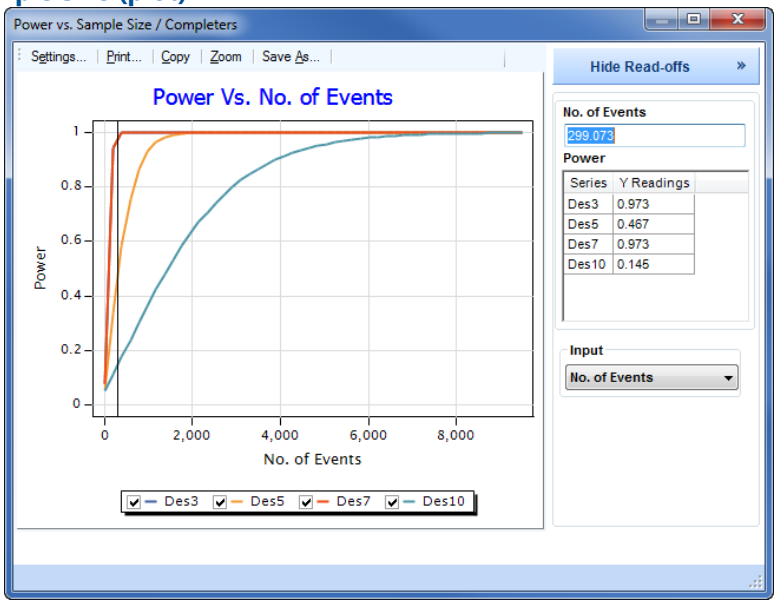

## Expected Sample Size (table)



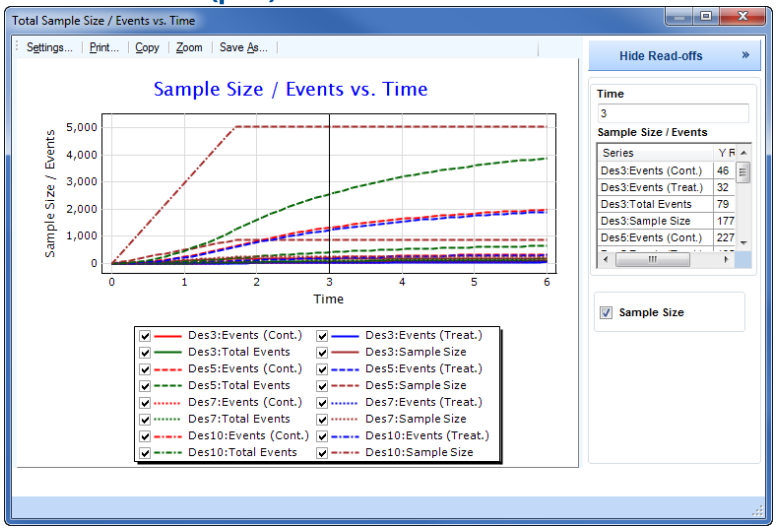

| Effect Size (ln(δ)) | Des3 | Des5 | Des7 | Des10 |
|---|---|---|---|---|
| -0.511 | 98.0543 | 274.1463 | 120.548 | 731.0397 |
| -0.49 | 100.1039 | 283.6335 | 123.7632 | 772.5465 |
| -0.469 | 102.1151 | 293.964 | 126.996 | 824.7602 |
| -0.448 | 104.0743 | 305.2224 | 130.2205 | 886.0035 |
| -0.427 | 105.9685 | 317.5214 | 133.4095 | 952.9916 |
| -0.407 | 107.7856 | 331.0004 | 136.5347 | 1021.3545 |
| -0.386 | 109.5148 | 345.8192 | 139.5681 | 1086.5691 |
| -0.365 | 111.1467 | 362.1425 | 142.4828 | 1145.0984 |
| -0.344 | 112.6736 | 380.1161 | 145.2542 | 1195.5042 |
| -0.323 | 114.09 | 399.8297 | 147.8607 | 1239.3112 |
| -0.302 | 115.3919 | 421.2725 | 150.2848 | 1281.3259 |
| -0.281 | 116.5778 | 444.2846 | 152.5131 | 1328.9859 |
| -0.261 | 117.6478 | 468.5196 | 154.5372 | 1390.5158 |
| -0.24 | 118.6039 | 493.4314 | 156.3533 | 1472.5567 |
| -0.219 | 119.4494 | 518.2999 | 157.9623 | 1578.9906 |
| -0.198 | 120.1892 | 542.2995 | 159.3689 | 1712.4001 |
| -0.177 | 120.829 | 564.602 | 160.5817 | 1877.5152 |
| -0.156 | 121.375 | 584.4919 | 161.6116 | 2084.006 |
| -0.136 | 121.834 | 601.4662 | 162.4716 | 2345.0551 |
| -0.115 | 122.2126 | 615.2911 | 163.1753 | 2666.8597 |
| -0.094 | 122.5171 | 626.0035 | 163.7364 | 3028.8096 |

# Chapter 10: *Tutorial: Survival Endpoint*

## Power vs. Sample Size (plot)




## Sample Size / Events vs. Time (plot)



This capability allows the user to explore a greater space of possibilities when determining the best choice of study design.

### 10.2.8 Events vs. Time plot

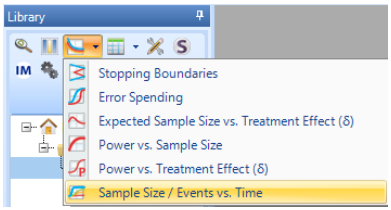
For survival studies, East provides a variety of charts and plots to visually validate and analyze the design. For example, the **Sample Size / Events vs. Time** plot allows the user to see the rate of increase in the number of events (control and treatment) over time (accrual duration, study duration). An additional feature of this particular chart is that a user can easily update key input parameters to determine how multiple different scenarios can directly impact a study. This provides significant benefits during the design phase, as the user can quickly examine how a variety of input values affect a study before the potentially lengthy task of simulation is employed.

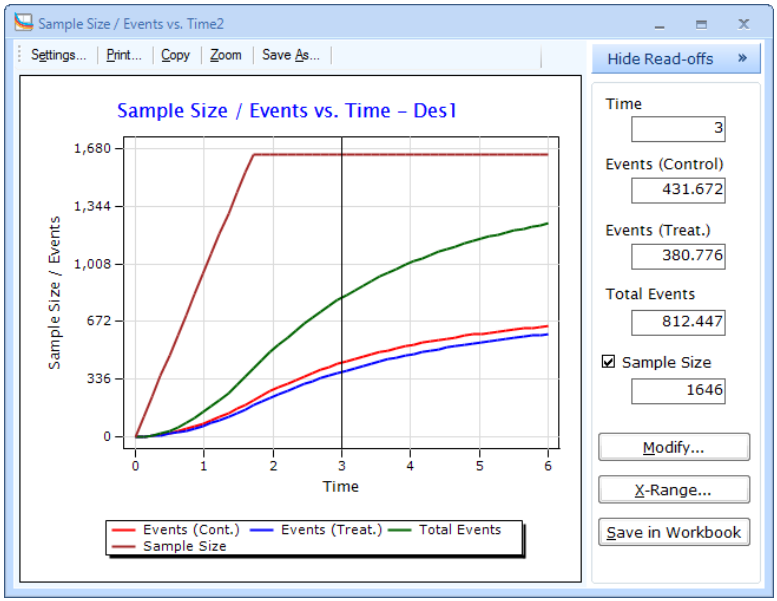
To illustrate this feature what follows is the example from the RALES study. For study details, refer to subsection **Background Information on the study** of this tutorial.

Currently there are ten designs in the **Output Preview** area. Select Des1 from them and save it to the current workbook. You may delete the remaining ones at this point.

To view the **Sample Size / Events vs. Time** plot, select the corresponding node in the **Library** and under the **Charts** icon choose **Sample Size / Events vs. Time**:

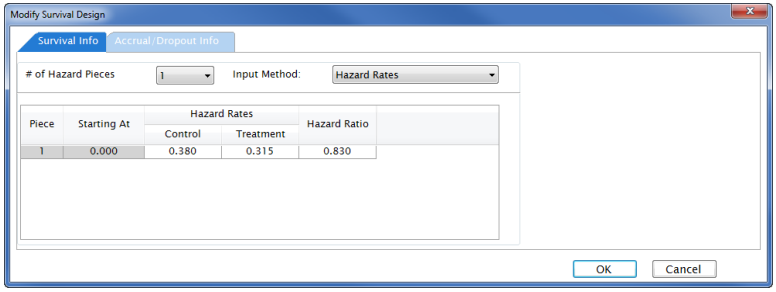
## Chapter 10: *Tutorial: Survival Endpoint*



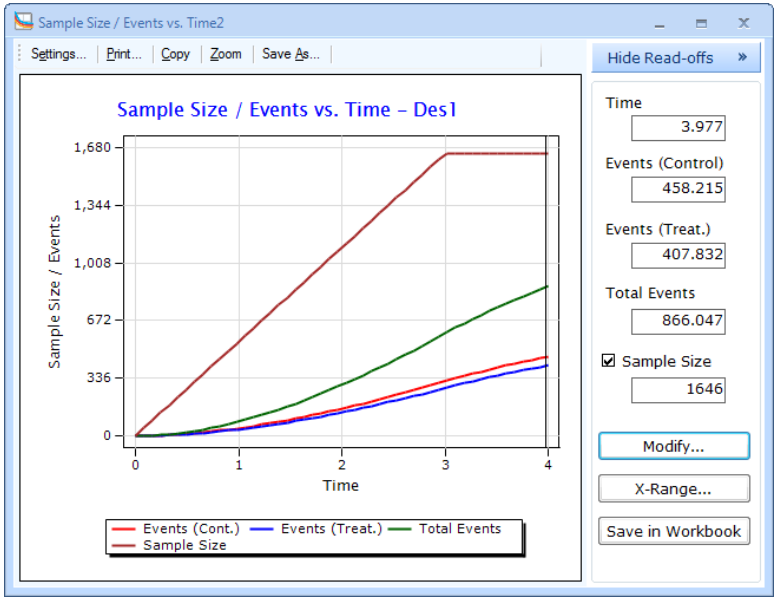




Survival parameters for this design can be edited directly through this chart by clicking the **Modify** button. The **Modify Survival Design** window is then displayed for the user to update

design parameters:




To illustrate the benefit of the modification feature, suppose at design time there is potential flexibility in the accrual and duration times for the study. To see how this may affect the number of subsequent events, modify the design to change the **Accrual Duration** to 3 and **Study Duration** to 4. Re-create the plot to view the effect of these new values on the shape and magnitude of the curves by clicking **OK**:

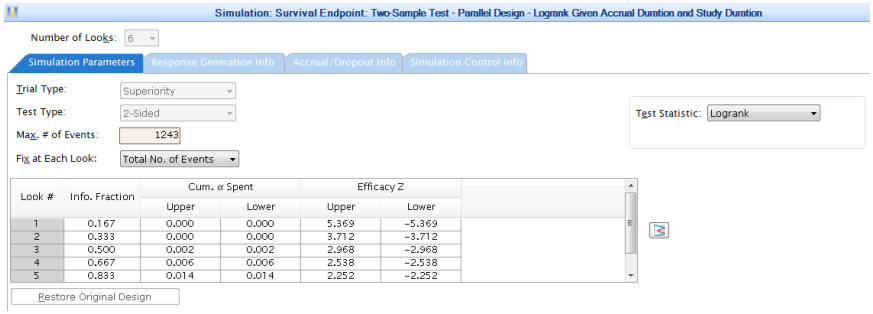


## Chapter 10: *Tutorial: Survival Endpoint*

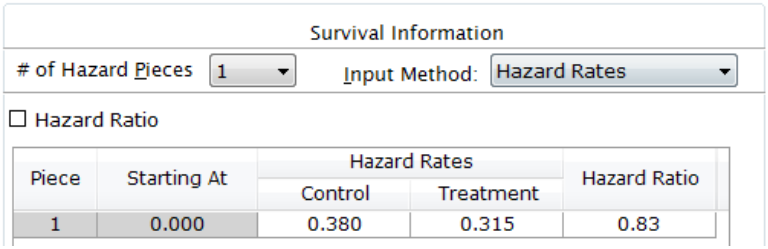
Similar steps can be taken to observe the effect of changing other parameter values on the number of events necessary to adequately power a study.

### 10.2.9 Simulation

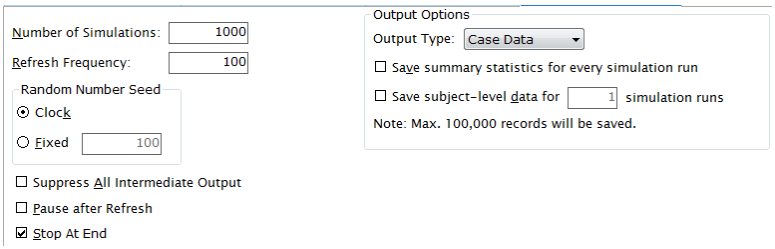
In the library, right-click on the node **Des1** and click **Simulate**. You will be presented with the following Simulation sheet.



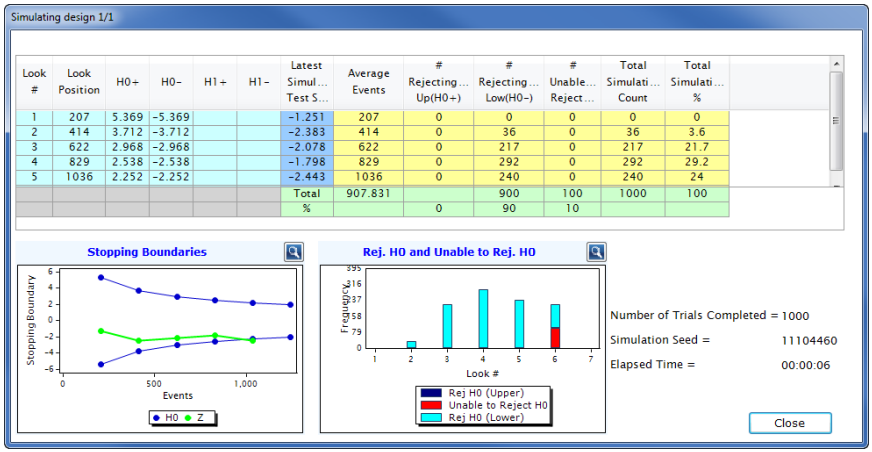

This sheet has four tabs - Simulation Parameters, Response Generation Info, Accrual/Dropout Info, and Simulation Control Info. Additionally, you can click **Include Options** and add some more tabs like Randomization Info or Stratification Info tab and so on. The first three tabs essentially contain the details of the parameters of the design. In the Simulation Control Info tab, you can specify the number of simulations to carry out and specify the file for storing simulation data. Let us first carry out 1000 simulations to check whether the design can reach the specified power of 90%. The Response Generation Info tab, by default, shows the hazard rates for control and treatment. We will use these values in our simulation.



In the Simulation Control tab, specify the number of simulations as 1000.




Let us keep the values in other tabs as they are and click **Simulate**. The progress of simulation process will appear in a temporary window as shown below.




This is the intermediate window showing the complete picture of simulations. Close this window after viewing it. You can see the complete simulation output in the details view. A new row, with the ID as Sim1, will be added in Output Preview.

| ID | Design Type | No. of Looks | Test Type | Specified α | Power | nt/nc | Spacing of Looks | Efficacy Boundary | Sample Size | Expected SS (H0) | Expected SS (H1) | Maximum Events | Exp. Events (H0) | Exp. Events (H1) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Des1 | Superiority | 6 | 2-Sided | 0.05 | 0.9 | 1 | Equal | LD (OF) | 1646 | 1645.996 | 1645.986 | 1243 | 1233.984 | 903.595 |
| Sim1 | Superiority | 6 | 2-Sided | | 0.9 | | User Specified | User Specified | 1646 | | | 1243 | | |

Click on Sim1 row and click the Output Summary icon  . You will see Simulation Output

summary appearing in the upper pane. It shows that the simulated power as 0.90, indicating that in 900 out of 1000 simulations the boundary was crossed.

| | Wbk1:Des1:Sim1 |
|---|---|
| Mnemonic | SU-2S-LRSD |
| **Test Parameters** | |
| Design Type | Superiority |
| Test Type | 2-Sided |
| Test Statistic | Logrank |
| Power | 0.9 |
| No. of Looks | 6 |
| **Model Parameters** | |
| No. of Hazard Pieces | 1 |
| **Boundary Parameters** | |
| Efficacy Boundary | User Specified |
| Spacing of Looks | User Specified |
| **Accrual & Dropout Parameters** | |
| Followup Duration | Until End of Study |
| No. of Accrual Periods | 1 |
| **Sample Size** | |
| Maximum | 1646 |
| **Events** | |
| Maximum | 1243 |
| **Simulation Results (Overall)** | |
| Average Study Duration | 3.741 |
| Average Sample Size | 1645.967 |
| Average Events | 907.831 |

You can save Sim1 as a node in the workbook. If you right-click on this node and then click Details, you will see the complete details of simulation appearing in the work area. Here is a

part of it.

**Simulation: Survival Endpoint: Two-Sample Test - Parallel Design - Logrank Given Accrual Duration and Study Duration**

| Simulation Parameters | |
|---|---|
| Simulation ID | Sim1 |
| Design Type | Superiority |
| Number of Looks | 6 |
| Test Type | 2-Sided |
| Sample Size (n) | 1646 |
| Fix at Each Look | Total No. of Events |
| Test Statistic | Logrank |
| Average Events | 907.831 |
| Total Accrual Duration | 1.7 |
| Avg. Power at Termination | 0.9 |
| **Simulation Control Parameters** | |
| Starting Seed | Clock |
| Number of Simulations | 1000 |

⊖Average Sample Size, Dropouts and Look Times

| Look # | Average Sample Size | Average Events | | Average Dropouts | | Average Look Time | Average Follow up |
|---|---|---|---|---|---|---|---|
| | | Control | Treatment | Control | Treatment | | |
| 1 | 1161.186 | 112.317 | 94.683 | 15.162 | 15.501 | 1.199 | 0.514 |
| 2 | 1644.098 | 223.105 | 190.895 | 30.077 | 31.235 | 1.755 | 0.727 |
| 3 | 1646 | 332.552 | 289.448 | 45.083 | 47.01 | 2.326 | 1.09 |
| 4 | 1646 | 436.799 | 392.201 | 59.967 | 63.001 | 3.069 | 1.454 |
| 5 | 1646 | 538.154 | 497.846 | 75.235 | 78.8 | 4.129 | 1.82 |
| 6 | 1646 | 636.921 | 606.079 | 90.102 | 94.479 | 5.981 | 2.182 |
| Average | 1645.967 | 480.572 | 427.259 | 65.096 | 69.339 | 3.741 | 1.593 |

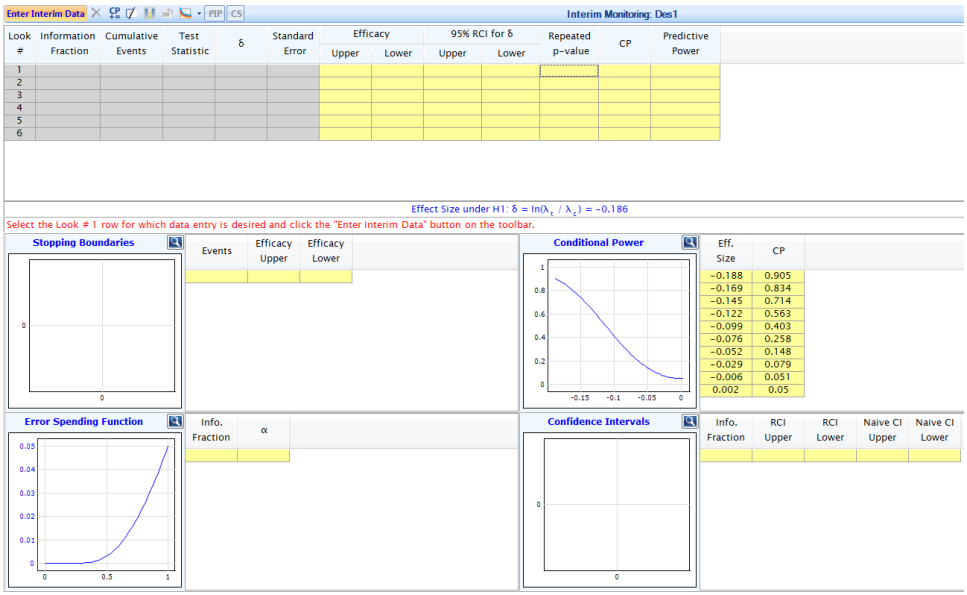
⊖Simulation Boundaries and Boundary Crossing Probabilities

| Look # | Events | Boundaries | | Stopping For | | Total Simulations | |
|---|---|---|---|---|---|---|---|
| | | Efficacy | | | | | |
| | | Upper | Lower | Upper Efficacy | Lower Efficacy | Count | % |
| 1 | 207 | 5.369 | -5.369 | 0 | 0 | 0 | 0.000% |
| 2 | 414 | 3.712 | -3.712 | 0 | 36 | 36 | 3.600% |
| 3 | 622 | 2.968 | -2.968 | 0 | 217 | 217 | 21.700% |
| 4 | 829 | 2.538 | -2.538 | 0 | 292 | 292 | 29.200% |
| 5 | 1036 | 2.252 | -2.252 | 0 | 240 | 240 | 24.000% |
| 6 | 1243 | 2.045 | -2.045 | 0 | 115 | 215 | 21.500% |
| Total | | | | 0 | 900 | 1000 | |
| % | | | | 0.000% | 90.000% | | |

## *10.2.10 Interim Monitoring*

Click Des1 node under workbook wbk1 and click the IM icon. Alternatively, you can right-click the Des1 node and select the item **Interim Monitoring**. In either case, you will see

the IM dashboard appearing as shown below.



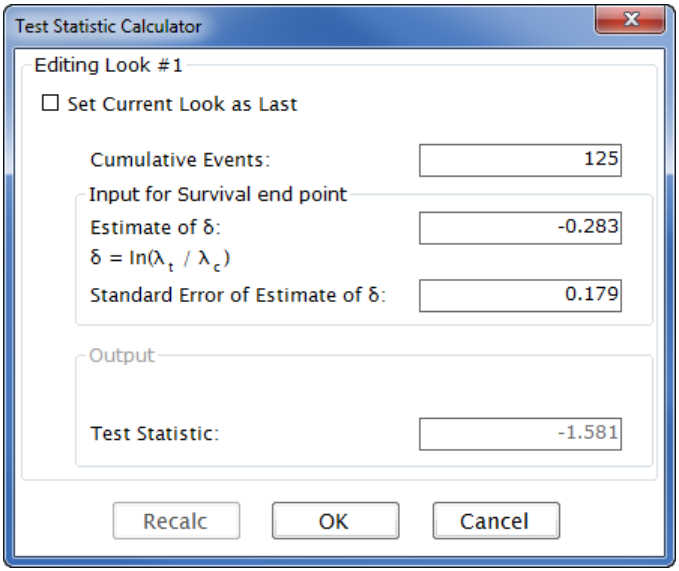

In the top row, you see a few icons. For now, we will discuss only the first icon Enter Interim Data which represents Test Statistic Calculator. Using this calculator, you will enter the details of interim look data analysis results into the IM dashboard.

Suppose we have the following data used by the Data Monitoring Committee during the first 5 looks of interim monitoring.

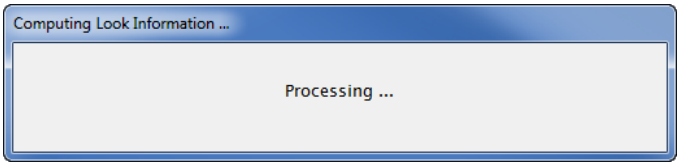
| Date | Total Deaths | $\hat{\delta}$ | SE($\hat{\delta}$) | Z-Statistic |
|---|---|---|---|---|
| Aug 96 | 125 | -0.283 | 0.179 | -1.581 |
| Mar 97 | 299 | -0.195 | 0.116 | -1.681 |
| Aug 97 | 423 | -0.248 | 0.097 | -2.557 |
| Mar 98 | 545 | -0.259 | 0.086 | -3.012 |
| Aug 98 | 670 | -0.290 | 0.077 | -3.766 |

The first look was taken at 125 events and the analysis of the data showed the value of $\delta$=

-0.283 and SE($\delta$)=0.179. First, click the blank row in the IM Dashboard and then click the Enter Interim Data icon. Now you can enter the first analysis results into the TS calculator and click Recalc. The Test Statistic value will be computed and the TS calculator will appear as shown below.




Now click on the button 'OK' to get the first look details into IM Dashboard. The following message will appear that some required computations are being carried out.

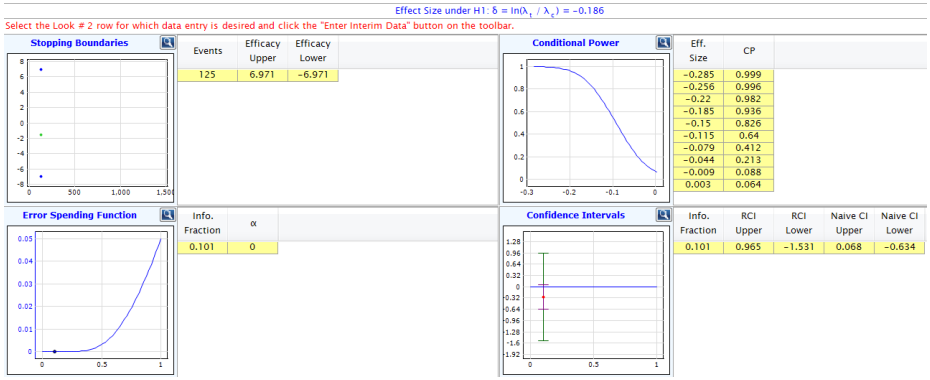



After the computations are over, the output for the first look will appear in the IM Dashboard

as shown below.

| Look # | Information Fraction | Cumulative Events | Test Statistic | δ | Standard Error | Efficacy | | 95% RCI for δ | | Repeated p-value | CP | Predictive Power |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | Upper | Lower | Upper | Lower | | | |
| 1 | 0.101 | 125 | -1.581 | -0.283 | 0.179 | 6.971 | -6.971 | 0.965 | -1.531 | 1 | 0.999 | 0.853 |
| 2 | | | | | | | | | | | | |
| 3 | | | | | | | | | | | | |
| 4 | | | | | | | | | | | | |
| 5 | | | | | | | | | | | | |
| 6 | | | | | | | | | | | | |

For the first look at total number of events, 125, the Information Fraction works out to be 0.101. The efficacy boundaries for this information fraction are newly computed. The Repeated 95% Confidence Interval limits and Repeated p-value are computed and displayed. You may also see that the charts at the bottom of the IM Dashboard have been updated with relevant details appearing on the side.

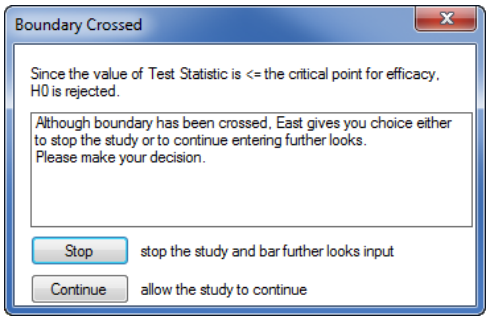



In a similar way, enter the interim analysis results for the next 3 looks in the IM Dashboard. Now the IM Dashboard will look like this:

| Look # | Information Fraction | Cumulative Events | Test Statistic | δ | Standard Error | Efficacy | | 95% RCI for δ | | Repeat ... p-value | CP | Predictive Power |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | Upper | Lower | Upper | Lower | | | |
| 1 | 0.101 | 125 | -1.581 | -0.283 | 0.179 | 6.971 | -6.971 | 0.965 | -1.531 | 1 | 0.999 | 0.853 |
| 2 | 0.241 | 299 | -1.681 | -0.195 | 0.116 | 4.423 | -4.423 | 0.318 | -0.708 | 0.66 | 0.95 | 0.795 |
| 3 | 0.34 | 423 | -2.557 | -0.248 | 0.097 | 3.672 | -3.672 | 0.108 | -0.604 | 0.212 | 0.999 | 0.962 |
| 4 | 0.438 | 545 | -3.012 | -0.259 | 0.086 | 3.206 | -3.206 | 0.017 | -0.535 | 0.069 | 1 | 0.993 |
| 5 | | | | | | | | | | | | |
| 6 | | | | | | | | | | | | |

Now again click on the fifth row in IM Dashboard, enter the fifth look results into the Test Statistic Calculator and click OK. This time, the boundary is crossed. A message window

appears as shown below.




Click Stop and you will see the details of all the looks in the IM Dashboard as shown below.

| Look # | Information Fraction | Cumulative Events | Test Statistic | δ | Standard Error | Efficacy | | 95% RCI for δ | | Repeated p-value | CP | Predictive Power |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | Upper | Lower | Upper | Lower | | | |
| 1 | 0.101 | 125 | -1.581 | -0.283 | 0.179 | 6.971 | -6.971 | 0.965 | -1.531 | 1 | 0.999 | 0.853 |
| 2 | 0.241 | 299 | 1.681 | 0.195 | 0.116 | 4.423 | -4.423 | 0.708 | -0.318 | 0.66 | 0.95 | 0.795 |
| 3 | 0.34 | 423 | -2.557 | -0.248 | 0.097 | 3.672 | -3.672 | 0.108 | -0.604 | 0.212 | 0.999 | 0.962 |
| 4 | 0.438 | 545 | -3.012 | -0.259 | 0.086 | 3.206 | -3.206 | 0.017 | -0.535 | 0.069 | 1 | 0.993 |
| 5 | 0.539 | 670 | -3.766 | -0.29 | 0.077 | 2.872 | -2.872 | -0.069 | -0.511 | 0.008 | NA | NA |

The final Adjusted Inference output also appears as displayed below.

| Final Inference | |
|---|---|
| Final Outputs at Look # | 5 |
| Adj. p-value | 0.001 |
| Adj. Pt. Est. for δ | -0.266 |
| **Adj. 95% CI for δ** | |
| Upper Confidence Bound | -0.104 |
| Lower Confidence Bound | -0.424 |
| Post-Hoc Power | |

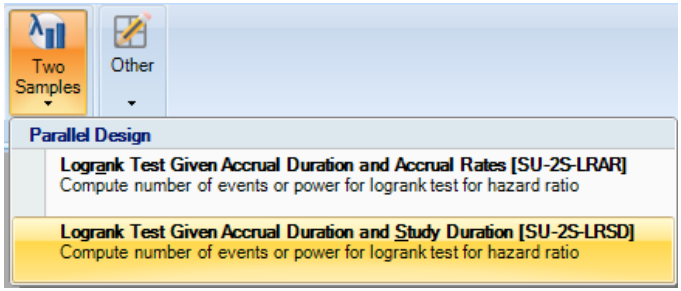
One important point to note here is that this study got over almost about 2 years ahead of planned schedule, because of the very favorable interim analysis results.

This completes the Interim Monitoring exercise in this trial.

## 10.3 User Defined R Function

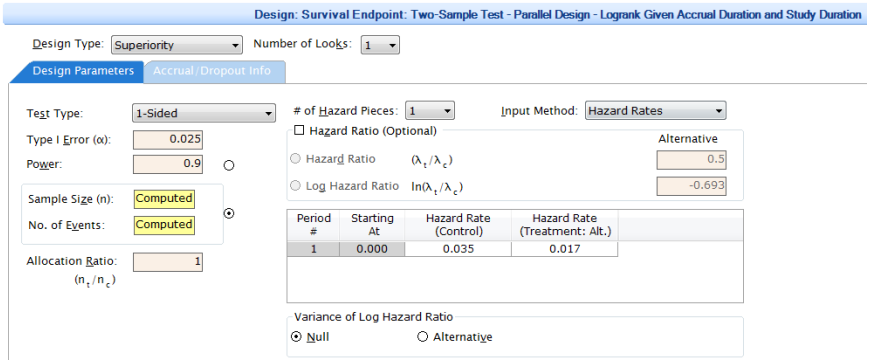
East allows you to customize simulations by inserting user-defined R functions for one or more of the following tasks: generate response, compute test statistic, randomize subjects, generate arrival times, and generate dropout information. The R functionality for arrivals and dropouts will be available only if you have entered such information at the design stage. Although the R functions are also available for all normal and binomial endpoints, we will illustrate this functionality for a time-to-event endpoint. Specifically, we will use an R function to generate Weibull survival responses.

Start East afresh. On the **Design** tab, click **Survival: Two Samples** and then **Logrank Test Given Accrual Duration and Study Duration**.

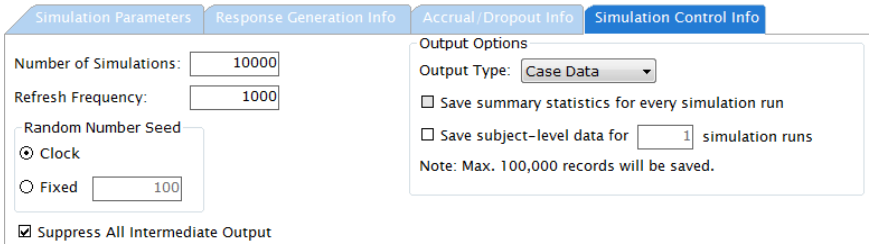



Choose the design parameters as shown below. In particular, select a one sided test with

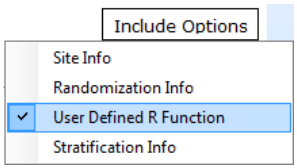
type-1 error of $\alpha = 0.025$.




Click **Compute** and save this design (Des1) to the **Library**. Right-click Des1 in the **Library** and click **Simulate**. In the **Simulation Control Info** tab, check the box for **`Suppress All Intermediate Input`**. Type $10000$ for **Number of Simulations** and select **`Clock`** for **Random Number Seed**.

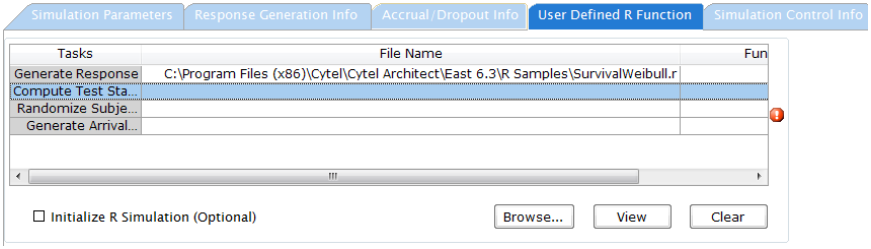



In the top right-hand corner for the input window, click **Include Options**, and then click **User Defined R Function**.

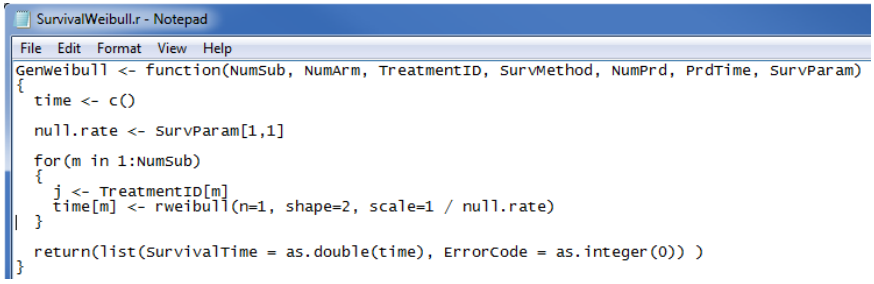


Go to the **User Defined R Function** tab. For now, leave the box **Initialize R simulation (optional)** unchecked. This optional task can be used to load required libraries, set seeds for simulations, and initialize global variables.

Select the row for **Generate Response**, click **Browse...**, and navigate to the folder containing your R file. Select the file and click **Open**. The path should now be displayed under **File Name**.




Click **View** to open a notepad application to view your R file. In this example, we are generating survival responses for both control and treatment arms from a Weibull with shape parameter $= 2$ (i.e. exponential), with the same hazard rate in both arms. This sample file is available in the folder named **`R Samples`** under installation directory of East 6.

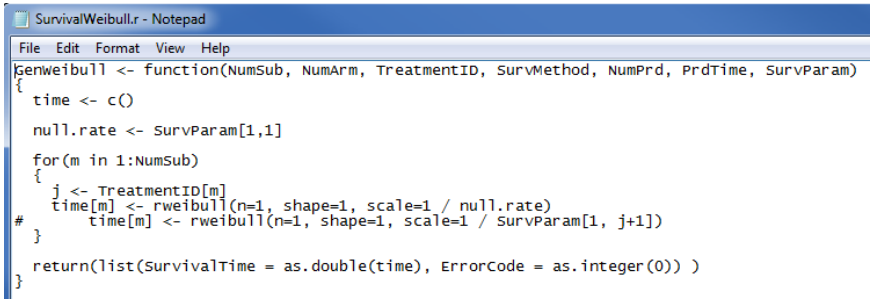



Copy the function name (in this case *GenWeibull*) and paste it into the cell for **Function Name**.

Save and close the R file, and click **Simulate**.

| Tasks | File Name | Function Name |
|---|---|---|
| Generate Response | C:\Program Files (x86)\Cytel\Cytel A... | GenWeibull |
| Compute Test Sta... | | |
| Randomize Subje... | | |
| Generate Arrival... | | |

Return to the tab for **User Defined R Function**, select the **Generate Response** row, and click **View**. In the R function, change the shape parameter $= 1$, to generate responses from a Weibull distribution with increasing hazards. Save and close the R file, and click **Simulate**. You may have to save this file on some other location.




Select both simulations (Sim1 and Sim2) from the **Output Preview**, and on the toolbar, click

 to display in the **Output Summary**.

| | Sim1 | Sim2 |
|---|---|---|
| Mnemonic | SU-2S-LRSD | SU-2S-LRSD |
| **Test Parameters** | | |
| Design Type | Superiority | Superiority |
| Test Type | 1-Sided | 1-Sided |
| Test Statistic | Logrank | Logrank |
| Power | 0.026 | 0.027 |
| No. of Looks | 1 | 1 |
| **Model Parameters** | | |
| No. of Hazard Pieces | 1 | 1 |
| **Accrual & Dropout Parameters** | | |
| Followup Duration | Until End of Study | Until End of Study |
| No. of Accrual Periods | 1 | 1 |
| **Sample Size** | | |
| Maximum | 182 | 182 |
| **Events** | | |
| Maximum | 88 | 88 |
| **Simulation Results (Overall)** | | |
| Average Study Duration | 34.637 | 30.681 |
| Average Sample Size | 182 | 182 |
| Average Events | 88 | 88 |

Notice that the type-1 error appears to be controlled in both cases. When we simulated from the exponential (Sim2), the average study duration (30.7 months) was close to what was calculated at Des1 for the expected study duration under the null. However, when we simulated from the Weibull with decreasing hazards (Sim1), the average study duration increased to 34.6 months.

The ability to use custom R functions for many simulation tasks allows considerable flexibility in performing sensitivity analyses and assessment of key operating characteristics.

# 11 Superiority Trials with Variable Follow-Up

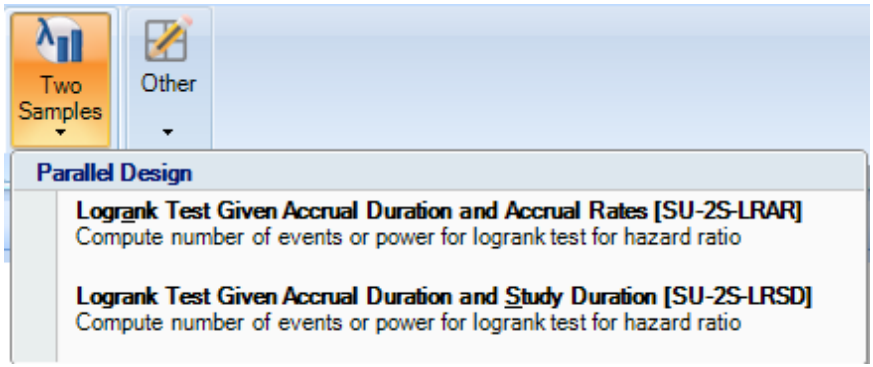
This chapter will illustrate through a worked example how to design, monitor and simulate a two-sample superiority trial with a time-to-event trial endpoint. Each subject who has not dropped out or experienced the event is followed until the trial ends. This implies that a subject who is enrolled earlier could potentially be followed for a longer time than a subject who is enrolled later on in the trial. In East we refer to such designs as **variable follow-up designs**.

## 11.1 The RALES Clinical Trial: Initial Design

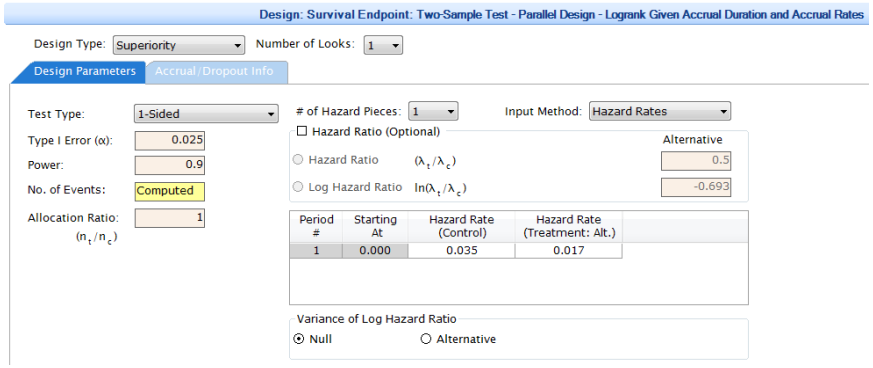
The RALES trial (Pitt et al., 1999) was a double blind study of aldosterone-receptor blocker spironolactone at a daily dose of 25 mg in combination with standard doses of an ACE inhibitor (treatment arm) versus standard therapy of an ACE inhibitor (control arm) in patients who had severe heart failure as a result of systolic left ventricular dysfunction. The primary endpoint was death from any cause. Six equally-spaced looks at the data using the Lan-DeMets-O'Brien-Fleming spending function were planned. The trial was designed to detect a hazard ratio of 0.83 with 90% power at a two-sided 0.05 level of significance. The hazard rate of the control arm was estimated to be 0.38/year. The trial was expected to enroll 960 patients/year.

We begin by using East to design RALES under these basic assumptions. Open East, click **Design** tab and then **Two Samples** button in **Survival** group. You will see the following

screen.

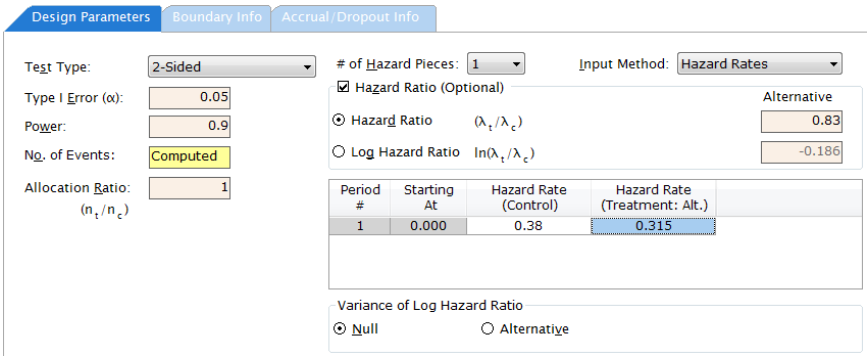



Note that there are two choices available in the above list; **Logrank Test Given Accrual Duration and Accrual Rates** and **Logrank Test Given Accrual Duration and Study Duration**. The option **Logrank Test Given Accrual Duration and Study Duration** is explained later in Chapter 13. Now click **Logrank Test Given Accrual Duration and Accrual Rates** and you will get the following input dialog box.

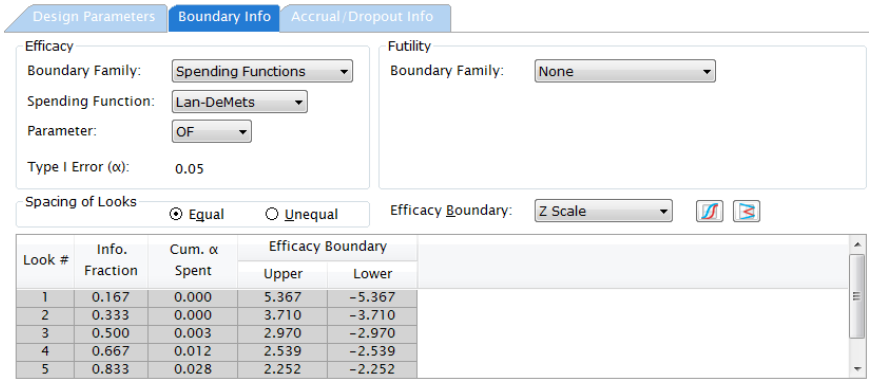



In the above dialog box, enter 6 for **Number of Looks** and keep the default choices of **Design Type:** `Superiority`, **Test Type:** `2-Sided`, **Type I Error ($\alpha$):** `0.05`, **Power :** `0.9`, and the **Allocation Ratio:** `1`.

Further, keep the default choices of **# of Hazard Pieces** as `1` and the **Input Method:** as `Hazard Rates`. Click the check box against **Hazard Ratio** and enter the **Hazard Ratio** as `0.83`. Enter **Hazard Rate (Control)** as `0.38`. You will see the **Hazard Rate (Treatment:Alt)** computed as `0.3154`. Also, keep the **Variance of Log Hazard Ratio** to be used as under `Null`. Now the **Design Parameters** tab of the input dialog will appear as shown below.

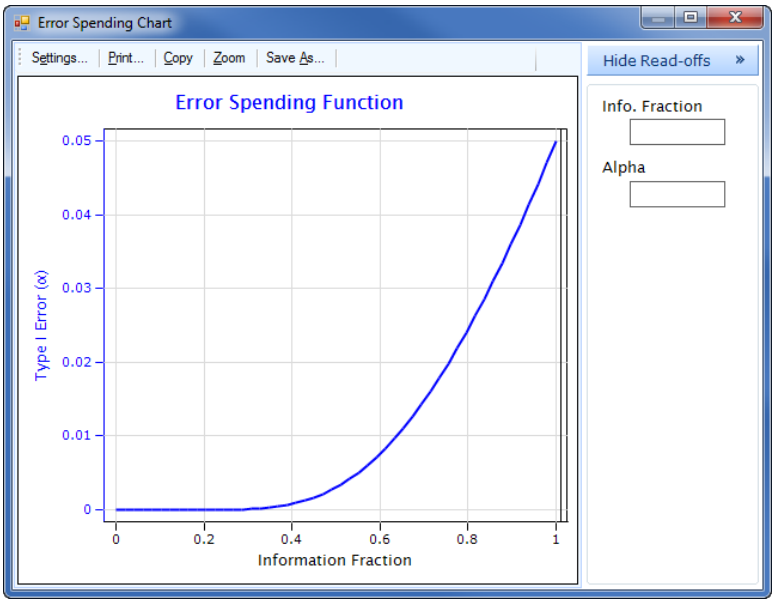



Now click on the tab **Boundary Info**. You will see the following input dialog box.

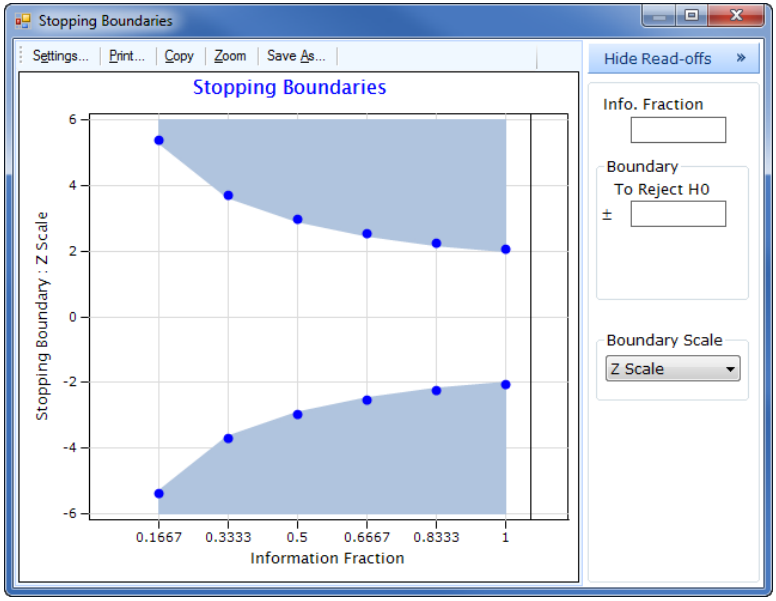



Keep all the default specifications for the boundaries to be used in the design. You can look at

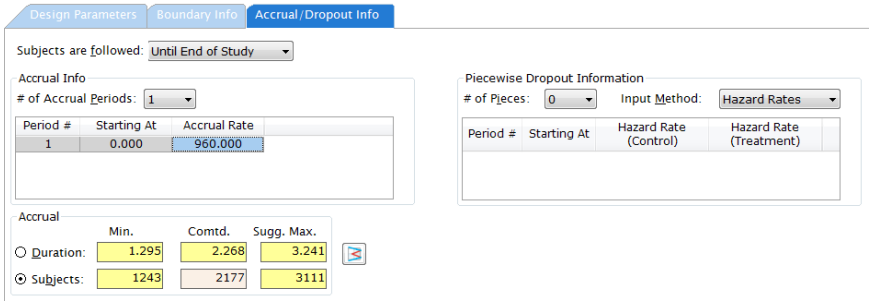
the Error Spending Chart by clicking on the icon 



If you click on the boundary chart icon , you will see the boundary chart as displayed below.




Now click **Accrual/Dropout Info** tab. Keep the default choice **`Until End of Study`** for the input **Subjects are followed:**. Keep the **# of Accrual Periods** as 1 and enter 960/year as the accrual rate. For this example, assume no dropouts. The dialog box will look as shown below.

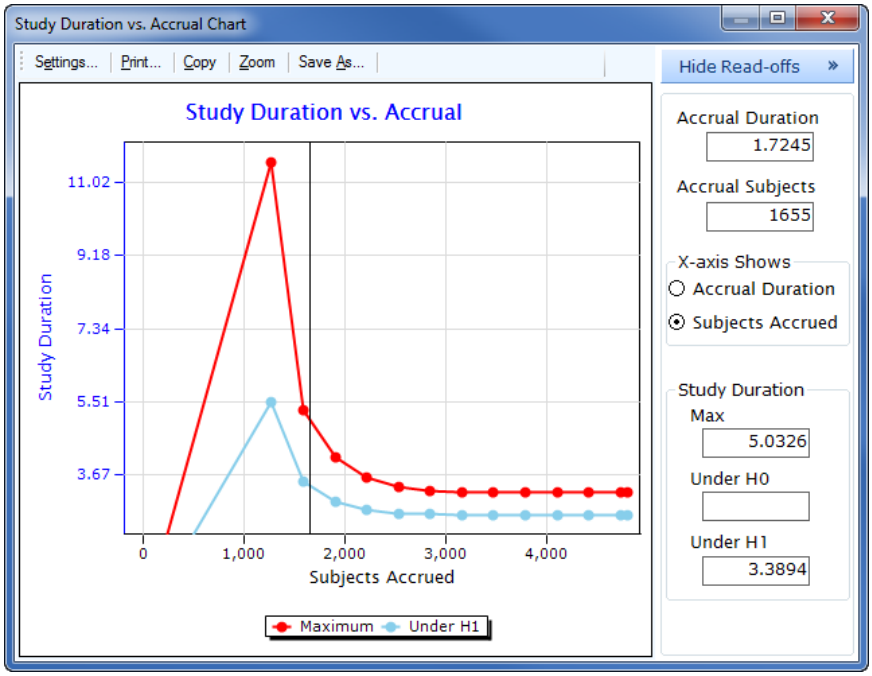



Under **Accrual** tab and in column titled **Comtd.** (commited) , you see two radio buttons

**Durations** and **Subjects** with the latter selected by default. The selected item will appear as the x-axis item in the **Study Duration vs. Accrual** chart, which you can get by clicking on the icon displayed on the side. Against **Durations** and **Subjects** you see two rows of three cells each. The first and third cells will show the min and max values for the row item and the middle cell, mid value between min and max values.

From the results displayed, you see that any sample size in the range 1243 to 3111 will suffice to attain the desired 90% power and selects 2177, the mid-point of the allowable range, as the default sample size. Depending on the needs of the study, you may wish to use a different sample size within the allowable range. The choice of sample size generally depends on how long you wish the study to last. The larger you make the patient accrual the shorter will be the total study duration, consisting of accrual time plus follow up time. To understand the essence of this trade-off, bring up the **Study Duration vs. Accrual** chart by clicking on the icon 

.



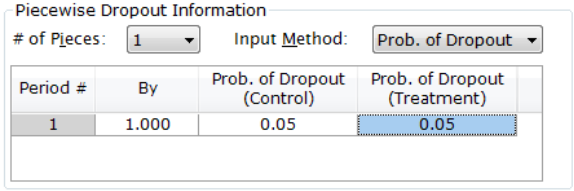

Based on this chart, a sample size of 1660 subjects is selected. Enter 1660 for **Committed Accrual (subjects)**. Click on **Compute** and see the results in the new plan created under **Output Preview**. This sample size ensures that the maximum study duration will be slightly more than 4.9 years. Additionally, under the alternative hypothesis, the expected study

duration will be only about 3.3 years.

| | Des1 |
|---|---|
| Mnemonic | SU-2S-LRAR |
| **Test Parameters** | |
| Design Type | Superiority |
| No. of Looks | 6 |
| Test Type | 2-Sided |
| Specified α | 0.05 |
| Power | 0.9 |
| **Model Parameters** | |
| Hazard Ratio (Alt.) | 0.83 |
| Var (Log HR) | Null |
| Allocation Ratio (nt/nc) | 1 |
| **Boundary Parameters** | |
| Spacing of Looks | Equal |
| Efficacy Boundary | LD (OF) |
| **Accrual & Dropout Parameters** | |
| Accrual Rate | 960 |
| Subjects are Followed | Until End of Study |
| No. of Accrual Periods | 1 |
| No. of Dropout Pieces | 0 |
| **Sample Size** | |
| Maximum | 1660 |
| Expected Under H0 | 1659.987 |
| Expected Under H1 | 1659.985 |
| **Events** | |
| Maximum | 1243 |
| Expected Under H0 | 1233.984 |
| Expected Under H1 | 903.595 |
| **Study Duration** | |
| Maximum | 4.905 |
| Expected Under H0 | 4.506 |
| Expected Under H1 | 3.337 |
| **Accrual Duration** | |
| Maximum | 1.729 |

## 11.2 Incorporating Drop-Outs

The investigators expect 5% of the patients in both the groups to drop out each year. To incorporate this drop-out rate into the design, in the **Piecewise Constant Dropout Rates** tab, select 1 for the number of pieces and change the Input Method from **Hazard Rates** to **Dropout Rates**. Then enter 5% dropouts at 1 year for both the groups.




We could have entered a hazard rate $\gamma$ for dropping out instead. By solving $1 - \exp(-\gamma) = 0.05$ we find $\gamma = -\ln(0.95) = 0.051$. This calculation is handled by East

To make Plan1 and Plan2 comparable change the sample size of Plan2 to 1660 by typing this value into the **Committed Accrual (Subjects)** cell. Click on **Compute** and see the

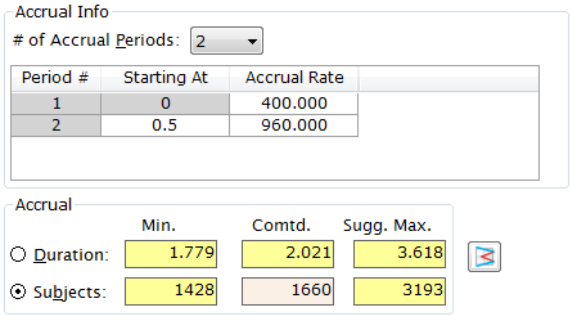
results in the new plan created under **`Output Preview`**.

| | Des1 | Des2 |
|---|---|---|
| Mnemonic | SU-2S-LRAR | SU-2S-LRAR |
| **Test Parameters** | | |
| Design Type | Superiority | Superiority |
| No. of Looks | 6 | 6 |
| Test Type | 2-Sided | 2-Sided |
| Specified α | 0.05 | 0.05 |
| Power | 0.9 | 0.9 |
| **Model Parameters** | | |
| Hazard Ratio (Alt.) | 0.83 | 0.83 |
| Var (Log HR) | Null | Null |
| Allocation Ratio (nt/nc) | 1 | 1 |
| **Boundary Parameters** | | |
| Spacing of Looks | Equal | Equal |
| Efficacy Boundary | LD (OF) | LD (OF) |
| **Accrual & Dropout Parameters** | | |
| Accrual Rate | 960 | 960 |
| Subjects are Followed | Until End of Study | Until End of Study |
| No. of Accrual Periods | 1 | 1 |
| No. of Dropout Pieces | 0 | 1 |
| **Sample Size** | | |
| Maximum | 1660 | 1660 |
| Expected Under H0 | 1659.987 | 1659.992 |
| Expected Under H1 | 1659.985 | 1659.986 |
| **Events** | | |
| Maximum | 1243 | 1243 |
| Expected Under H0 | 1233.984 | 1233.984 |
| Expected Under H1 | 903.595 | 903.595 |
| **Study Duration** | | |
| Maximum | 4.905 | 5.87 |
| Expected Under H0 | 4.506 | 5.258 |
| Expected Under H1 | 3.337 | 3.687 |
| **Accrual Duration** | | |
| Maximum | 1.729 | 1.729 |
| Expected Under H0 | 1.729 | 1.729 |
| Expected Under H1 | 1.729 | 1.729 |

A comparison of the first and second plans reveals that, because of the drop-outs, the maximum study duration will be prolonged from 4.9 years under Plan1 to 5.9 years under Plan2. The expected study duration will likewise be prolonged from 3.3 years to 3.7 years under the alternative hypothesis, and from 4.5 years to 5.3 years under the null hypothesis.

## 11.3 Incorporating Non-Constant Accrual Rates

In many clinical trials, the enrollment rate is low in the beginning and reaches its maximum expected level a few months later when all the sites enrolling patients have been recruited. Suppose that patients are expected to enroll at an average rate of 400/year for the first six months and at an average rate of 960/year thereafter. Now in **Accrual Info** tab, specify that there are two accrual periods and enter the accrual rate for each period in the dialog box as shown below.




Once again change the sample size to 1660 to make Plan3 comparable to the other two plans.

Click on **Compute** to complete the design.

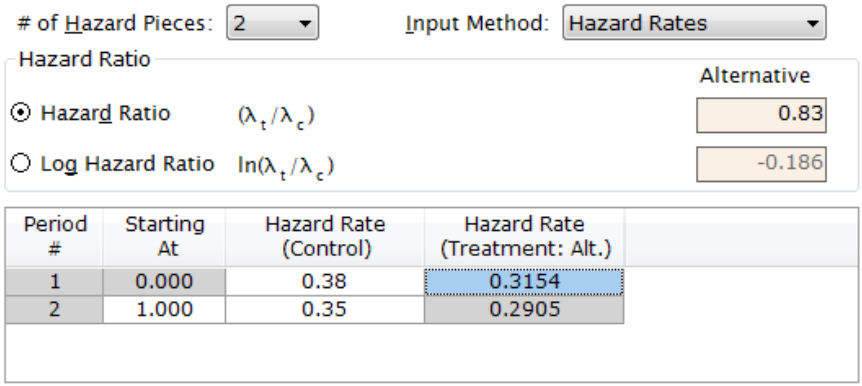
| | Des1 | Des2 | Des3 |
|---|---|---|---|
| Mnemonic | SU-2S-LRAR | SU-2S-LRAR | SU-2S-LRAR |
| **Test Parameters** | | | |
| Design Type | Superiority | Superiority | Superiority |
| No. of Looks | 6 | 6 | 6 |
| Test Type | 2-Sided | 2-Sided | 2-Sided |
| Specified α | 0.05 | 0.05 | 0.05 |
| Power | 0.9 | 0.9 | 0.9 |
| **Model Parameters** | | | |
| Hazard Ratio (Alt.) | 0.83 | 0.83 | 0.83 |
| Var (Log HR) | Null | Null | Null |
| Allocation Ratio (nt/nc) | 1 | 1 | 1 |
| **Boundary Parameters** | | | |
| Spacing of Looks | Equal | Equal | Equal |
| Efficacy Boundary | LD (OF) | LD (OF) | LD (OF) |
| **Accrual & Dropout Parameters** | | | |
| Accrual Rate | 960 | 960 | Multiple |
| Subjects are Followed | Until End of Study | Until End of Study | Until End of Study |
| No. of Accrual Periods | 1 | 1 | 2 |
| No. of Dropout Pieces | 0 | 1 | 1 |
| **Sample Size** | | | |
| Maximum | 1660 | 1660 | 1660 |
| Expected Under H0 | 1659.987 | 1659.992 | 1659.99 |
| Expected Under H1 | 1659.985 | 1659.986 | 1659.985 |
| **Events** | | | |
| Maximum | 1243 | 1243 | 1243 |
| Expected Under H0 | 1233.984 | 1233.984 | 1233.984 |
| Expected Under H1 | 903.595 | 903.595 | 903.595 |
| **Study Duration** | | | |
| Maximum | 4.905 | 5.87 | 6.15 |
| Expected Under H0 | 4.506 | 5.258 | 5.538 |
| Expected Under H1 | 3.337 | 3.687 | 3.966 |
| **Accrual Duration** | | | |
| Maximum | 1.729 | 1.729 | 2.021 |
| Expected Under H0 | 1.729 | 1.729 | 2.021 |
| Expected Under H1 | 1.729 | 1.729 | 2.021 |

Notice that the enrollment period has increased from 1.7 years to 2 years. Likewise, the maximum study duration and the expected study durations under $H_0$ and $H_1$ have also increased relative to Plans 1 and 2. Now the maximum study duration is 6.15 years.

## 11.4 Incorporating Piecewise Constant Hazards

Prior studies had suggested that the survival curves might not follow an exponential distribution. Suppose it is believed that the hazard rate for failure on the control arm decreases after the first 12 months from 0.38 to 0.35. We will assume that the hazard ratio is

still 0.83. We can enter the appropriate piecewise hazard rates into East as follows.




Change the sample size to 1660 for comparability with the previous plans. Click on **Compute** and see the results of the plan in the **Output Preview**.

| | Des1 | Des2 | Des3 | Des4 |
|---|---|---|---|---|
| Mnemonic | SU-2S-LRAR | SU-2S-LRAR | SU-2S-LRAR | SU-2S-LRAR |
| **Test Parameters** | | | | |
| Design Type | Superiority | Superiority | Superiority | Superiority |
| No. of Looks | 6 | 6 | 6 | 6 |
| Test Type | 2-Sided | 2-Sided | 2-Sided | 2-Sided |
| Specified α | 0.05 | 0.05 | 0.05 | 0.05 |
| Power | 0.9 | 0.9 | 0.9 | 0.9 |
| **Model Parameters** | | | | |
| Hazard Ratio (Alt.) | 0.83 | 0.83 | 0.83 | 0.83 |
| Var (Log HR) | Null | Null | Null | Null |
| Allocation Ratio (nt/nc) | 1 | 1 | 1 | 1 |
| **Boundary Parameters** | | | | |
| Spacing of Looks | Equal | Equal | Equal | Equal |
| Efficacy Boundary | LD (OF) | LD (OF) | LD (OF) | LD (OF) |
| **Accrual & Dropout Parameters** | | | | |
| Accrual Rate | 960 | 960 | Multiple | Multiple |
| Subjects are Followed | Until End of Study | Until End of Study | Until End of Study | Until End of Study |
| No. of Accrual Periods | 1 | 1 | 2 | 2 |
| No. of Dropout Pieces | 0 | 1 | 1 | 1 |
| **Sample Size** | | | | |
| Maximum | 1660 | 1660 | 1660 | 1660 |
| Expected Under H0 | 1659.987 | 1659.992 | 1659.99 | 1659.991 |
| Expected Under H1 | 1659.985 | 1659.986 | 1659.985 | 1659.985 |
| **Events** | | | | |
| Maximum | 1243 | 1243 | 1243 | 1243 |
| Expected Under H0 | 1233.984 | 1233.984 | 1233.984 | 1233.984 |
| Expected Under H1 | 903.595 | 903.595 | 903.595 | 903.595 |
| **Study Duration** | | | | |
| Maximum | 4.905 | 5.87 | 6.15 | 6.555 |
| Expected Under H0 | 4.506 | 5.258 | 5.538 | 5.868 |
| Expected Under H1 | 3.337 | 3.687 | 3.966 | 4.136 |
| **Accrual Duration** | | | | |
| Maximum | 1.729 | 1.729 | 2.021 | 2.021 |
| Expected Under H0 | 1.729 | 1.729 | 2.021 | 2.021 |
| Expected Under H1 | 1.729 | 1.729 | 2.021 | 2.021 |

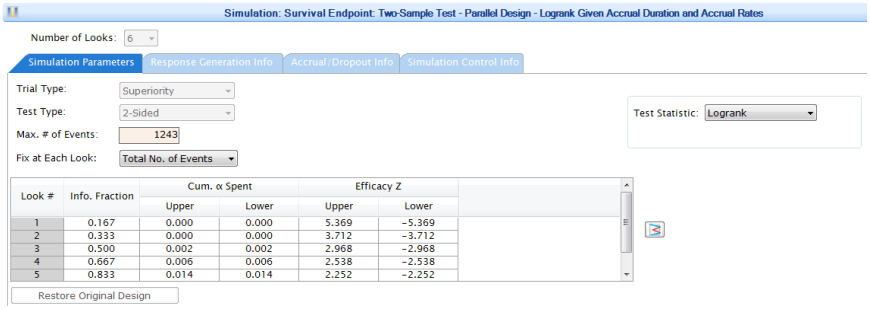
We observe that the impact of changing from a constant hazard rate to a piecewise constant

hazard rate is substantial. The maximum study duration has increased from 6.15 years for Plan3 to 6.56 years for Plan4.

## 11.5 Simulating a Trial with Proportional Hazards

▪ **11.5.1 Simulation Worksheet** ▪ **11.5.2 Simulating Under $H_1$** ▪ **11.5.3 Simulating...**

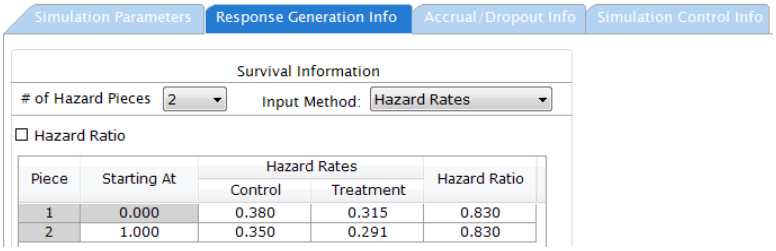
It would be useful to verify the operating characteristics of the various plans created in the previous section by simulation. The new survival simulation capabilities in East permit this. Let us use these capabilities to simulate Plan4. Save this design in the workbook. Right-click on this design node and select the menu item **Simulate**. You'll see the following **Survival Simulation** worksheet.



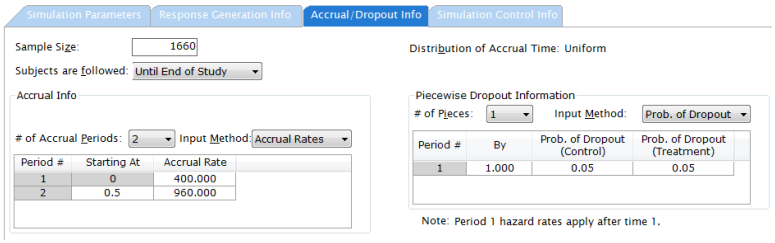

### *11.5.1 Components of the Simulation Worksheet*

This simulation worksheet consists four tabs - **`Simulation Parameters`**, **Response Generation Info**, **Accrual/Dropout Info**, and **Simulation Control Info**. The **Simulation Parameters** tab displays all the parameters of the simulation. If desired, you may modify one or more of these parameter values before carrying out simulation. The second tab **Response**

**Generation Info** will appear as shown below.

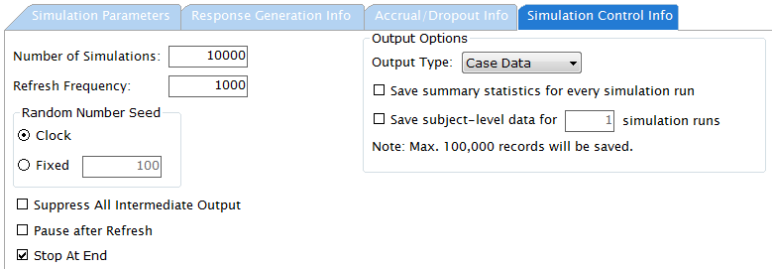



In this tab, you may modify values of response parameters before carrying out simulation. The third tab **`Accrual/Dropout Info`** will display information relating to accrual and dropouts.



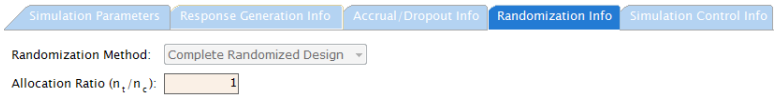

As in the case of other tabs, you may modify one or more values appearing in this tab before simulation is carried out.

In the **`Simulation Control Info`**, you may specify the simulation parameters like number of simulations required and the desired simulation seed etc.

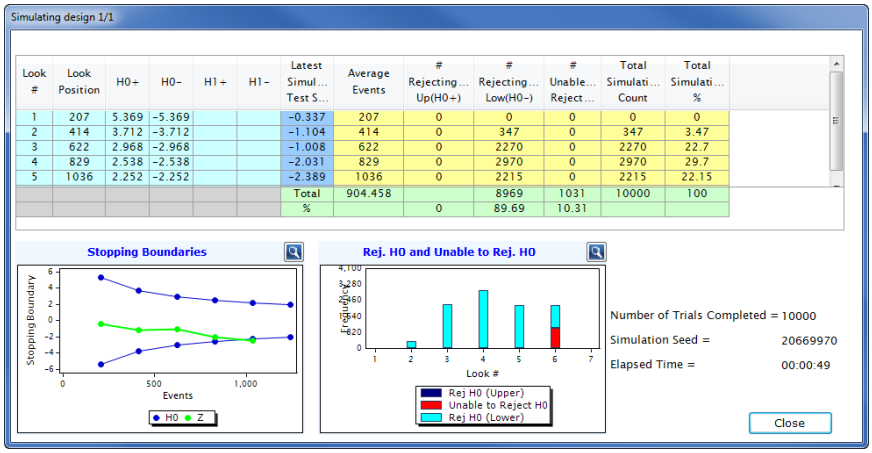


## Chapter 11: *Superiority Trials with Variable Follow-Up*

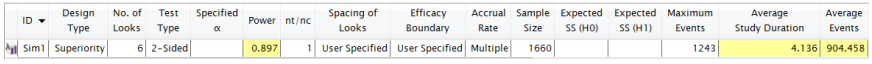
Also optionally, you may bring out one more tab **Randomization Info** by clicking on **Include Options**. In the **Randomization Info**, you may alter the allocation ratio of the design before carrying out simulation.




Keeping all the default parameter values same as in the different tabs, click **Simulate**. You can see the progress of the simulation process summarized as shown in the following screen shot.




At the end of simulation, the simulation results appear in a row in the **Output Preview** as shown below.




The output summary can be seen by clicking on the icon  after selecting the simulation

row in the Output Preview.

| | Sim1 |
|---|---|
| Mnemonic | SU-2S-LRAR |
| **Test Parameters** | |
| Design Type | Superiority |
| Test Type | 2-Sided |
| Test Statistic | Logrank |
| Power | 0.897 |
| No. of Looks | 6 |
| **Model Parameters** | |
| No. of Hazard Pieces | 2 |
| Allocation Ratio (nt/nc) | 1 |
| **Boundary Parameters** | |
| Efficacy Boundary | User Specified |
| Spacing of Looks | User Specified |
| **Accrual & Dropout Parameters** | |
| Followup Duration | Until End of Study |
| Accrual Rate | Multiple |
| No. of Accrual Periods | 2 |
| **Sample Size** | |
| Maximum | 1660 |
| **Events** | |
| Maximum | 1243 |
| **Simulation Results (Overall)** | |
| Average Study Duration | 4.136 |
| Average Sample Size | 1659.83 |
| Average Events | 904.458 |

Now save the simulation results to the workbook by selecting the simulation results row and then clicking on  . On this newly added workbook node for simulation, right-click and

select **Details**. You will see the complete details simulation appearing on the output pane.

**Simulation: Survival Endpoint: Two-Sample Test - Parallel Design - Logrank Given Accrual Duration and Accrual Rates**

| Simulation Parameters | |
|---|---|
| Simulation ID | Sim1 |
| Design Type | Superiority |
| Number of Looks | 6 |
| Test Type | 2-Sided |
| Sample Size (n) | 1660 |
| Fix at Each Look | Total No. of Events |
| Test Statistic | Logrank |
| Average Events | 904.458 |
| Total Accrual Duration | 2.0208 |
| Avg. Power at Termination | 0.897 |
| **Randomization Parameters** | |
| Method | Complete Randomized Design |
| Allocation Ratio ($n_t/n_c$) | 1 |
| **Simulation Control Parameters** | |
| Starting Seed | Clock |
| Number of Simulations | 10000 |

**Average Sample Size, Dropouts and Look Times**

| Look # | Average Sample Size | Average Events | | Average Dropouts | | Average Look Time | Average Follow up |
|---|---|---|---|---|---|---|---|
| | | Control | Treatment | Control | Treatment | | |
| 1 | 1137.766 | 111.911 | 95.089 | 15.066 | 15.532 | 1.476 | 0.525 |
| 2 | 1654.066 | 222.799 | 191.201 | 30.401 | 31.511 | 2.051 | 0.73 |
| 3 | 1660 | 332.382 | 289.618 | 45.931 | 48.045 | 2.641 | 1.104 |
| 4 | 1660 | 436.983 | 392.017 | 61.799 | 65.188 | 3.429 | 1.492 |
| 5 | 1660 | 538.567 | 497.433 | 77.752 | 82.34 | 4.556 | 1.883 |
| 6 | 1660 | 637.458 | 605.542 | 93.69 | 99.92 | 6.531 | 2.272 |
| Average | 1659.83 | 479.254 | 425.205 | 67.282 | 72.059 | 4.136 | 1.637 |

**Simulation Boundaries and Boundary Crossing Probabilities**

| Look # | Events | Boundaries | | Stopping For | | Total Simulations | |
|---|---|---|---|---|---|---|---|
| | | Efficacy | | | | | |
| | | Upper | Lower | Upper Efficacy | Lower Efficacy | Count | % |
| 1 | 207 | 5.369 | -5.369 | 0 | 0 | 0 | 0.000% |
| 2 | 414 | 3.712 | -3.712 | 0 | 347 | 347 | 3.470% |
| 3 | 622 | 2.968 | -2.968 | 0 | 2270 | 2270 | 22.700% |
| 4 | 829 | 2.538 | -2.538 | 0 | 2970 | 2970 | 29.700% |
| 5 | 1036 | 2.252 | -2.252 | 0 | 2215 | 2215 | 22.150% |
| 6 | 1243 | 2.045 | -2.045 | 0 | 1167 | 2198 | 21.980% |
| Total | | | | 0 | 8969 | 10000 | |
| % | | | | 0.000% | 89.690% | | |

## 11.5.2 Simulating Under $H_1$

We illustrate by running 1000 simulations for the current design with a fixed number of events at each look. Select a look time definition based on the number of events and click on the **Simulate** button. You will see a new row added in the Output Preview. Select this row and save it to Library node. If you double-click this node, you will see the following detailed

output. (The actual values may differ, depending on the starting seed used).

**Simulation: Survival Endpoint: Two-Sample Test - Parallel Design - Logrank Given Accrual Duration and Accrual Rates**

| Simulation Parameters | |
|---|---|
| Simulation ID: | Sim2 |
| Design Type: | Superiority |
| Number of Looks: | 6 |
| Test Type: | 2-Sided |
| Fix at Each Look: | Total No. of Events |
| Test Statistic: | Logrank |
| Average Events: | 901.004 |
| Total Accrual Duration: | 2.0208 |
| Avg. Power at Termination: | 0.911 |
| **Simulation Control Parameters** | |
| Starting Seed: | Clock |
| Number of Simulations: | 1000 |

**Simulation Boundaries and Boundary Crossing Probabilities:**

| Look # | Events | Boundaries Efficacy Upper | Boundaries Efficacy Lower | Early Stopping For Upper Efficacy | Early Stopping For Lower Efficacy | Total Simulations Count | Total Simulations % |
|---|---|---|---|---|---|---|---|
| 1 | 207 | 5.3688 | -5.3688 | 0 | 0 | 0 | 0 |
| 2 | 414 | 3.712 | -3.712 | 0 | 32 | 32 | 3.2 |
| 3 | 622 | 2.9683 | -2.9683 | 0 | 232 | 232 | 23.2 |
| 4 | 829 | 2.5382 | -2.5382 | 0 | 310 | 310 | 31 |
| 5 | 1036 | 2.252 | -2.252 | 0 | 208 | 208 | 20.8 |
| 6 | 1243 | 2.0448 | -2.0448 | 0 | 129 | 218 | 21.8 |
| Total | | | | 0 | 911 | 1000 | |
| % | | | | 0 | 91.1 | | |

**Average Sample Size, Dropouts and Look Times:**

| Look # | Average Sample Size | Average Events Control | Average Events Treatment | Average Dropouts Control | Average Dropouts Treatment | Average Look Time | Average Follow up |
|---|---|---|---|---|---|---|---|
| 1 | 1138.227 | 111.737 | 95.263 | 15.24 | 15.426 | 1.4765 | 0.5256 |
| 2 | 1653.936 | 222.491 | 191.509 | 30.628 | 31.365 | 2.0503 | 0.7298 |
| 3 | 1660 | 332.3264 | 289.6736 | 45.9246 | 47.9587 | 2.6428 | 1.1053 |
| 4 | 1660 | 437.3967 | 391.6033 | 61.9918 | 65.375 | 3.4323 | 1.494 |
| 5 | 1660 | 538.7418 | 497.2582 | 78.4484 | 82.5399 | 4.5669 | 1.8863 |
| 6 | 1660 | 637.6147 | 605.3853 | 93.9404 | 100.2248 | 6.5446 | 2.2754 |
| Average | 1659.935 | 477.596 | 423.408 | 67.091 | 71.981 | 4.1226 | 1.6333 |

**Response Generation Parameters**

No. of Hazard Pieces: 2
Input Method: Hazard Rates

| Piece # | Starting At | Control | Treatment | Hazard Ratio |
|---|---|---|---|---|
| 1 | 0 | 0.38 | 0.3154 | 0.83 |
| 2 | 1 | 0.35 | 0.2905 | 0.83 |

**Accrual/Dropout Parameters**

Sample Size: 1660
Subjects are Followed: Until End of Study
Accrual Input Method: Accrual Rates

| Period # | Starting At | Accrual Rate |
|---|---|---|
| 1 | 0 | 400 |
| 2 | 0.5 | 960 |

No. of Dropout Pieces: 1
Dropout Input Method: Dropout Rates

| Period # | At | Control | Treatment |
|---|---|---|---|
| 1 | 1 | 5 | 5 |

**Overall Simulation Results**

Average Study Duration: 4.123
Starting Seed: 426374642
Total Number of Simulations: 1000
Elapsed Time: 00:00:05

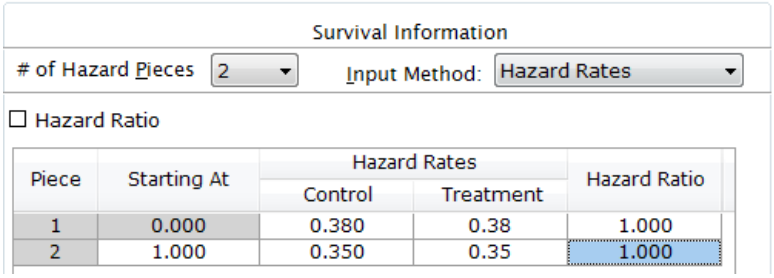
Let us examine these 1000 simulations more closely.

The column labeled **`Events`** in the first table, displays the number of events after which each interim look was taken. The column labeled **`Avg. Look Time`** in the second table, displays the average calendar times at which each interim look was taken. Thus, the first interim look (taken after observing 207 events) occurred after an average elapse of about 1.5 years; the second interim look (taken after observing 414 events) occurred after an average elapse of about 2.1 years; etc. The remaining columns of the simulation output are self-explanatory. The columns labeled **`Early Stopping For`** show that 911 of the 1000 simulations crossed the

lower stopping boundary, thus confirming (up to Monte Carlo accuracy) that this design has 90% power. The detailed output tables also show how the events, drop-outs, accruals, and average follow-up times were observed at each interim analysis.

### 11.5.3 Simulating Under $H_0$

To simulate under the null hypothesis we must go to the **`Response Generation Info`** tab. In this pane change the hazard rate for the treatment arm to 0.38 for the first piece and to 0.35 for the second piece of the hazard function.




| Piece | Starting At | Hazard Rates | | Hazard Ratio |
|---|---|---|---|---|
| | | Control | Treatment | |
| 1 | 0.000 | 0.380 | 0.38 | 1.000 |
| 2 | 1.000 | 0.350 | 0.35 | 1.000 |

This change implies that we will be simulating under the null hypothesis. Click on the **`Simulate`** button. A new row in Output Preview will be added now. Select this row and add to the library node. By double-clicking on this node, you will see the detailed simulation

output as shown below.The results are displayed below.

**Simulation: Survival Endpoint: Two-Sample Test - Parallel Design - Logrank Given Accrual Duration and Accrual Rates**

| Simulation Parameters | |
|---|---|
| Simulation ID: | Sim3 |
| Design Type: | Superiority |
| Number of Looks: | 6 |
| Test Type: | 2-Sided |
| Fix at Each Look: | Total No. of Events |
| Test Statistic: | Logrank |
| Average Events: | 1230.994 |
| Total Accrual Duration: | 2.0208 |
| Avg. Power at Termination: | 0.051 |

| Simulation Control Parameters | |
|---|---|
| Starting Seed: | Clock |
| Number of Simulations: | 1000 |

**Simulation Boundaries and Boundary Crossing Probabilities:**

| Look # | Events | Boundaries | | Early Stopping For | | Total Simulations | |
|---|---|---|---|---|---|---|---|
| | | Efficacy | | | | | |
| | | Upper | Lower | Upper Efficacy | Lower Efficacy | Count | % |
| 1 | 207 | 5.3688 | -5.3688 | 0 | 0 | 0 | 0 |
| 2 | 414 | 3.712 | -3.712 | 0 | 0 | 0 | 0 |
| 3 | 622 | 2.9683 | -2.9683 | 3 | 4 | 7 | 0.7 |
| 4 | 829 | 2.5382 | -2.5382 | 5 | 6 | 11 | 1.1 |
| 5 | 1036 | 2.252 | -2.252 | 6 | 9 | 15 | 1.5 |
| 6 | 1243 | 2.0448 | -2.0448 | 9 | 9 | 967 | 96.7 |
| Total | | | | 23 | 28 | 1000 | |
| % | | | | 2.3 | 2.8 | | |

**Average Sample Size, Dropouts and Look Times:**

| Look # | Average Sample Size | Average Events | | Average Dropouts | | Average Look Time | Average Follow up |
|---|---|---|---|---|---|---|---|
| | | Control | Treatment | Control | Treatment | | |
| 1 | 1089.595 | 103.372 | 103.628 | 13.924 | 14.346 | 1.4264 | 0.502 |
| 2 | 1614.547 | 207.072 | 206.929 | 28.183 | 28.46 | 1.9763 | 0.6816 |
| 3 | 1660 | 311.296 | 310.704 | 42.797 | 43.352 | 2.5136 | 1.0074 |
| 4 | 1660 | 414.5136 | 414.4864 | 57.8016 | 58.3162 | 3.2203 | 1.3596 |
| 5 | 1660 | 518.001 | 517.999 | 73.0183 | 73.5682 | 4.225 | 1.7154 |
| 6 | 1660 | 621.8925 | 621.1075 | 88.1303 | 88.7301 | 5.9331 | 2.0706 |
| Average | 1660 | 615.994 | 615 | 87.22 | 87.842 | 5.8538 | 2.0502 |

**Response Generation Parameters**

No. of Hazard Pieces: 2
Input Method: Hazard Rates

| Piece # | Starting At | Control | Treatment | Hazard Ratio |
|---|---|---|---|---|
| 1 | 0 | 0.38 | 0.38 | 1 |
| 2 | 1 | 0.35 | 0.35 | 1 |

**Accrual/Dropout Parameters**

Sample Size: 1660
Subjects are Followed: Until End of Study
Accrual Input Method: Accrual Rates

| Period # | Starting At | Accrual Rate |
|---|---|---|
| 1 | 0 | 400 |
| 2 | 0.5 | 960 |

No. of Dropout Pieces: 1
Dropout Input Method: Dropout Rates

| Period # | At | Control | Treatment |
|---|---|---|---|
| 1 | 1 | 5 | 5 |

**Overall Simulation Results**

Average Study Duration: 5.854
Starting Seed: 427433290
Total Number of Simulations: 1000
Elapsed Time: 00:00:06

Out of 1000 simulated trials only 23 crossed the upper stopping boundary and 28 crossed the lower stopping boundary thus confirming (up to Monte Carlo accuracy) that the type-1 error is preserved for this design.

## 11.6 Simulating a Trial with Non-Proportional Hazards

▪ 11.6.1 Single-Look Design ▪ 11.6.2 Single-Look Design ▪ 11.6.3 Group Seq. Design

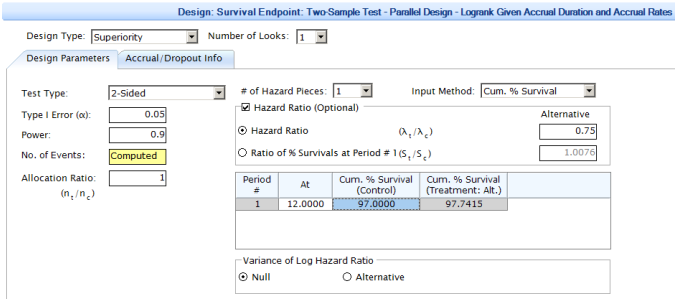
A new agent is to be tested against placebo in a large cardiovascular study with the endpoint being time to stroke, MI or death. The control arm has a 12-month event-free rate of 97%. We

wish to design the study to detect a hazard ratio of 0.75 with 90% power, using a two-sided test conducted at the 0.05 level. An important design consideration is that treatment differences are expected to emerge only after one year of therapy. Subjects will enroll at the rate of 1000/month and be followed to the end of the study. The dropout rate is expected to be 10% per year for both treatment arms. Finally, the study should be designed for maximum study duration of 50 months.

The usual design options in East are not directly applicable to this trial because they require the hazard ratio to be constant under the alternative hypothesis. Here, however, we are required to power the trial to detect a hazard ratio of 0.75 that only emerges after patients have been on the study for 12 months. The simulation capabilities of East can help us with the design.

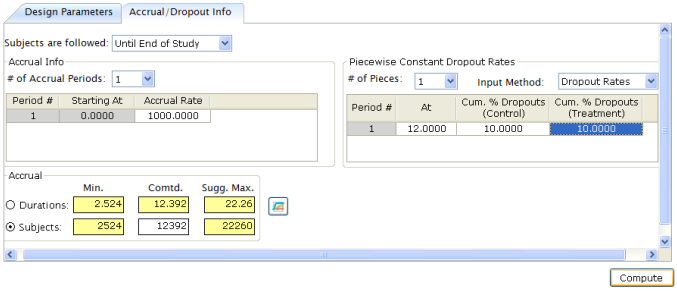
### *11.6.1 Single-Look Design with Proportional Hazards*

We begin by creating a single-look design powered to detect hazard ratio of 0.75, ignoring the fact that the two survival curves separate out only after 12 months. Open a new survival design worksheet by clicking on **Design–>Time to Event–>Logrank Test Given Accrual Duration and Accrual Rates**. In the resulting `Design Parameters` tab, enter the parameters values as shown below.

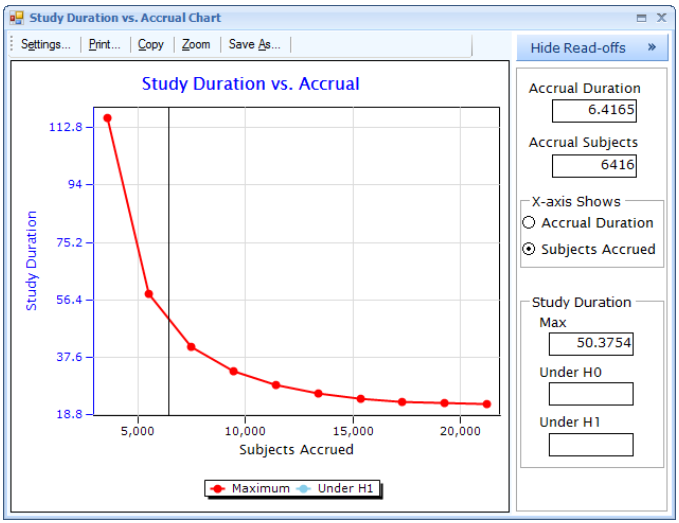



Click on the tab `Accrual/Dropout Info` and enter the values as shown below, excluding

the Accrual tab.




East informs you in the Accrual tab, that any sample size in the range 2524 to 22260 will suffice to attain the desired 90% power. However, the study will end sooner if we enroll more patients. Recall that we wish the trial to last no more than 50 months, inclusive of accrual and follow-up. The **Accrual-Duration** chart can provide guidance on sample size selection. This chart reveals that if 6400 subjects are enrolled, the expected maximum duration of a trial is close to 50 months.




Now change the **Comtd.** number of subjects to 6400 and click on **Compute** to complete the

design. A new row is added for this design in the Output Preview. Select this row and add it to a library node under a workbook. Now you double-click on this node, you will see the detailed output as shown below.

**Design: Survival Endpoint: Two-Sample Test - Parallel Design - Logrank Given Accrual Duration and Accrual Rates**

| Test Parameters | |
|---|---|
| Design ID: | Des1 |
| Design Type: | Superiority |
| Number of Looks: | 1 |
| Test Type: | 2-Sided |
| Specified α: | 0.05 |
| Power: | 0.9 |
| **Model Parameters** | |
| HR = $\lambda_t/\lambda_c$ | |
| Under H0: | 1 |
| Under H1: | 0.75 |
| Ratio of % Surv. at Period #1: | 1.0076 |
| Var (Log HR): | Null |
| Allocation Ratio ($n_t/n_c$): | 1 |
| **Accrual/Dropout Parameters** | |
| Accrual Duration: | 6.4 |
| Max Study Duration: | 48.677 |
| Dropout: | Yes |

**Output:**

| | |
|---|---|
| Upper Critical Point | 1.96 |
| Lower Critical Point | -1.96 |
| Sample Size (n) | 6400 |
| Sample Size Treatment (n_t) | 3201 |
| Sample Size Control (n_c) | 3199 |
| Events (s) | 508 |
| Events Treatment (s_t) | 220 |
| Events Control (s_c) | 288 |
| Dropouts (d) | 2010 |
| Dropouts Treatment (d_t) | 1012 |
| Dropouts Control (d_c) | 998 |
| Accrual Duration | 6.4 |
| Study Duration | 48.677 |
| Max. Information (I_max) | 127 |

**Survival Info. : Cum. % Survival**

| Period # | At | Cum. % Survival | | Hazard Ratio |
|---|---|---|---|---|
| | | Control (λc) | Treatment (λt1) | Alt. (λt1/λc) |
| 1 | 12 | 97 | 97.741 | 0.75 |

**Accrual Info. :**

| Period # | At | Accrual Rate |
|---|---|---|
| 1 | 0 | 1000 |

**Dropout Info. : %Dropout Rates**

| Period # | At | Control | Treatment |
|---|---|---|---|
| 1 | 12 | 10 | 10 |

We can verify the operating characteristics of Plan1 by simulation. With the cursor on Plan1 node, Click on Simulation icon from the library menu bar. You'll be taken to the survival simulation worksheet. In the **`Simulation Control Info`** tab, specify the number of simulations to be 1000. Now click on **`Simulate`** button. This will generate 1000 simulations from the survival curves specified in the design. Each simulation will consist of survival data on 6400 subjects entering the trial uniformly at the rate of 1000/month. Events (failures) will be tracked and the simulated trial will be terminated when the total number of events equals 508. Subjects surviving past this termination time point will have their survival times censored. The resulting survival data will be summarized in terms of the logrank test statistic. Each simulation records two important quantities:

- the calendar time at which the last of the specified 508 events arrived;
- whether or not the logrank test statistic rejected the null hypothesis.

We would expect that, on average, the 508 events will occur in about 48.7 months and about 90% of the simulations will reject the null hypothesis. The simulation summary is shown in the

following screen shot.

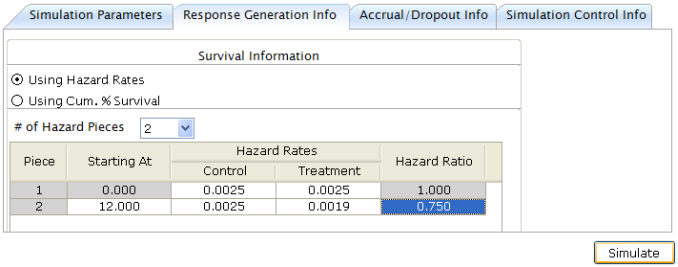
| | Sim1 |
|---|---|
| Mnemonic | SU-2S-LRAR |
| **Test Parameters** | |
| Design Type | Superiority |
| Test Type | 2-Sided |
| Test Statistic | Logrank |
| Power | 0.891 |
| No. of Looks | 1 |
| **Model Parameters** | |
| No. of Hazard Pieces | 1 |
| **Accrual & Dropout Parameters** | |
| Followup Duration | Until End of Study |
| Accrual Rate | 1000 |
| No. of Accrual Periods | 1 |
| **Sample Size** | |
| Maximum | 6400 |
| **Events** | |
| Maximum | 508 |
| **Study Duration** | |
| Maximum | 48.7946 |

Indeed we observe that the average study duration for this set of 1000 simulations was 48.8 months, and that 891 of the 1000 simulated trials crossed the critical value and rejected $H_0$ and hence the power attained is 0.891. This serves as an independent verification of the operating characteristics of Plan1, up to Monte Carlo accuracy.

### 11.6.2 Single-Look Design with Non-Proportional Hazards

Were it not for the fact that the hazard ratio of 0.75 only emerges after 12 months of therapy, Plan1 would meet the goals of this study. However, the impact of the late separation of the survival curves must be taken into consideration. This is accomplished, once again, by simulation. Click the Edit Simulation icon while the cursor is on the last simulation node. In the resulting simulation sheet click on Response Generation Info tab. In this tab, specify that the hazard rates for the control and treatment arms are identical and equal to 0.0025 for the first 12 months and the hazard ratio is 0.75 thereafter. This is done by making appropriate entries

in this tab as shown below.




Click on the **`Simulate`** button. This will generate 1000 simulations from survival curves specified in the **`Survival Parameters Pane`**. As before, each simulation will consist of survival data on 6400 subjects entering the trial uniformly at the rate of 1000/month. Events (failures) will be tracked and the simulated trial will be terminated when the total number of

events equals 508. The summary output of this simulation run as shown below.

| | Wbk2:Des1:Sim2 |
|---|---|
| Mnemonic | SU-2S-LRAR |
| **Test Parameters** | |
| Design Type | Superiority |
| Test Type | 2-Sided |
| Test Statistic | Logrank |
| Power | 0.565 |
| No. of Looks | 1 |
| **Model Parameters** | |
| No. of Hazard Pieces | 2 |
| **Accrual & Dropout Parameters** | |
| Followup Duration | Until End of Study |
| Accrual Rate | 1000 |
| No. of Accrual Periods | 1 |
| **Sample Size** | |
| Maximum | 6400 |
| **Events** | |
| Maximum | 508 |
| **Study Duration** | |
| Maximum | 46.9637 |

This time only 565 of the 1000 trials were able to reject $H_0$.The drop in power is of course due to the fact that the two survival curves do not separate out until 12 months have elapsed. Thus events that arise within the first 12 months arrive at the same rate for both arms and are not very informative about treatment differences.

We need to increase the power of the study to 90%. This can be accomplished in one of two ways:

1. Prolonging the study duration until a sufficient number of events are obtained to achieve 90% power.
2. Increasing the sample size.

The first approach cannot be used because the study duration is not permitted to exceed 50 months. The simulations have shown that the study duration is already almost 50 months, and it has only achieved 56.5% power. Thus we must resort to increasing the sample size.

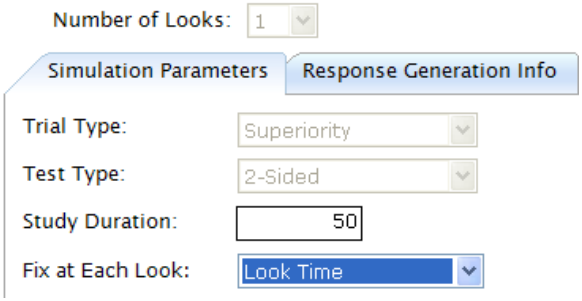
Now if we increase the sample size while keeping the total number of events fixed at 508, the average study duration will drop. The power, however, may not increase. In fact it might even decrease since a larger fraction of the 508 events will arise in the first 12 months, before the two survival curves have separated. To see this, increase the sample size from 6400 to 10000 in the **Accrual/Dropout Info** tab. Then click on **Simulate** button. From this simulation run, you will get the output summary as shown below.

| | Wbk2:Des1:Sim3 |
|---|---|
| Mnemonic | SU-2S-LRAR |
| **Test Parameters** | |
| Design Type | Superiority |
| Test Type | 2-Sided |
| Test Statistic | Logrank |
| Power | 0.297 |
| No. of Looks | 1 |
| **Model Parameters** | |
| No. of Hazard Pieces | 2 |
| **Accrual & Dropout Parameters** | |
| Followup Duration | Until End of Study |
| Accrual Rate | 1000 |
| No. of Accrual Periods | 1 |
| **Sample Size** | |
| Maximum | 10000 |
| **Events** | |
| Maximum | 508 |
| **Study Duration** | |
| Maximum | 29.7351 |

Notice that the average study duration has dropped to 29.7 months. But the power has dropped also. This time only 297 of the 1000 simulations could reject the null hypothesis.

To increase power we must increase sample size while keeping the study duration fixed at about 50 months. This is accomplished by selecting the **Look Time** option from the drop-down box in the **Fix at Each Look** section of the **Survival Parameters Pane** and choosing a 50 month Total Study Durn., while keeping the sample size increase from 6400

to 10000.




We will now run 1000 simulations in each of which 10000 subjects are enrolled at the rate of 1000/year. Each simulated trial will be terminated at the end of 50 months of calendar time and a logrank test statistic will be derived from the data.Click on the **Simulate** button. Add

the simulation run output to library node and see the following output summary.

| | Wbk2:Des1:Sim4 |
|---|---|
| Mnemonic | SU-2S-LRAR |
| **Test Parameters** | |
| Design Type | Superiority |
| Test Type | 2-Sided |
| Test Statistic | Logrank |
| Power | 0.752 |
| No. of Looks | 1 |
| **Model Parameters** | |
| No. of Hazard Pieces | 2 |
| **Accrual & Dropout Parameters** | |
| Followup Duration | Until End of Study |
| Accrual Rate | 1000 |
| No. of Accrual Periods | 1 |
| **Sample Size** | |
| Maximum | 10000 |
| **Events** | |
| Maximum | 808.229 |
| **Study Duration** | |
| Maximum | 50 |

For more details, you can click  icon after selecting the saved simulation node.

**Simulation: Survival Endpoint: Two-Sample Test - Parallel Design - Logrank Given Accrual Duration and Accrual Rates**

| Simulation Parameters | |
|---|---|
| Simulation ID: | Sim4 |
| Design Type: | Superiority |
| Number of Looks: | 1 |
| Test Type: | 2-Sided |
| Fix at Each Look: | Look Time |
| Test Statistic: | Logrank |
| Average Events: | 808.229 |
| Total Accrual Duration: | 10 |
| Avg. Power at Termination: | 0.752 |
| **Simulation Control Parameters** | |
| Starting Seed: | Clock |
| Number of Simulations: | 1000 |

**Simulation Boundaries and Boundary Crossing Probabilities:**

| Look # | Look Time | Boundaries Efficacy Upper | Boundaries Efficacy Lower | Early Stopping For Upper Efficacy | Early Stopping For Lower Efficacy | Total Simulations Count | Total Simulations % |
|---|---|---|---|---|---|---|---|
| 1 | 50 | 1.96 | -1.96 | 0 | 752 | 1000 | 100 |
| Total | | | | 0 | 752 | 1000 | |
| % | | | | 0 | 75.2 | | |

**Average Sample Size, Dropouts and Look Times:**

| Look # | Average Sample Size | Average Events Control | Average Events Treatment | Average Dropouts Control | Average Dropouts Treatment | Average Look Time | Average Follow up |
|---|---|---|---|---|---|---|---|
| 1 | 10000 | 440.503 | 367.726 | 1546.643 | 1560.357 | 50 | 35.3794 |
| Average | 10000 | 440.503 | 367.726 | 1546.643 | 1560.357 | 50 | 35.3794 |

**Response Generation Parameters**

No. of Hazard Pieces: 2
Input Method: Hazard Rates

| Piece # | Starting At | Control | Treatment | Hazard Ratio |
|---|---|---|---|---|
| 1 | 0 | 0.003 | 0.003 | 1 |
| 2 | 12 | 0.003 | 0.002 | 0.75 |

**Accrual/Dropout Parameters**

Sample Size: 10000
Subjects are Followed: Until End of Study
Accrual Input Method: Accrual Rates

| Period # | Starting At | Accrual Rate |
|---|---|---|
| 1 | 0 | 1000 |

No. of Dropout Pieces: 1
Dropout Input Method: Dropout Rates

| Period # | At | Control | Treatment |
|---|---|---|---|
| 1 | 12 | 10 | 10 |

**Overall Simulation Results**

Average Study Duration: 50
Starting Seed: 431473585
Total Number of Simulations: 1000
Elapsed Time: 00:00:07

Now you can see, the power of the study has increased to 75.2%. On average 808 events occurred during the 50 months that the study remained open. Since we require 90% power, the sample size must be increased even further. This can be done by trial and error over several simulation experiments. Eventually we discover that a sample size of 17200 patients

will provide about 90% power with an average of 1304 events.

**Simulation: Survival Endpoint: Two-Sample Test - Parallel Design - Logrank Given Accrual Duration and Accrual Rates**

| Simulation Parameters | |
|---|---|
| Simulation ID: | Sim5 |
| Design Type: | Superiority |
| Number of Looks: | 1 |
| Test Type: | 2-Sided |
| Fix at Each Look: | Look Time |
| Test Statistic: | Logrank |
| Average Events: | 1303.896 |
| Total Accrual Duration: | 17.2 |
| Avg. Power at Termination: | 0.906 |
| **Simulation Control Parameters** | |
| Starting Seed: | Clock |
| Number of Simulations: | 1000 |

**Simulation Boundaries and Boundary Crossing Probabilities:**

| Look # | Look Time | Boundaries | | Early Stopping For | | Total Simulations | |
|---|---|---|---|---|---|---|---|
| | | Efficacy | | | | | |
| | | Upper | Lower | Upper Efficacy | Lower Efficacy | Count | % |
| 1 | 50 | 1.96 | -1.96 | 0 | 906 | 1000 | 100 |
| Total | | | | 0 | 906 | 1000 | |
| % | | | | 0 | 90.6 | | |

**Average Sample Size, Dropouts and Look Times:**

| Look # | Average Sample Size | Average Events | | Average Dropouts | | Average Look Time | Average Follow up |
|---|---|---|---|---|---|---|---|
| | | Control | Treatment | Control | Treatment | | |
| 1 | 17200 | 710.04 | 593.856 | 2488.851 | 2506.108 | 50 | 33.0892 |
| Average | 17200 | 710.04 | 593.856 | 2488.851 | 2506.108 | 50 | 33.0892 |

**Response Generation Parameters**

No. of Hazard Pieces: 2
Input Method: Hazard Rates

| Piece # | Starting At | Control | Treatment | Hazard Ratio |
|---|---|---|---|---|
| 1 | 0 | 0.003 | 0.003 | 1 |
| 2 | 12 | 0.003 | 0.002 | 0.75 |

**Accrual/Dropout Parameters**

Sample Size: 17200
Subjects are Followed: Until End of Study
Accrual Input Method: Accrual Rates

| Period # | Starting At | Accrual Rate |
|---|---|---|
| 1 | 0 | 1000 |

No. of Dropout Pieces: 1
Dropout Input Method: Dropout Rates

| Period # | At | Control | Treatment |
|---|---|---|---|
| 1 | 12 | 10 | 10 |

**Overall Simulation Results**

Average Study Duration: 50
Starting Seed: 431880926
Total Number of Simulations: 1000
Elapsed Time: 00:00:10

It is evident from these simulations that the proportional hazards assumption is simply not appropriate if the survival curves separate out late. In the present example the proportional hazards assumption would have led to a sample size of 6400 whereas the sample size actually needed was 17200.

### *11.6.3 Group Sequential Design with Non-Proportional Hazards*

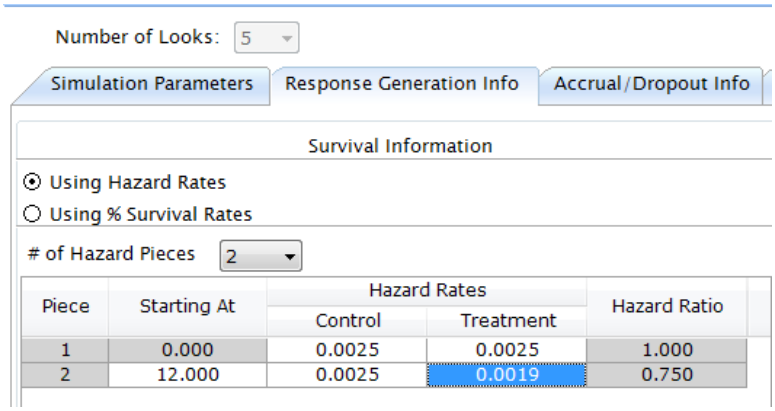
The single-look design discussed in the previous section required a sample size of 17200 subjects. A group sequential design, monitored by an independent data monitoring committee, is usually more efficient for large studies of this type. Such a trial can be designed with efficacy stopping boundaries or with efficacy and futility stopping boundaries. Consider first a design with five equally spaced efficacy boundaries. Go back to the library, click on Des1

node, and then click on  . In the resulting design input dialog window, change the entry in the **Number of Looks** cell from 1 to 5. Click on **Compute** button and save the plan as Des2 in the library. Select Des1 and Des2 nodes and then click on  to see the following details for both the plans.

| | Wbk2:Des1 | Wbk2:Des2 |
|---|---|---|
| Mnemonic | SU-2S-LRAR | SU-2S-LRAR |
| **Test Parameters** | | |
| Design Type | Superiority | Superiority |
| No. of Looks | 1 | 5 |
| Test Type | 2-Sided | 2-Sided |
| Specified α | 0.05 | 0.05 |
| Power | 0.9 | 0.9002 |
| **Model Parameters** | | |
| Hazard Ratio (Alt.) | 0.75 | 0.75 |
| Var (Log HR) | Null | Null |
| Allocation Ratio (nt/nc) | 1 | 1 |
| **Boundary Parameters** | | |
| Spacing of Looks | | Equal |
| Efficacy Boundary | | LD (OF) |
| **Accrual & Dropout Parameters** | | |
| Accrual Rate | 1000 | 1000 |
| Subjects are Followed | Until End of Study | Until End of Study |
| No. of Accrual Periods | 1 | 1 |
| No. of Dropout Pieces | 1 | 1 |
| **Sample Size** | | |
| Maximum | 6400 | 12555 |
| Expected Under H0 | 6400 | 12554.9964 |
| Expected Under H1 | 6400 | 12554.1223 |
| **Events** | | |
| Maximum | 508 | 520 |
| Expected Under H0 | 508 | 516.5858 |
| Expected Under H1 | 508 | 385.5073 |
| **Accrual Duration** | | |
| Maximum | 6.4 | 12.555 |
| Expected Under H0 | 6.4 | 12.555 |
| Expected Under H1 | 6.4 | 12.5541 |
| **Study Duration** | | |
| Maximum | 48.677 | 27.2317 |
| Expected Under H0 | 41.817 | 24.2646 |
| Expected Under H1 | 48.677 | 21.4502 |

Des2 reveals that a group sequential design, with five equally spaced looks, taken after observing 104, 208, 312, 416 and 520 events, respectively, utilizing the default Lan-DeMets-O'Brien-Fleming (**LD (OF)**) spending function, achieves 90% power with a

maximum sample size of 12555 and a maximum study duration of 27.232 months. The expected study duration under $H_1$ is 21.451 months. However, these operating characteristics are based on the assumption that the hazard ratio is constant and equals 0.75. Since in fact the hazard ratio is 0.75 only after 12 months of treatment, the actual power of this design is unlikely to be 90%. We can use simulation to determine the actual power. With the cursor in any cell of Des2 node, select  from the menu bar. You will be taken to the simulation worksheet. In the **`Response Generation Info`** tab, make the changes in the hazard rates as shown below.



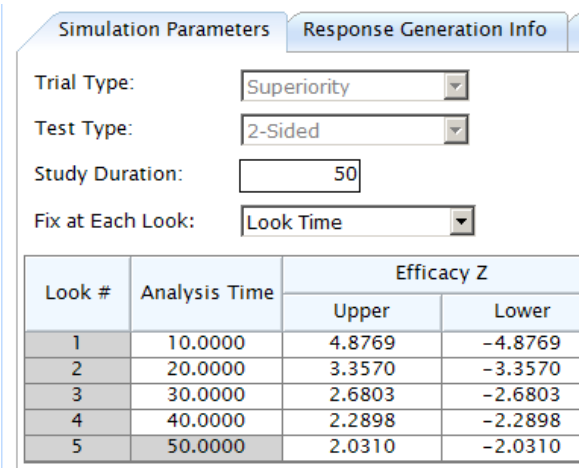

After changing the number of simulations as 1000 in the **Simulation Control Info**, click on the **`Simulate`** button to run 1000 simulations of Des2 with data being generated from the survival distributions that were specified in the **`Response Generation Info`** tab. The

results of this simulation run are as shown below.

| | Wbk2:Des2:Sim6 |
|---|---|
| Mnemonic | SU-2S-LRAR |
| **Test Parameters** | |
| Design Type | Superiority |
| Test Type | 2-Sided |
| Test Statistic | Logrank |
| Power | 0.187 |
| No. of Looks | 5 |
| **Model Parameters** | |
| No. of Hazard Pieces | 2 |
| **Boundary Parameters** | |
| Efficacy Boundary | User Specified |
| Spacing of Looks | User Specified |
| **Accrual & Dropout Parameters** | |
| Followup Duration | Until End of Study |
| Accrual Rate | 1000 |
| No. of Accrual Periods | 1 |
| **Sample Size** | |
| Maximum | 12555 |
| **Events** | |
| Maximum | 520 |
| **Study Duration** | |
| Maximum | 25.2768 |

Only 187 of the 1000 simulated trials were able to reject the null hypothesis indicating that the study is grossly underpowered. We can improve on this performance by extending the total study duration so that additional events may be observed. To increase study duration, go to the **`Simulation Parameters`** tab and select the **`Look Time`** option under **`Fix at Each Look`**. We had specified at the outset that the total study duration should not exceed 50 months. Let us therefore fix the total study duration at 50 months and space each interim look

10 months apart by editing the **Study Duration**.




| Look # | Analysis Time | Efficacy Z | |
|---|---|---|---|
| | | Upper | Lower |
| 1 | 10.0000 | 4.8769 | -4.8769 |
| 2 | 20.0000 | 3.3570 | -3.3570 |
| 3 | 30.0000 | 2.6803 | -2.6803 |
| 4 | 40.0000 | 2.2898 | -2.2898 |
| 5 | 50.0000 | 2.0310 | -2.0310 |

We are now ready to simulate a 5-look group sequential trial in which the **LD(OF)** stopping boundaries are applied and the looks are spaced 10 months apart. Each simulated trial will enroll 12555 subjects at the rate of 1000/month. The simulation data will be generated from survival distributions in which the hazard rates of both arms are 0.0025 for the first 12 months and the hazard ratio is 0.75 thereafter. To generate 1000 simulations of this design click on the **Simulate** button. These simulations do indeed show a substantial increase in power, from

18.7% previously to 79.9% .

| | Wbk2:Des2:Sim6 | Wbk2:Des2:Sim7 |
|---|---|---|
| Mnemonic | SU-2S-LRAR | SU-2S-LRAR |
| **Test Parameters** | | |
| Design Type | Superiority | Superiority |
| Test Type | 2-Sided | 2-Sided |
| Test Statistic | Logrank | Logrank |
| Power | 0.187 | 0.799 |
| No. of Looks | 5 | 5 |
| **Model Parameters** | | |
| No. of Hazard Pieces | 2 | 2 |
| **Boundary Parameters** | | |
| Efficacy Boundary | User Specified | User Specified |
| Spacing of Looks | User Specified | User Specified |
| **Accrual & Dropout Parameters** | | |
| Followup Duration | Until End of Study | Until End of Study |
| Accrual Rate | 1000 | 1000 |
| No. of Accrual Periods | 1 | 1 |
| **Sample Size** | | |
| Maximum | 12555 | 12555 |
| **Events** | | |
| Maximum | 520 | 879.219 |
| **Study Duration** | | |
| Maximum | 25.2768 | 50 |

The design specifications stated, however, that the trial should have 90% power. In order to achieve this amount of power we will have to increase the sample size. By trial and error, upon increasing the sample size to 18200 on the **`Simulation Parameters`** tab we observe that

the power has increased to 90 % (up to Monte Carlo accuracy).

| | Wbk2:Des2:Sim6 | Wbk2:Des2:Sim7 | Wbk2:Des2:Sim8 |
|---|---|---|---|
| Mnemonic | SU-2S-LRAR | SU-2S-LRAR | SU-2S-LRAR |
| **Test Parameters** | | | |
| Design Type | Superiority | Superiority | Superiority |
| Test Type | 2-Sided | 2-Sided | 2-Sided |
| Test Statistic | Logrank | Logrank | Logrank |
| Power | 0.187 | 0.799 | 0.904 |
| No. of Looks | 5 | 5 | 5 |
| **Model Parameters** | | | |
| No. of Hazard Pieces | 2 | 2 | 2 |
| **Boundary Parameters** | | | |
| Efficacy Boundary | User Specified | User Specified | User Specified |
| Spacing of Looks | User Specified | User Specified | User Specified |
| **Accrual & Dropout Parameters** | | | |
| Followup Duration | Until End of Study | Until End of Study | Until End of Study |
| Accrual Rate | 1000 | 1000 | 1000 |
| No. of Accrual Periods | 1 | 1 | 1 |
| **Sample Size** | | | |
| Maximum | 12555 | 12555 | 18300 |
| **Events** | | | |
| Maximum | 520 | 879.219 | 1169.605 |
| **Study Duration** | | | |
| Maximum | 25.2768 | 50 | 50 |

## 11.7 Simulating a Trial with Stratification variables

The data presented in Appendix I of Kalbfleisch and Prentice (1980) on lung cancer patients were used as a basis for this example. We will design a trial to compare two treatments (Standard and Test) in a target patient group where patients had some prior therapy. The response variable is the survival time in days of lung cancer patients. First, we will create a design for 3 looks, to compare the two treatment groups. Next, using this design, we will carry out simulation with stratification variables. Three covariates in the data are used here as stratum variables: a) type of cancer cell (small, adeno, large, squamous,), b) age in years (<=50, >50), and c) performance status score (<=50, >50 and <=70, >70).

The input data for base design are as follows: Trial type:superiority; test type:2-sided; type I error:0.05; power:0.90; allocation ratio:1; hazard rate (control):0.009211; hazard rate (treatment):0.004114; number of looks:3; Boundary family:spending functions; spending function:Lan-DeMets (OF); subjects are followed:until end of study; subjects accrual rate:12 per day.

The input data for stratified simulation are as given below: The number of stratum variables=3 (cell type; age group; performance status score).

### *11.7.1 Creating the design*

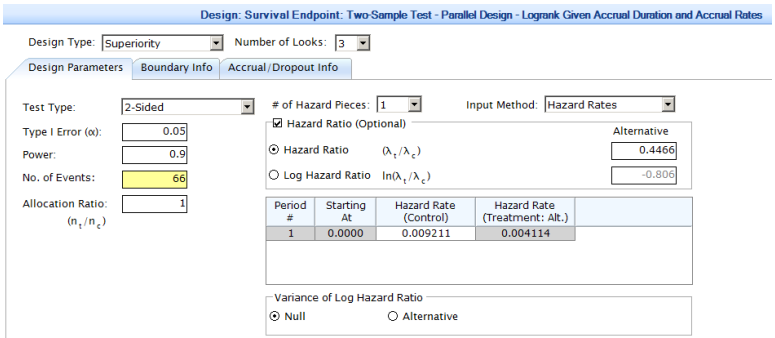
First we will create a design using the input data. Open East, click **Design** tab and then **Time to Event** button in **Survival** group. Now click **Logrank Test: Given Accrual Duration and Accrual Rates**. In the resulting screen, enter the input data in the dialog boxes under the different tabs. Finally click on **Compute** button. Now the dialog boxes under the different tabs will appear as shown below.

The **Design Parameters** tab is shown below, where you can see the computed value of **No.of**

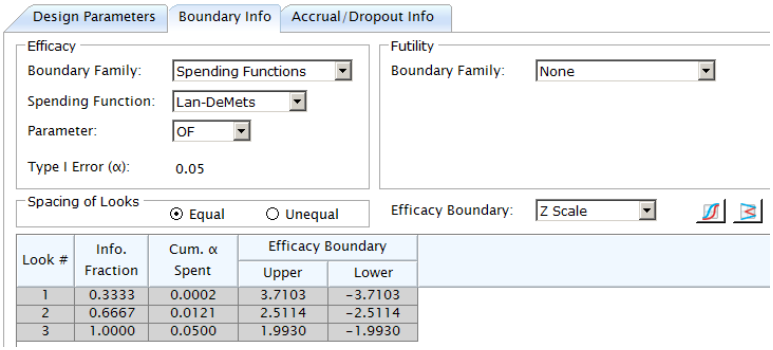
Table 11.4: Input data for stratified simulation

| Cell type | Proportion | Hazard ratio |
|---|---|---|
| small | 0.28 | Baseline |
| adeno | 0.13 | 2.127 |
| large | 0.25 | 0.528 |
| squamous | 0.34 | 0.413 |
| **Age group** | **Proportion** | **Hazard ratio** |
| $\leq$ 50 years | 0.28 | Baseline |
| $>$ 50 years | 0.72 | 0.438 |
| **Performance status score group** | **Proportion** | **Hazard ratio** |
| $\leq$ 50 | 0.43 | Baseline |
| > 50 and $\leq$ 70 | 0.37 | 0.164 |
| > 70 | 0.20 | 0.159 |

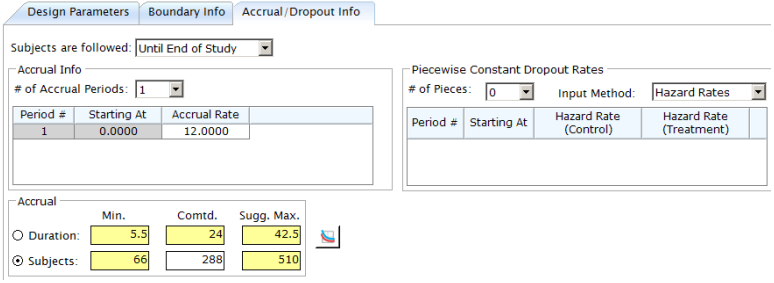
**Events**.



The **Boundary Info** will appear as shown below, where all the input data are seen.




The **Accrual/Dropout Info** tab containing the input data will be as shown below.




After the design is completed and saved in a workbook, select the design node and click on

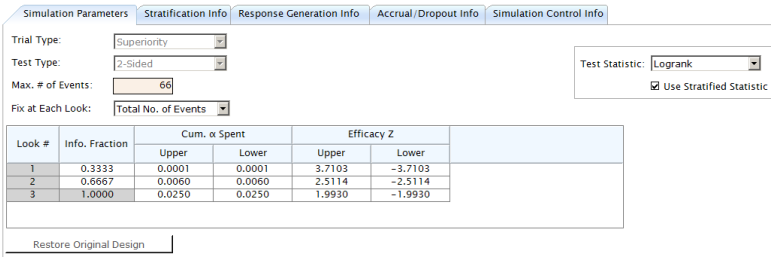
the **output summary** icon to see the following output display.

| | Wbk3:Des1 |
|---|---|
| Mnemonic | SU-2S-LRAR |
| **Test Parameters** | |
| Design Type | Superiority |
| No. of Looks | 3 |
| Test Type | 2-Sided |
| Specified α | 0.05 |
| Power | 0.9023 |
| **Model Parameters** | |
| Hazard Ratio (Alt.) | 0.4466 |
| Var (Log HR) | Null |
| Allocation Ratio (nt/nc) | 1 |
| **Boundary Parameters** | |
| Spacing of Looks | Equal |
| Efficacy Boundary | LD (OF) |
| **Accrual & Dropout Parameters** | |
| Accrual Rate | 12 |
| Subjects are Followed | Until End of Study |
| No. of Accrual Periods | 1 |
| No. of Dropout Pieces | 0 |
| **Sample Size** | |
| Maximum | 288 |
| Expected Under H0 | 287.9915 |
| Expected Under H1 | 288 |
| **Events** | |
| Maximum | 66 |
| Expected Under H0 | 65.7293 |
| Expected Under H1 | 52.8227 |
| **Accrual Duration** | |
| Maximum | 24 |
| Expected Under H0 | 23.9993 |
| Expected Under H1 | 24 |
| **Study Duration** | |
| Maximum | 52.0169 |
| Expected Under H0 | 40.3527 |
| Expected Under H1 | 43.2957 |

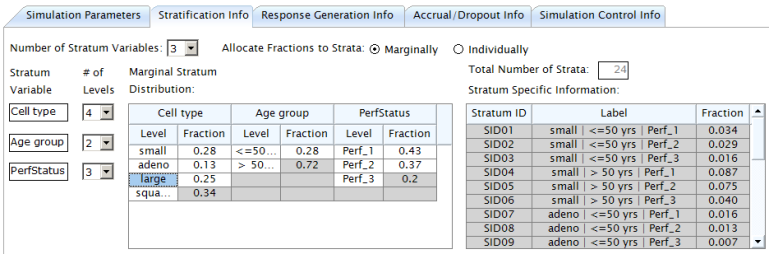
## *11.7.2 Running Stratified Simulation*

After selecting the design node, click on **Simulate** icon. You will see simulation screen with the dialog boxes under different tabs. Click on **Include Options** and select **Stratification Info**.

The dialog box under **Simulation Parameters** will be as shown below. Keep the default test statistic **LogRank** and the default choice of **Use Stratified Statistic**.



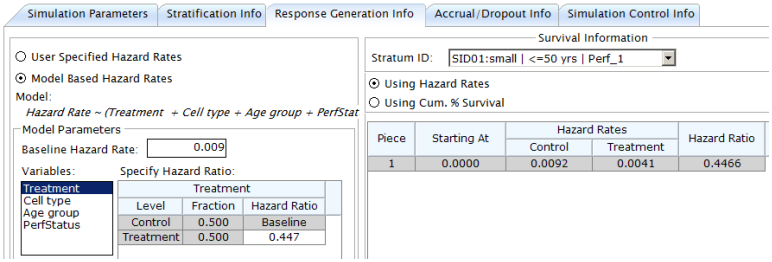

After entering the stratification input information, the dialog box under **Stratification Info** will appear as shown below.




After entering adding response related input information, the dialog box under **Response**

# Chapter 11: *Superiority Trials with Variable Follow-Up*

**Generation Info** will display details as shown in the following screen shots.




Variables:

Treatment
Cell type
Age group
PerfStatus

Specify Hazard Ratio:

| Cell type | | |
|---|---|---|
| Level | Fraction | Hazard Ratio |
| small | 0.280 | Baseline |
| adeno | 0.130 | 2.127 |
| large | 0.250 | 0.528 |
| squamous | 0.340 | 0.413 |

Variables:

Treatment
Cell type
Age group
PerfStatus

Specify Hazard Ratio:

| Age group | | |
|---|---|---|
| Level | Fraction | Hazard Ratio |
| <=50 yrs | 0.280 | Baseline |
| > 50 yrs | 0.720 | 0.438 |

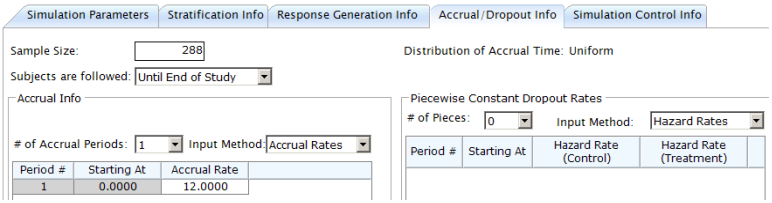
Variables:

Treatment
Cell type
Age group
PerfStatus

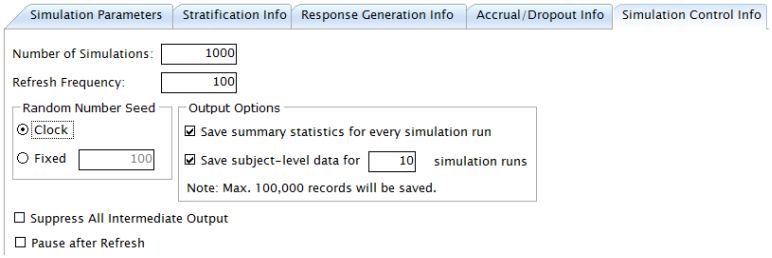
Specify Hazard Ratio:

| PerfStatus | | |
|---|---|---|
| Level | Fraction | Hazard Ratio |
| Perf_1 | 0.430 | Baseline |
| Perf_2 | 0.370 | 0.164 |
| Perf_3 | 0.200 | 0.159 |

The **Accrual/Dropout Info** dialog box will appear as shown below.




In the **Simulation Control Info** tab, specify number of simulations as 1000 and select the choices under output options to save simulation data. The dialog box will appear as shown below.




After clicking on **Simulate** button, the results will appear in the Output Preview row. Click on it and save it in the workbook. Select this simulation node and click on **Output Summary** icon

to see the following stratification simulation output summary.

| | Wbk3:Des1:Sim1 |
|---|---|
| Mnemonic | SU-2S-LRAR |
| **Test Parameters** | |
| Design Type | Superiority |
| Test Type | 2-Sided |
| Test Statistic | Stratified Logrank |
| Power | 0.856 |
| No. of Looks | 3 |
| **Boundary Parameters** | |
| Efficacy Boundary | User Specified |
| Spacing of Looks | User Specified |
| **Accrual & Dropout Parameters** | |
| Followup Duration | Until End of Study |
| Accrual Rate | 12 |
| No. of Accrual Periods | 1 |
| **Sample Size** | |
| Maximum | 288 |
| **Events** | |
| Maximum | 66 |
| **Study Duration** | |
| Maximum | 172.4444 |
| **Stratum Information** | |
| No. of Stratum Variables | 3 |
| No. of Strata | 24 |
| Allocate Fractions to Strata | Marginally |
| Specification of Hazard Rates | Model based |

The stratified simulation results show that the attained power 0.856 is slightly less than the design specified power of 0.90.

# 12 Non-Inferiority Trials Given Accrual Duration and Accrual Rates

This chapter will illustrate through a worked example how to design, monitor and simulate a two-sample non-inferiority trial with a time-to-event trial endpoint, when the accrual duration and accrual rates are fixed.

## 12.1 Establishing the Non-Inferiority Margin

The first step in designing a non-inferiority trial is to establish a suitable non-inferiority margin. This is typically done by performing a meta-analysis on past clinical trials of the active control versus placebo. Regulatory agencies then require the sponsor of the clinical trial to demonstrate that a fixed percentage of the active control effect (usually 50%) is retained by the new treatment. A further complication arises because the active control effect can only be estimated with error. We illustrate below with an example provided by reviewers at the FDA.

Rothman et al. (2003) have discussed a clinical trial to establish the non-inferiority of the test drug Xeloda (treatment $t$) relative to the active control (treatment $c$) consisting of 5-fluorouracil with leucovarin (5FU+LV) for metastatic colorectal cancer. In order to establish a suitable non-inferiority margin for this trial it is necessary to first establish the effect of 5FU+LV relative to the reference therapy of 5FU alone (treatment $p$, here regarded as placebo). To establish this effect the FDA conducted a ten-study random effects meta-analysis (FDA Medical-Statistical review for Xeloda, NDA 20-896, April 2001) of randomized comparisons of 5-FU alone versus 5-FU+LV. Letting $\lambda_t, \lambda_c$ and $\lambda_p$ denote the constant hazard rates for the new treatment, the active control and the placebo, respectively, the FDA meta-analysis established that

$$\ln\left(\widehat{\lambda_p/\lambda_c}\right) = 0.234$$

with standard error

$$\text{se}[\ln(\widehat{\lambda_p/\lambda_c})] = 0.075 \ .$$

Thus with $100\gamma$% confidence the active control effect lies inside the interval

$$[0.234 - 0.075\Phi^{-1}(\frac{1+\gamma}{2}), 0.234 + 0.075\Phi^{-1}(\frac{1+\gamma}{2})] \tag{12.1}$$

The new study is required to demonstrate that some fraction (usually 50%) of the active control effect is retained. Rothman et al. (2003) state that the claim of non-inferiority for the new treatment relative to the active control can be demonstrated if the upper limit of a two-sided $100(1-\alpha)$% confidence interval for $\ln(\lambda_t/\lambda_c)$ is less than a pre-specified fraction of the lower limit of a two-sided $100\gamma$% confidence interval for the active control effect established by the meta-analysis. This is known as the "two confidence intervals procedure". Specifically in order to claim non-inferiority in the current trial it is necessary to show that

$$\ln(\widehat{\lambda_t/\lambda_c}) + \Phi^{-1}(1-\alpha/2)\text{se}[\ln(\widehat{\lambda_t/\lambda_c})] < (1-f_0)\{\ln(\widehat{\lambda_p/\lambda_c}) - \Phi^{-1}(\frac{1+\gamma}{2})\text{se}[\ln(\widehat{\lambda_p/\lambda_c})]\} \ . \tag{12.2}$$

We may re-write the non-inferiority condition (12.2) in terms of a one-sided Wald test of the form

$$\frac{\ln(\widehat{\lambda_t/\lambda_c}) - \delta_0}{\text{se}[\ln(\widehat{\lambda_t/\lambda_c})]} < \Phi^{-1}(1-\alpha/2) \ , \tag{12.3}$$

where

$$\delta_0 = (1-f_0)\{\ln(\widehat{\lambda_p/\lambda_c}) - \Phi^{-1}(\frac{1+\gamma}{2})\text{se}[\ln(\widehat{\lambda_p/\lambda_c})]\} \tag{12.4}$$

is the non-inferiority margin.

The choice $f_0 = 1$ implies that the entire active control effect must be retained in the new trial and amounts to running a superiority trial. At the other end of the spectrum, the choice $f_0 = 0$ implies that none of the active control effect need be retained; i.e., the new treatment is only required to demonstrate effectiveness relative to placebo. The usual choice is $f_0 = 0.5$, implying that the new treatment is required to retain at least 50% of the active control effect. The usual choice for $\alpha$ is $\alpha = 0.05$. A conservative choice for the coefficient $\gamma$ is $\gamma = (1-\alpha) = 0.95$. Rothman et al. (2003) refer to this method of establishing the non-inferiority margin as the "two 95 percent two-sided confidence interval procedure" or the

"95-95 rule". In general this approach leads to rather tight margins unless the active control effect is substantial. Rothman et al. (2003) have also proposed more lenient margins that vary with the amount of power desired. Fleming (2007), however, argues for the stricter 95-95 rule on the grounds that it offers greater protection against an ineffective medical compound being approved in the event that the results of the previous trials used to establish the active control effect are of questionable relevance to the current setting. Accordingly we evaluate (12.4) with $\gamma = 0.95$, $f_0 = 0.5$, $\ln\left(\widehat{\lambda_p/\lambda_c}\right) = 0.234$ and $\text{se}[\ln\left(\widehat{\lambda_p/\lambda_c}\right)] = 0.075$ thereby obtaining the non-inferiority margin to be $\delta_0 = 0.044$ for the log hazard ratio and $\exp(0.044) = 1.045$ for the hazard ratio.

## 12.2 Design of Metastatic Colorectal Cancer Trial

▪ 12.2.1 Single-Look Design ▪ 12.2.2 Early Stopping for Futility

In this section we will use East to design a single-look non-inferiority trial comparing the test drug Xeloda (treament $t$) to the active control 5FU+LV (treatment $c$) for the treatment of metastatic colorectal cancer. On the basis of a meta-analysis of ten previous studies of the active control versus placebo (Rothman et al., 2003), a non-inferiority margin of 1.045 for $\lambda_t/\lambda_c$ has been established. Thus we are interested in testing the null hypothesis of inferiority $H_0$: $\lambda_t/\lambda_c \geq 1.045$ versus the one-sided alternative hypothesis that $\lambda_t/\lambda_c < 1.045$. Subjects are expected to enroll at the rate of 60/month and the median survival time for patients randomized to the active control arm is expected to be 18 months.

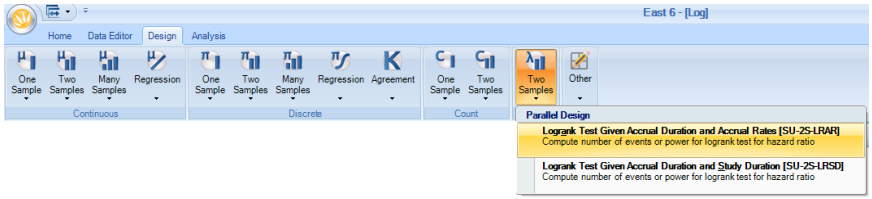
### 12.2.1 Single-Look Design

We will use East to create an initial single-look design having 80% power to detect the alternative hypothesis $H_1$: $\lambda_t/\lambda_c = 1$ with a one sided level 0.025 non-inferiority test.

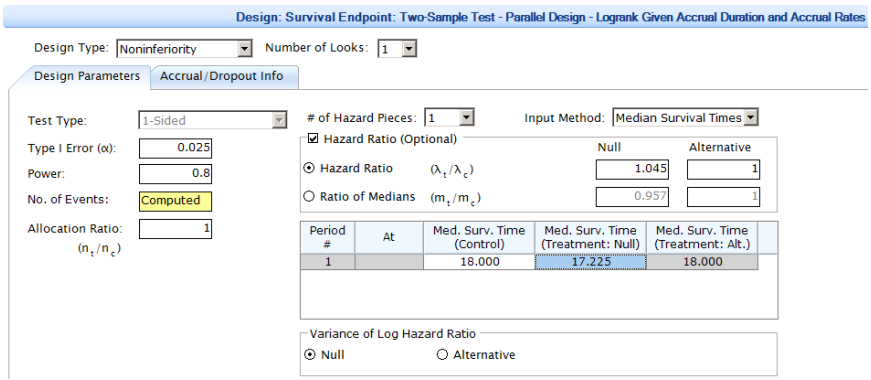
To begin click **Survival: Two Samples** on the **Design** tab and then click **Parallel Design: Log**

# Chapter 12: *Non-Inferiority Trials Given Accrual Duration and Accrual Rates*

**Rank Test Given Accrual Duration and Accrual Rates**.



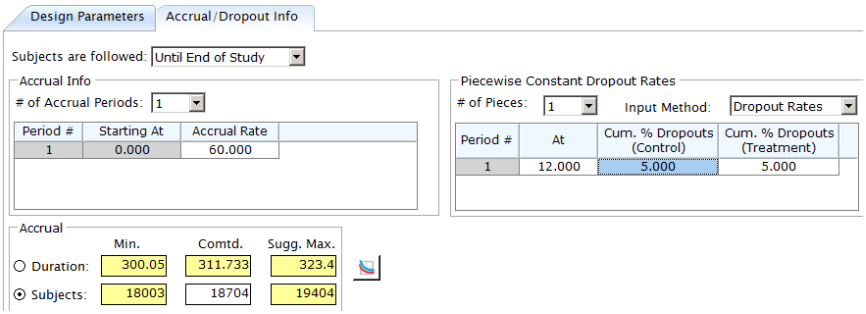

A new screen will appear. Enter the appropriate design parameters into the dialog box as shown below.




The box labeled **Variance of Log Hazard Ratio** specifies whether the calculation of the required number of events is to be based on the variance estimate of the log hazard ratio under the null hypothesis or the alternative hypothesis. The default choice in East is **Null**. Most textbooks recommend this choice as well (see, for example Collett, 1994, equation (2.21) specialized to no ties). It will usually not be necessary to change this default. For a technical discussion of this issue refer to Appendix **??**, Section **??**.

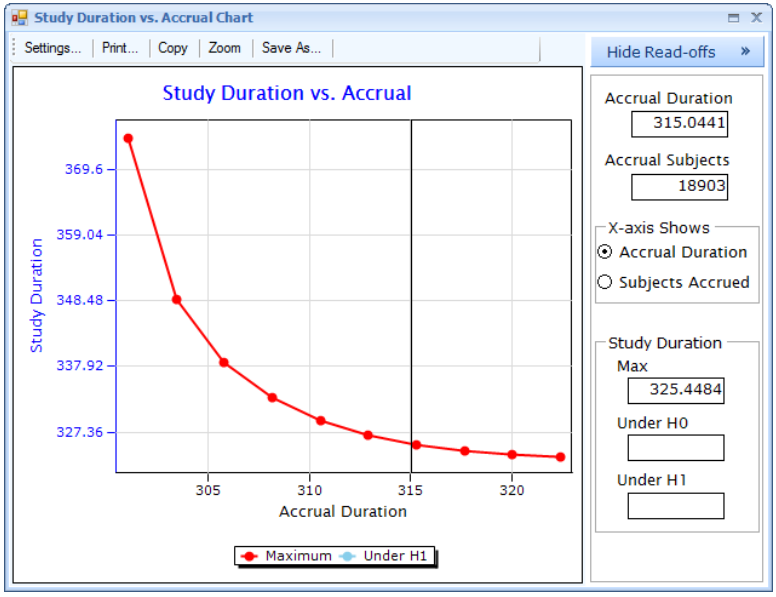
Next click on the **Accrual/Dropout Info** tab. Here we will specify the accrual information and dropout rates. Enter an accrual rate of 60. Suppose that there are 5% drop-outs per year in

each arm. Enter these values as shown below.




On the bottom of this screen is where you can specify the accrual duration or number of subjects. East automatically computes a range that is necessary to achieve the desired power of the study and selects the midpoint of the range, as the committed accrual duration or subjects. If your study has a restriction on accrual duration or subject accrual, you may enter this value in the **Comtd.** column. In our example, East computes a minimum accrual duration of 300.05 months and a suggested maximum of 323.4 months. Also, if you click the  icon a chart which shows the relationship between accrual duration (or subject accrual, depending

on whether you choose to specify accrual duration or subject accrual) and study duration.




Looking at this chart, choosing an accrual duration longer that 315 months will not result in a substantial decrease in study duration. Thus, we commit to an accrual duration of 315 months. Close this chart, select the radio button next to **Duration** and enter 315 in the **Comtd.** column.

Click on **Compute** to complete the design. The design is shown as a row in the **Output Preview** located in the lower pane of this window. You can select this design by clicking anywhere along the row in the **Output Preview**. With Des1 selected, click the icon to display the details of this design in the upper pane, which are shown below. You may also wish to save this design. Select Des1 in the **Output Preview** window and click the to save

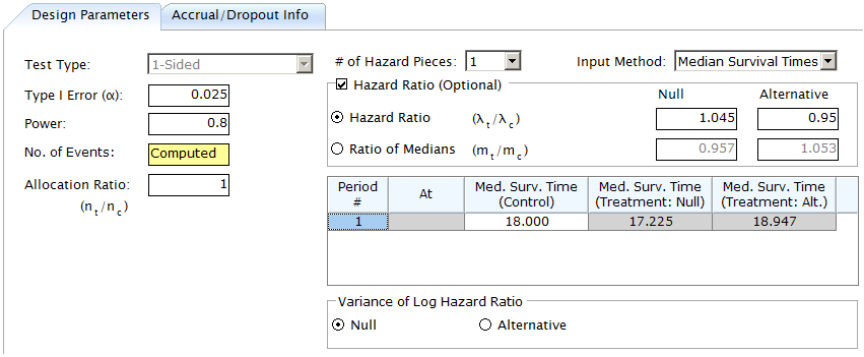
this design to Workbook1 in the **Library**.

| | Wbk1:Des1 |
|---|---|
| Mnemonic | SU-2S-LRAR |
| **Test Parameters** | |
| Design Type | Noninferiority |
| No. of Looks | 1 |
| Test Type | 1-Sided |
| Specified α | 0.025 |
| Power | 0.8 |
| **Model Parameters** | |
| Hazard Ratio (Null) | 1.045 |
| Hazard Ratio (Alt.) | 1 |
| Var (Log HR) | Null |
| Allocation Ratio (nt/nc) | 1 |
| **Accrual & Dropout Parameters** | |
| Accrual Rate | 60 |
| Subjects are Followed | Until End of Study |
| No. of Accrual Periods | 1 |
| No. of Dropout Pieces | 1 |
| **Sample Size** | |
| Maximum | 18900 |
| Expected Under H0 | 18900 |
| Expected Under H1 | 18900 |
| **Events** | |
| Maximum | 16205 |
| Expected Under H0 | 16205 |
| Expected Under H1 | 16205 |
| **Accrual Duration** | |
| Maximum | 315 |
| Expected Under H0 | 315 |
| Expected Under H1 | 315 |
| **Study Duration** | |
| Maximum | 325.466 |
| Expected Under H0 | 323.839 |
| Expected Under H1 | 325.466 |

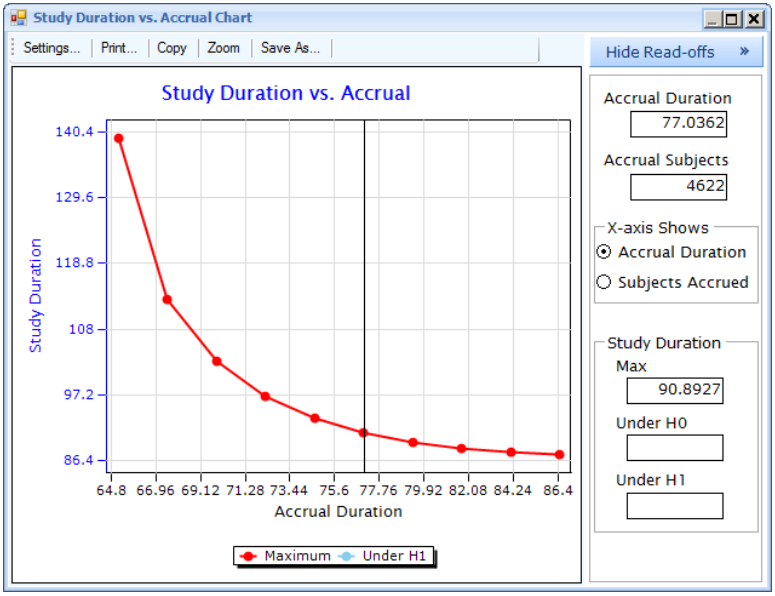
It is immediately evident that Des1 is untenable. It requires 16,205 events to be fully powered. The problem lies with trying to power the trial to detect a hazard ratio of 1 under the alternative hypothesis. Suppose instead that the investigators actually believe that the treatment is slightly superior to the active control, but the difference is too small to be

detected in a superiority trial. In that case a non-inferiority design powered at a hazard ratio less than 1 (0.95, say) would be a better option because such a trial would require fewer events.

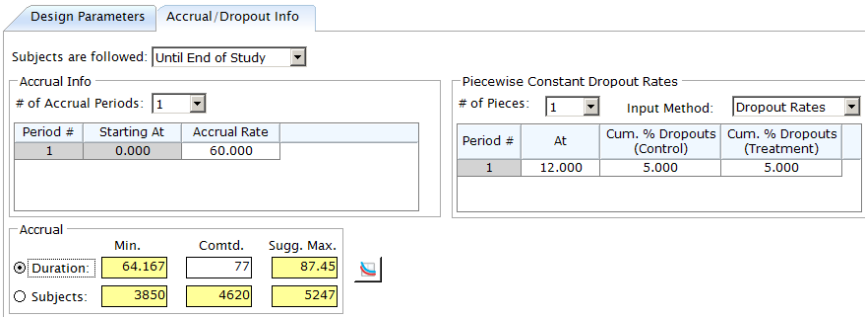
To see this create a new design by selecting Des1 in the **Library**, and clicking the  icon on the **Library** toolbar. Then edit this design by specifying a hazard ratio of 0.95 under the alternative hypothesis as shown below.




Next, click on the **Accrual/Dropout Info** tab. Notice that the minimum and suggested maximum accrual have changed to 64.167 and 87.45 months, respectively. Click the  icon to display the study duration versus accrual chart.




Suppose that after examining this chart, you decide that an accrual duration longer than 77 months is not worth the small decrease in study duration one would gain from a longer accrual duration. Close this chart. Select the radio button next to **Duration** and enter 77 in the **Comtd.** column.




Click the **Compute** button to generate output for Des2. With Des2 selected in the **Output**

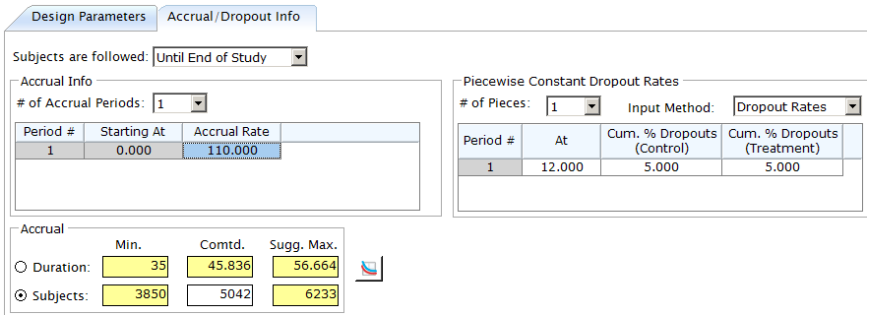
**Preview**, click the  icon to save Des2 to the **Library**. In the **Library**, select the rows for Des1 and Des2, by holding the Ctrl key, and then click the  icon. The upper pane will display the details of the two designs side-by-side:

| | Wbk1:Des1 | Wbk1:Des2 |
|---|---|---|
| Mnemonic | SU-2S-LRAR | SU-2S-LRAR |
| **Test Parameters** | | |
| Design Type | Noninferiority | Noninferiority |
| No. of Looks | 1 | 1 |
| Test Type | 1-Sided | 1-Sided |
| Specified α | 0.025 | 0.025 |
| Power | 0.8 | 0.8 |
| **Model Parameters** | | |
| Hazard Ratio (Null) | 1.045 | 1.045 |
| Hazard Ratio (Alt.) | 1 | 0.95 |
| Var (Log HR) | Null | Null |
| Allocation Ratio (nt/nc) | 1 | 1 |
| **Accrual & Dropout Parameters** | | |
| Accrual Rate | 60 | 60 |
| Subjects are Followed | Until End of Study | Until End of Study |
| No. of Accrual Periods | 1 | 1 |
| No. of Dropout Pieces | 1 | 1 |
| **Sample Size** | | |
| Maximum | 18900 | 4620 |
| Expected Under H0 | 18900 | 4620 |
| Expected Under H1 | 18900 | 4620 |
| **Events** | | |
| Maximum | 16205 | 3457 |
| Expected Under H0 | 16205 | 3457 |
| Expected Under H1 | 16205 | 3457 |
| **Accrual Duration** | | |
| Maximum | 315 | 77 |
| Expected Under H0 | 315 | 77 |
| Expected Under H1 | 315 | 77 |
| **Study Duration** | | |
| Maximum | 325.466 | 90.946 |
| Expected Under H0 | 323.839 | 88.962 |
| Expected Under H1 | 325.466 | 90.946 |

Des2 is clearly easier to implement than Des1. It requires only 3,457 events and 4620 subjects to be fully powered. Also note the marked decrease in study duration under either the null or alternative hypothesis. Nevertheless, Des2 is also unsatisfactory. The maximum study duration for Des2 (accrual plus follow-up) is 90.9 months with 77 months of that amount of time being utilized to enroll 4620 patients. It is necessary to shorten the maximum study duration further.

One possible way to shorten the maximum study duration is to increase the rate of enrollment. Suppose that additional sites can be enlisted to enroll patients after the study is activated so that six months later the average rate of enrollment is increased to 110/month. To see the impact of the increased rate of enrollment select Des2 in the **Library**, and click on the icon on the **Library** toolbar.

Next, click on the **Accrual/Dropout Info** tab. Change the accrual rates as shown below.




Notice how East automatically updates the accrual duration and subject accrual. An accrual duration in the range of 35 to 56.664 months is sufficient to achieve the desired power. Suppose that after examining the study duration versus accrual chart, we decide on an accrual duration of 49 months. Enter 49 in the **Comtd.** column.

Click the **Compute** button to generate output for Des3. With Des3 selected in the **Output Preview**, click the icon to save Des3 to the **Library**. In the **Library**, select the rows for Des1, Des2, and Des3 by holding the Ctrl key, and then click the icon. The upper pane

will display the details of the three designs side-by-side:

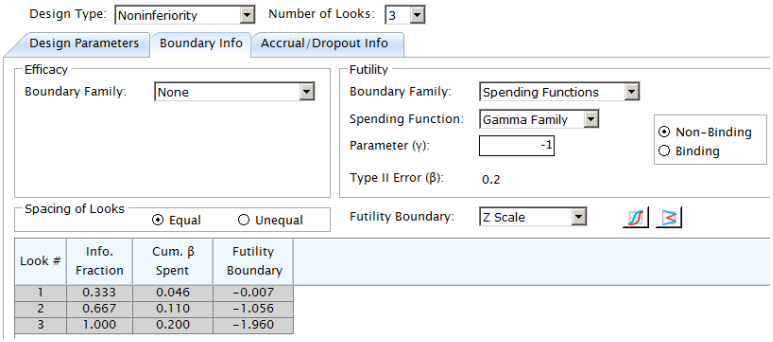
| | Wbk1:Des1 | Wbk1:Des2 | Wbk1:Des3 |
|---|---|---|---|
| Mnemonic | SU-2S-LRAR | SU-2S-LRAR | SU-2S-LRAR |
| **Test Parameters** | | | |
| Design Type | Noninferiority | Noninferiority | Noninferiority |
| No. of Looks | 1 | 1 | 1 |
| Test Type | 1-Sided | 1-Sided | 1-Sided |
| Specified α | 0.025 | 0.025 | 0.025 |
| Power | 0.8 | 0.8 | 0.8 |
| **Model Parameters** | | | |
| Hazard Ratio (Null) | 1.045 | 1.045 | 1.045 |
| Hazard Ratio (Alt.) | 1 | 0.95 | 0.95 |
| Var (Log HR) | Null | Null | Null |
| Allocation Ratio (nt/nc) | 1 | 1 | 1 |
| **Accrual & Dropout Parameters** | | | |
| Accrual Rate | 60 | 60 | 110 |
| Subjects are Followed | Until End of Study | Until End of Study | Until End of Study |
| No. of Accrual Periods | 1 | 1 | 1 |
| No. of Dropout Pieces | 1 | 1 | 1 |
| **Sample Size** | | | |
| Maximum | 18900 | 4620 | 5390 |
| Expected Under H0 | 18900 | 4620 | 5390 |
| Expected Under H1 | 18900 | 4620 | 5390 |
| **Events** | | | |
| Maximum | 16205 | 3457 | 3457 |
| Expected Under H0 | 16205 | 3457 | 3457 |
| Expected Under H1 | 16205 | 3457 | 3457 |
| **Accrual Duration** | | | |
| Maximum | 315 | 77 | 49 |
| Expected Under H0 | 315 | 77 | 49 |
| Expected Under H1 | 315 | 77 | 49 |
| **Study Duration** | | | |
| Maximum | 325.466 | 90.946 | 58.523 |
| Expected Under H0 | 323.839 | 88.962 | 57.158 |
| Expected Under H1 | 325.466 | 90.946 | 58.523 |

Des3 also requires 3457 events. However, because of the faster rate of enrollment the time that it takes to obtain these events is cut down to 58.5 months.

### *12.2.2 Early Stopping for Futility*

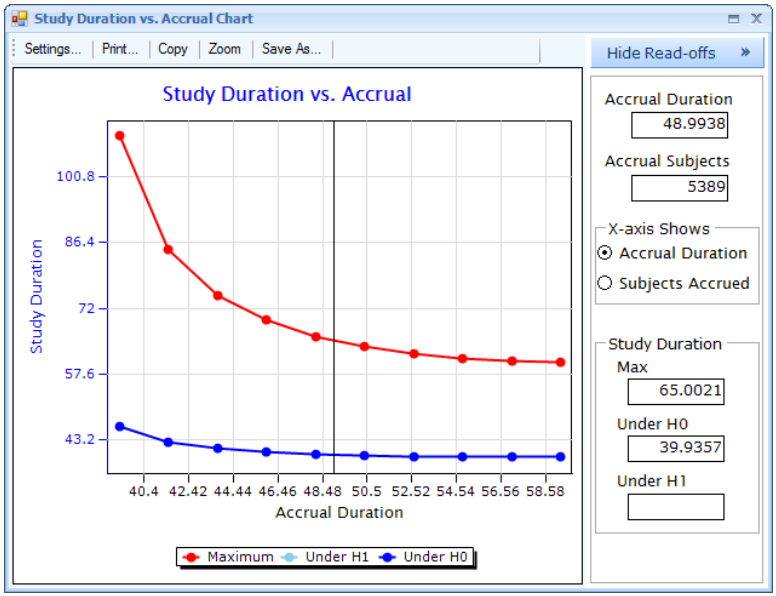
Under the null hypothesis Des3, with 3457 events, has an expected study duration of 57.2 months. This is a very long time commitment for a trial that is unlikely to be successful. Therefore it would be a good idea to introduce a futility boundary for possible early stopping. Since we wish to be fairly aggressive about early stopping for futility we will generate the futility boundary from the Gamma$(-1)$ $\beta$-spending function. On the other hand, since there is no interest in early stopping for efficacy, we will not use an efficacy boundary.

Create a new design by selecting Des3 in the **Library**, and clicking the icon on the **Library** toolbar. Change the number of looks from 1 to 3. Next, click on the **Boundary Info** tab. Enter the parameters as shown below. Be sure to select the **Non-Binding** option. This choice gives us the flexibility to continue the trial even if a futility boundary has been crossed. Data monitoring committees usually want this flexibility; for example, to follow a secondary endpoint.




Next click on the **Accrual/Dropout Info** tab. Once again, East automatically computes the minimum and suggested maximum values for the accrual duration and subject accrual. Click the icon to display the study duration versus accrual chart. Notice that another line is added to the chart. Now, we can see the maximum study duration vs accrual under the null

hypothesis.




Suppose that after examining this chart, you decide to set the accrual duration at 49 months. Any increase in accrual duration past 49 months will not result in a substantial decrease in study duration. Close this chart. Select the radio button for **Duration** and enter 49 in the **Comtd.** column.

Click the **Compute** button to generate output for Des4. With Des4 selected in the **Output Preview**, click the  icon to save Des4 to the **Library**. In the **Library**, select the rows for Des3 and Des4 by holding the Ctrl key, and then click the  icon. The upper pane will

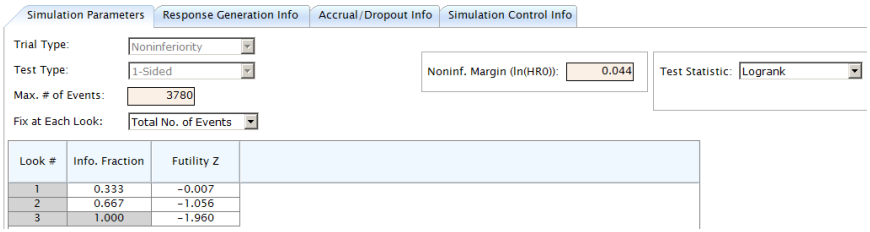
display the details of the two designs side-by-side:

| | Wbk1:Des3 | Wbk1:Des4 |
|---|---|---|
| Mnemonic | SU-2S-LRAR | SU-2S-LRAR |
| **Test Parameters** | | |
| Design Type | Noninferiority | Noninferiority |
| No. of Looks | 1 | 3 |
| Test Type | 1-Sided | 1-Sided |
| Specified α | 0.025 | 0.025 |
| Attained α | | 0.022 |
| Power | 0.8 | 0.8 |
| **Model Parameters** | | |
| Hazard Ratio (Null) | 1.045 | 1.045 |
| Hazard Ratio (Alt.) | 0.95 | 0.95 |
| Var (Log HR) | Null | Null |
| Allocation Ratio (nt/nc) | 1 | 1 |
| **Boundary Parameters** | | |
| Futility Boundary | | Gm (-1) (NB) |
| Spacing of Looks | | Equal |
| **Accrual & Dropout Parameters** | | |
| Accrual Rate | 110 | 110 |
| Subjects are Followed | Until End of Study | Until End of Study |
| No. of Accrual Periods | 1 | 1 |
| No. of Dropout Pieces | 1 | 1 |
| **Sample Size** | | |
| Maximum | 5390 | 5390 |
| Expected Under H0 | 5390 | 4140.262 |
| Expected Under H1 | 5390 | 5271.053 |
| **Events** | | |
| Maximum | 3457 | 3780 |
| Expected Under H0 | 3457 | 2056.327 |
| Expected Under H1 | 3457 | 3583.036 |
| **Accrual Duration** | | |
| Maximum | 49 | 49 |
| Expected Under H0 | 49 | 37.639 |
| Expected Under H1 | 49 | 47.919 |
| **Study Duration** | | |
| Maximum | 58.523 | 64.893 |
| Expected Under H0 | 57.158 | 39.546 |
| Expected Under H1 | 58.523 | 62.059 |

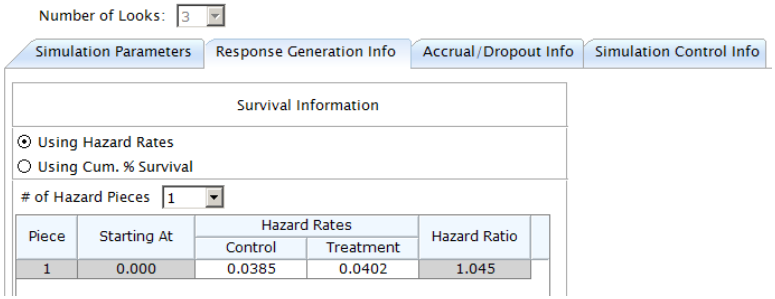
Observe that while the maximum study duration has been inflated by about 6 months compared to Des3, the expected study duration under $H_0$ has been cut down by almost 18 months.

It would be useful to simulate Des4 under a variety of scenarios for the hazard ratio. Select Des4 in the **Library** and click the  icon. You will be taken to the following simulation

worksheet.




We wish to simulate this trial under the null hypothesis that the hazard ratio is $\exp(0.044) = 1.045$. To this end click on the **Response Generation Info** tab. In this tab change the control and treatment hazard rates as shown below.




Next, click the **Simulate** button to simulate 10000 trials. A new row labeled Sim1 will appear in the **Output Preview** window. Select Sim1 in the **Output Preview** and click the  icon to save it to the **Library**. In the **Library**, double-click Sim1. A portion of the output is displayed

below. (The actual values may differ, depending on the starting seed used).

**Simulation: Survival Endpoint: Two-Sample Test - Parallel Design - Logrank Given Accrual Duration and Accrual Rates**

| Simulation Parameters | |
|---|---|
| Simulation ID: | Sim1 |
| Design Type: | Noninferiority |
| Number of Looks: | 3 |
| Test Type: | 1-Sided |
| Fix at Each Look: | Total No. of Events |
| Noninferiority Margin (ln(HR0)): | 0.044 |
| Test Statistic: | Logrank |
| Average Events: | 2061.486 |
| Total Accrual Duration: | 49 |
| Avg. Power at Termination: | 0.0205 |
| **Simulation Control Parameters** | |
| Starting Seed: | Clock |
| Number of Simulations: | 10000 |

**Simulation Boundaries and Boundary Crossing Probabilities:**

| Look # | Events | Boundaries | | Early Stopping For | Unable To Reject H1 | Total Simulations | |
|---|---|---|---|---|---|---|---|
| | | Futility | | | | | |
| | | Upper | | Futility | | Count | % |
| 1 | 1260 | -0.007 | | 5008 | | 5008 | 50.08 |
| 2 | 2520 | -1.0555 | | 3623 | | 3623 | 36.23 |
| 3 | 3780 | -1.96 | | 1164 | 205 | 1369 | 13.69 |
| Total | | | | 9795 | 205 | 10000 | |
| % | | | | 97.95 | 2.05 | | |

**Average Sample Size, Dropouts and Look Times:**

| Look # | Average Sample Size | Average Events | | Average Dropouts | | Average Look Time | Average Follow up |
|---|---|---|---|---|---|---|---|
| | | Control | Treatment | Control | Treatment | | |
| 1 | 3213.0046 | 620.4686 | 639.5314 | 68.9234 | 67.9725 | 29.2054 | 9.9645 |
| 2 | 4962.6635 | 1249.6314 | 1270.3686 | 137.1993 | 136.4663 | 45.1094 | 12.8999 |
| 3 | 5390 | 1883.5486 | 1896.4514 | 205.3798 | 206.0351 | 63.168 | 17.8147 |
| Average | 4144.9649 | 1014.1761 | 1047.3099 | 112.5943 | 111.2671 | 39.6178 | 12.1053 |

**Response Generation Parameters**

No. of Hazard Pieces: 1
Input Method: Hazard Rates

| Piece # | Starting At | Control | Treatment | Hazard Ratio |
|---|---|---|---|---|
| 1 | 0 | 0.039 | 0.04 | 1.045 |

**Accrual/Dropout Parameters**

Sample Size: 5390
Subjects are Followed: Until End of Study
Accrual Input Method: Accrual Rates

| Period # | Starting At | Accrual Rate |
|---|---|---|
| 1 | 0 | 110 |

No. of Dropout Pieces: 1
Dropout Input Method: Dropout Rates

| Period # | At | Control | Treatment |
|---|---|---|---|
| 1 | 12 | 5 | 5 |

**Overall Simulation Results**

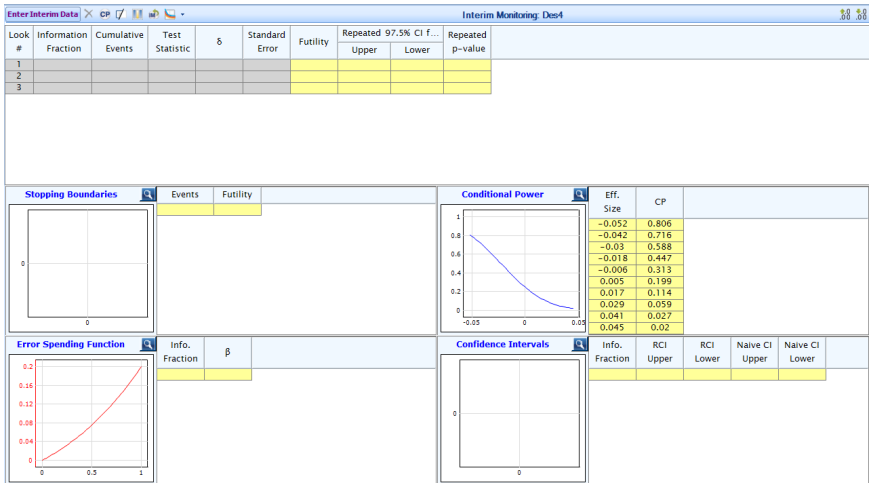
Average Study Duration: 39.618
Starting Seed: 63985542
Total Number of Simulations: 10000
Elapsed Time: 00:00:56

Note that 205 out of the 10000 simulations were unable to reject the alternative hypothesis, thus confirming (up to Monte Carlo accuracy) that this design achieves a type-1 error of $2.5\%$. Also, observe that 50.08% of these trials have crossed the futility boundary at the very first interim look after only 29.205 months of study duration.

## 12.3 Interim Monitoring

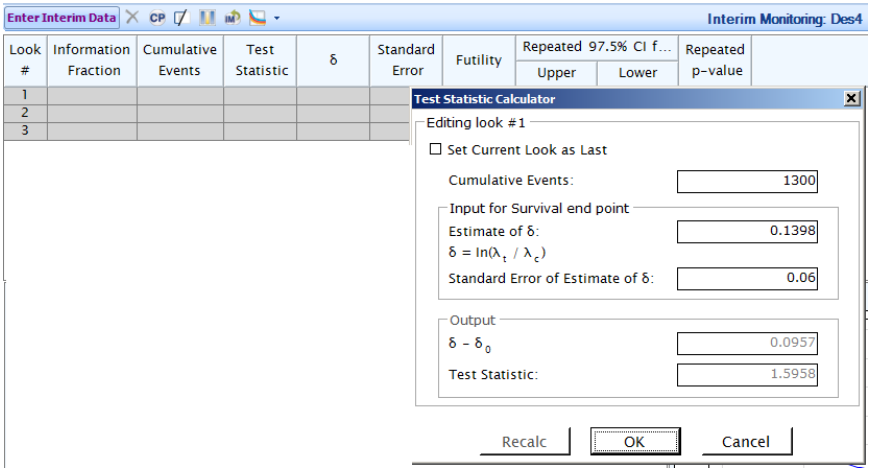
Suppose we have adopted Des4. Let us monitor the trial with the help of the Interim Monitoring Worksheet. Select Des4 in the **Library**, and click the IM icon from the Library toolbar. Alternatively, right-click on Des4 and select **IM Dashboard**. The interim monitoring dashboard contains various controls for monitoring the trial, and is divided into two sections. The top section contains several columns for displaying output values based on the interim

inputs. The bottom section contains four charts, each with a corresponding table to its right. These charts provide graphical and numerical descriptions of the progress of the clinical trial and are useful tools for decision making by a data monitoring committee.

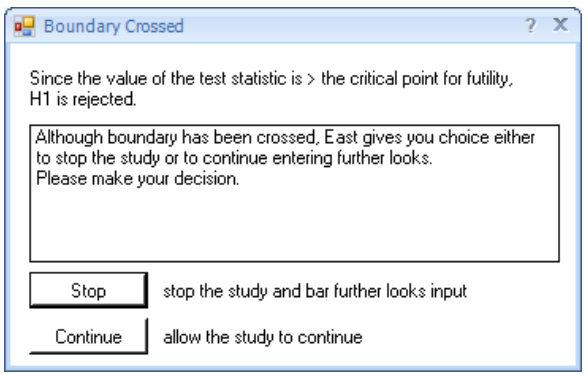



Suppose that the first interim look is taken after observing 1300 events. The observed hazard ratio is 1.15 and the standard error of the log hazard ratio is 0.06. Enter this information into the interim monitoring worksheet using Test Statistic calculator. Click on Enter Interim Data and

enter the data in the test statistic calculator as shown below.



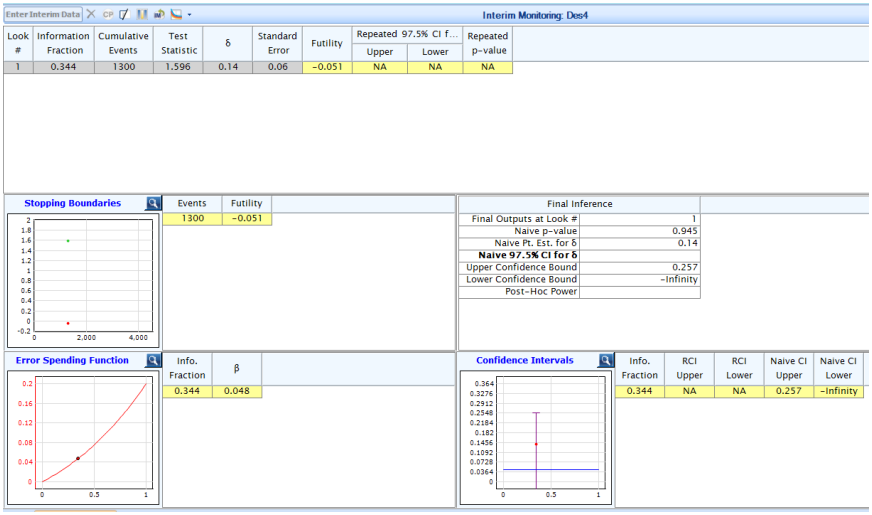

Next, click **OK**. East will indicate that the $H_1$ (futility) boundary has been crossed and hence, the alternative hypothesis of non-inferiority is rejected in favor of the null hypothesis of inferiority.




Click the **Stop** button to terminate the trial. You will see the IM sheet output including Final

## Chapter 12: *Non-Inferiority Trials Given Accrual Duration and Accrual Rates*

Inference details as shown below.




Observe that the upper 97.5% confidence bound for $\delta$, 0.257, is above the non-inferiority margin of 0.044 (on the log hazard ratio scale).

# 13 Superiority Trials Given Accrual Duration and Study Duration

This chapter will illustrate through a worked example how to design and simulate a two-sample superiority trial with a time-to-event trial endpoint, where the accrual duration and study duration are constrained. Most trials in the pharmaceutical industry setting are designed in this manner, time being a more rigid constraint than the accrual rate of patients. The duration of a clinical trial impacts the duration of a drug development program, and thus time to market and potential revenues. Therefore it is of interest to fix the study duration as well as the accrual duration to finish the clinical trial according to schedule. The option to design a trial in this way is available in East.

## 13.1 Calculating a Sample Size

For this design, East obtains the maximum number of events $D_{max}$ from the maximum information $I_{max}$, as described in Appendix sections **??** and **??**. To calculate the sample size, we first equate the expected number of events $d(S_a + S_f)$ (as calculated in Appendix **??** which depends on the accrual duration ($S_a$) and the duration of follow-up ($S_f$) to the maximum number of events $D_{max}$.

$$d(S_a + S_f) = D_{max} \tag{13.1}$$

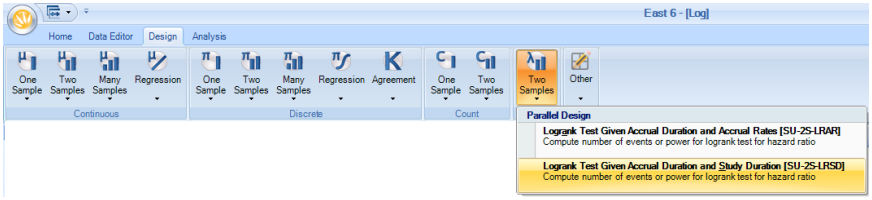
In this type of design the accrual duration $S_a$ and the study duration $S_a + S_f$ are given as input. East iterates between sample sizes, increasing onwards from a minimum value of $D_{max}$, enrolled over a duration of $S_a$ until $D_{max}$ events are found to occur within a study duration of $S_a + S_f$. The result is the unique sample size required to obtain the proper power for the study.

## 13.2 The RALES Clinical Trial: Initial Design

The RALES trial (Pitt et. al., 1999) was a double blind study of aldosterone-receptor blocker spironolactone at a daily dose of 25 mg in combination with standard doses of an ACE inhibitor (treatment arm) versus standard therapy of an ACE inhibitor (control arm) in patients who had severe heart failure as a result of systolic left ventricular dysfunction. The primary endpoint was death from any cause. Six equally-spaced looks at the data using the Lan-DeMets-O'Brien-Fleming spending function were planned. The trial was designed to detect a hazard ratio of 0.83 with 90% power at a two-sided 0.05 level of significance. The hazard rate of the control arm was estimated to be 0.38.

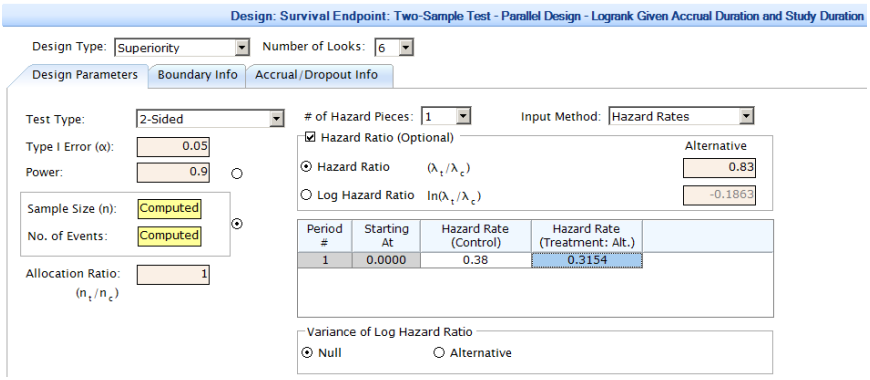
Randomization was scheduled to begin in March 1995 and complete in December 1996 for a total of 1.8 years of enrollment. Follow-up was planned through December 1999, so that the total study duration from first patient enrolled to last patient visit should be 4.8 years.

We begin by using East to design RALES under these basic assumptions. To begin click **Survival: Two Samples** on the **Design** tab and then click **Parallel Design: Logrank Test Given Accrual Duration and Study Duration** as shown below



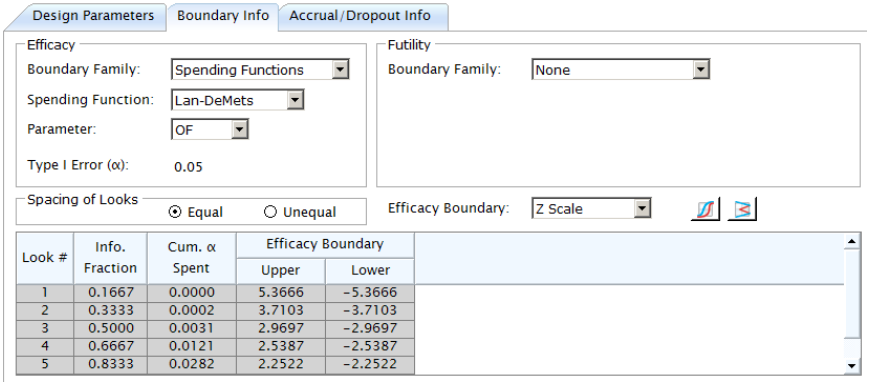

A new screen will appear. Enter the appropriate design parameters into the dialog box as

shown below.

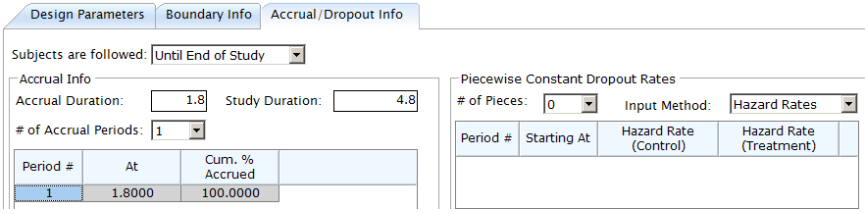



The box labeled **Variance of Log Hazard Ratio** specifies whether the calculation of the required number of events is to be based on the variance estimate of the log hazard ratio under the null hypothesis or the alternative hypothesis. The default choice in East is **Null**. Most textbooks recommend this choice as well (see, for example Collett, 1994, equation (2.21) specialized to no ties). It will usually not be necessary to change this default. For a technical discussion of this issue refer to Appendix **??**, Section **??**.

Next, click on the **Boundary Info** tab. We will take six equally spaced looks at the data using the Lan-DeMets O'Brien-Fleming spending function. These are the default setting in East.



Note that we do not select a futility boundary in this case. Next click on the **Accrual/Dropout Info** tab. Here we will specify the accrual information and dropout rates. The software allows a specification of piecewise constant hazards and variable accrual rates but we start by looking at an example that does not require any of these options. In the drop-down menu next to **Subjects are followed:** select `Until End of Study`. Set the **Accrual Duration** to 1.8 years and the **Study Duration** to 4.8 years. Notice that East has changed the settings so that at 1.8 years the study should be 100% accrued. Keep the number of accrual periods equal to the default of 1. To the right of the **Accrual Info** box is the **Piecewise Constant Dropout Rates** box. This box is used to enter that rate at which we expect patients to drop out of the study. For the present we will assume that there are no drop-outs.




Click on **Compute** to complete the design. The design is shown as a row in the **Output Preview** located in the lower pane of this window. You can select this design by clicking anywhere along the row in the **Output Preview**. With Des1 selected, click the icon to display the details of this design in the upper pane, which are shown below. You may also wish to save this design. Select Des1 in the **Output Preview** window and click the to save

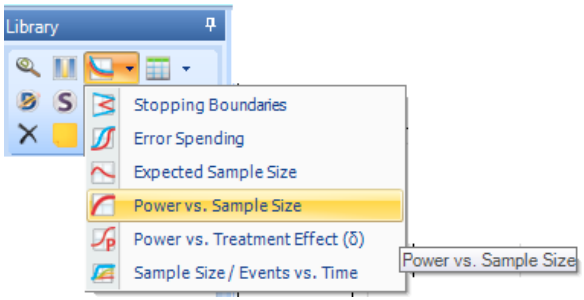
this design to Workbook1 in the **Library**.

| | Wbk1:Des1 |
|---|---|
| Mnemonic | SU-2S-LRSD |
| **Test Parameters** | |
| Design Type | Superiority |
| No. of Looks | 6 |
| Test Type | 2-Sided |
| Specified α | 0.05 |
| Power | 0.9 |
| **Model Parameters** | |
| Hazard Ratio (Alt.) | 0.83 |
| Var (Log HR) | Null |
| Allocation Ratio (nt/nc) | 1 |
| **Boundary Parameters** | |
| Spacing of Looks | Equal |
| Efficacy Boundary | LD (OF) |
| **Accrual & Dropout Parameters** | |
| Subjects are Followed | Until End of Study |
| No. of Accrual Periods | 1 |
| No. of Dropout Pieces | 0 |
| **Sample Size** | |
| Maximum | 1689 |
| Expected Under H0 | 1688.978 |
| Expected Under H1 | 1687.5564 |
| **Events** | |
| Maximum | 1243 |
| Expected Under H0 | 1233.9843 |
| Expected Under H1 | 903.5945 |
| **Accrual Duration** | |
| Maximum | 1.8 |
| Expected Under H0 | 1.8 |
| Expected Under H1 | 1.7985 |
| **Study Duration** | |
| Maximum | 4.8 |
| Expected Under H0 | 4.4164 |
| Expected Under H1 | 3.3044 |

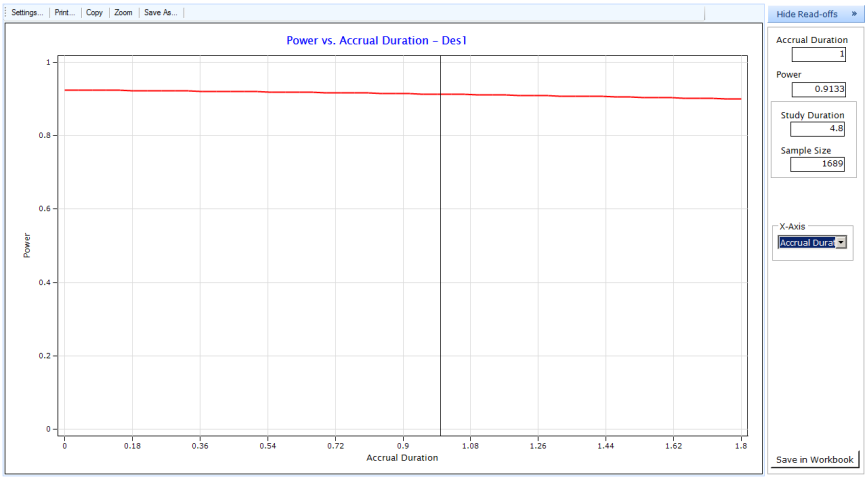
East notifies you that 1243 events and a sample size of 1689 are required to attain the desired 90% power in the allotted time.

East provides charts to examine the trade-offs between power and accrual duration, study

duration, sample size or number of events. Select Des1 in the **Library** click the icon and select **Power vs. Sample Size** as shown below.



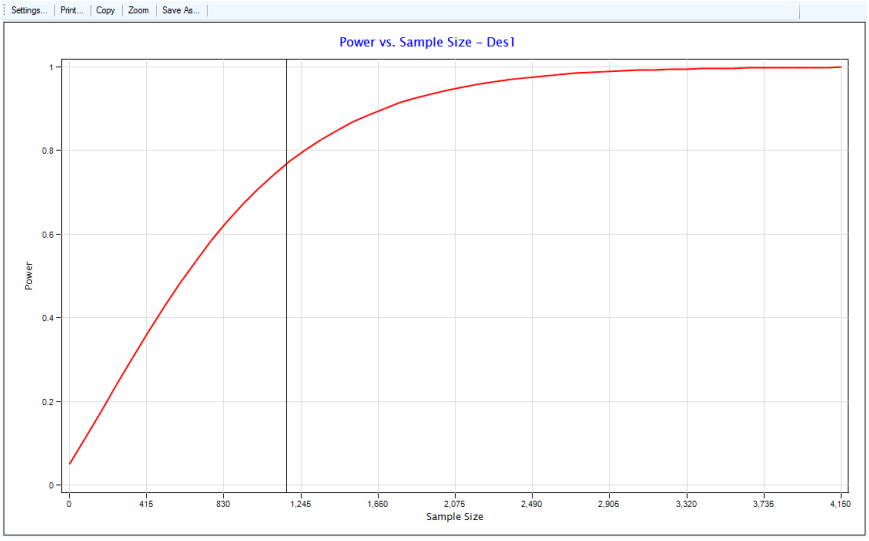

To the right of the graph, swith the **X-Axis** to **`Accrual Duration`**.




This graph shows for a fixed study duration of 4.8 years and a fixed sample size of 1689, the trade-off between power and accrual duration. For 1 year accrual, we see that the power will be 91.3%.

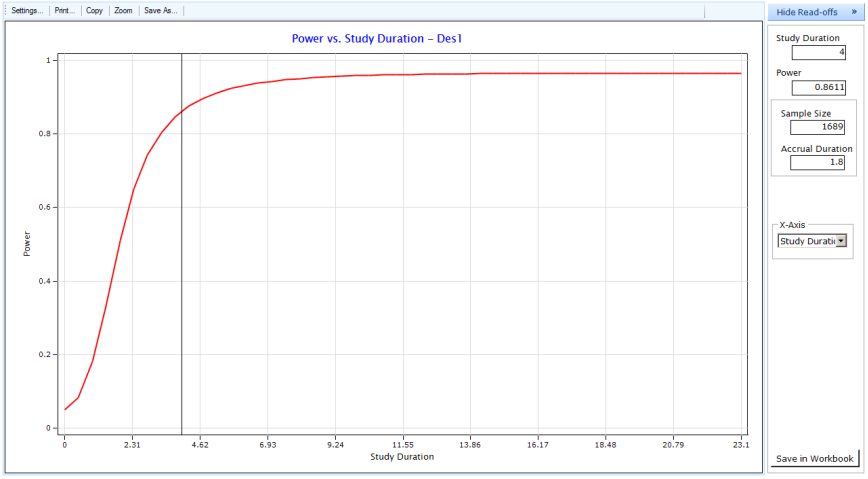
Now switch the **X-Axis** option from **`Accrual Duration`** to **`Sample Size`**. You will see the

following chart.



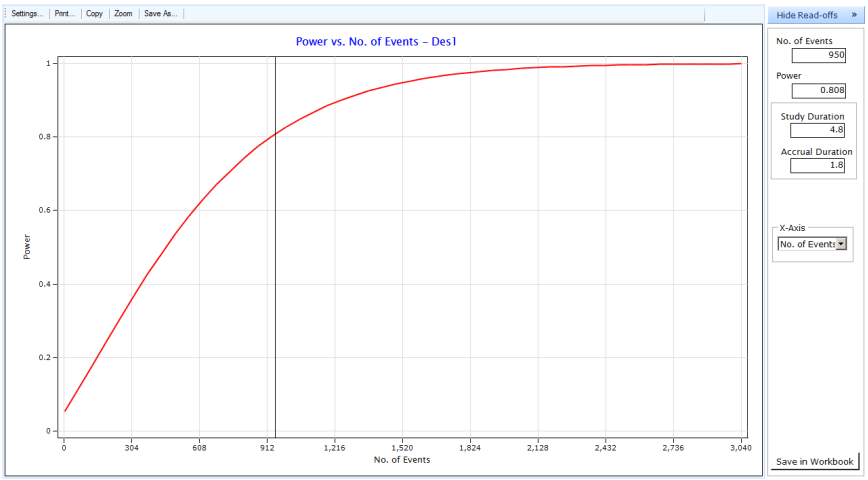

You can see that for a fixed Accrual Duration of 1.8 years and a fixed Study Duration of 4.8 years, 1170 subjects would provide you with 76.9% power. Switch the **X-Axis** option again,

this time to **Study Duration**. The following chart will appear.




Here we see that closing the trial early after 4 years, given an accrual of 1689 patients over 1.8 years, we only have 86% power to detect the alternative hypothesis of interest. Finally, switch the **X-Axis** to **No. of Events**. The power of the study is really tied to the number of events that are observed. This chart shows the direct relationship between power and number

of events.

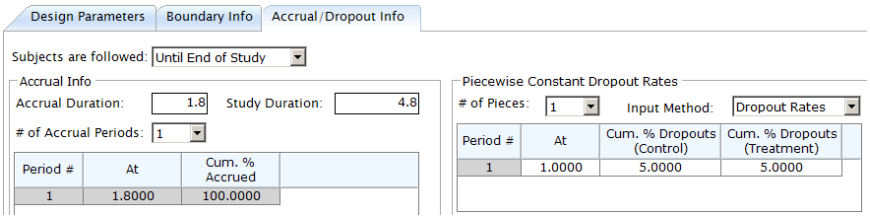



Note that 950 events give us about 81% power. You may wish to save some or all of these charts to the **Library** by clicking on the **Save in Workbook** button.

## 13.3 Incorporating Drop-Outs

The investigators expect 5% of the patients in the spironolactone group and the control group to drop out each year. Create a new design by selecting Des1 in the **Library**, and clicking the icon on the **Library** toolbar. Next, click on the **Accrual/Dropout Info** tab. In the **Piecewise Constant Dropout Rates** box, select 1 for the number of pieces and change the **Input Method** from `Hazard Rates` to `Dropout Rates`. Then enter 5% dropouts at 1 year for the treatment and control arm as shown below. Although East allows you to specify different dropout (hazard) rates for the two groups, it is recommended that you select equal

dropout (hazard) rates.




Click the **Compute** button to generate output for Des2. With Des2 selected in the **Output Preview**, click the icon to save Des2 to the **Library**. In the **Library**, select the rows for Des1 and Des2, by holding the Ctrl key, and then click the icon. The upper pane will

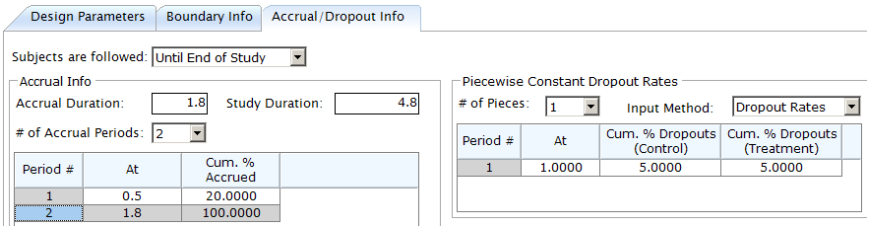
display the details of the two designs side-by-side:

| | Wbk1:Des1 | Wbk1:Des2 |
|---|---|---|
| Mnemonic | SU-2S-LRSD | SU-2S-LRSD |
| **Test Parameters** | | |
| Design Type | Superiority | Superiority |
| No. of Looks | 6 | 6 |
| Test Type | 2-Sided | 2-Sided |
| Specified α | 0.05 | 0.05 |
| Power | 0.9 | 0.9 |
| **Model Parameters** | | |
| Hazard Ratio (Alt.) | 0.83 | 0.83 |
| Var (Log HR) | Null | Null |
| Allocation Ratio (nt/nc) | 1 | 1 |
| **Boundary Parameters** | | |
| Spacing of Looks | Equal | Equal |
| Efficacy Boundary | LD (OF) | LD (OF) |
| **Accrual & Dropout Parameters** | | |
| Subjects are Followed | Until End of Study | Until End of Study |
| No. of Accrual Periods | 1 | 1 |
| No. of Dropout Pieces | 0 | 1 |
| **Sample Size** | | |
| Maximum | 1689 | 1824 |
| Expected Under H0 | 1688.978 | 1823.9663 |
| Expected Under H1 | 1687.5564 | 1820.7454 |
| **Events** | | |
| Maximum | 1243 | 1243 |
| Expected Under H0 | 1233.9843 | 1233.9843 |
| Expected Under H1 | 903.5945 | 903.5945 |
| **Accrual Duration** | | |
| Maximum | 1.8 | 1.8 |
| Expected Under H0 | 1.8 | 1.8 |
| Expected Under H1 | 1.7985 | 1.7968 |
| **Study Duration** | | |
| Maximum | 4.8 | 4.8 |
| Expected Under H0 | 4.4164 | 4.3608 |
| Expected Under H1 | 3.3044 | 3.242 |

A comparison of the two plans reveals that, because of the drop-outs, we require 1,824 subjects to be enrolled under Des2 rather than 1689 under Des1. Also, the expected study duration will not change much under the alternative and null hypotheses between Des1 and Des2.

## 13.4 Incorporating Non-Constant Accrual Rates

In many clinical trials the enrollment rate is low in the beginning and reaches its maximum expected level a few months later when all the sites enrolling patients are onboard. Suppose that 20% of the total accrual is expected to occur during the first six months with the rest happening during the remaining 1.3 years. Create a new design by selecting Des2 in the **Library**, and clicking the icon on the **Library** toolbar. Next, click on the **Accrual/Dropout Info** tab. Specify that there are two accrual periods and enter the cumulative accrual for each period in the dialog box as shown below.




Click the **Compute** button to generate output for Des3. With Des3 selected in the **Output Preview**, click the icon to save Des3 to the **Library**. In the **Library**, select the rows for Des1, Des2, and Des3 by holding the Ctrl key, and then click the icon. The upper pane

will display the details of the three designs side-by-side:

| | Wbk1:Des1 | Wbk1:Des2 | Wbk1:Des3 |
|---|---|---|---|
| Mnemonic | SU-2S-LRSD | SU-2S-LRSD | SU-2S-LRSD |
| **Test Parameters** | | | |
| Design Type | Superiority | Superiority | Superiority |
| No. of Looks | 6 | 6 | 6 |
| Test Type | 2-Sided | 2-Sided | 2-Sided |
| Specified α | 0.05 | 0.05 | 0.05 |
| Power | 0.9 | 0.9 | 0.9 |
| **Model Parameters** | | | |
| Hazard Ratio (Alt.) | 0.83 | 0.83 | 0.83 |
| Var (Log HR) | Null | Null | Null |
| Allocation Ratio (nt/nc) | 1 | 1 | 1 |
| **Boundary Parameters** | | | |
| Spacing of Looks | Equal | Equal | Equal |
| Efficacy Boundary | LD (OF) | LD (OF) | LD (OF) |
| **Accrual & Dropout Parameters** | | | |
| Subjects are Followed | Until End of Study | Until End of Study | Until End of Study |
| No. of Accrual Periods | 1 | 1 | 2 |
| No. of Dropout Pieces | 0 | 1 | 1 |
| **Sample Size** | | | |
| Maximum | 1689 | 1824 | 1837 |
| Expected Under H0 | 1688.978 | 1823.9663 | 1836.9775 |
| Expected Under H1 | 1687.5564 | 1820.7454 | 1835.7774 |
| **Events** | | | |
| Maximum | 1243 | 1243 | 1243 |
| Expected Under H0 | 1233.9843 | 1233.9843 | 1233.9843 |
| Expected Under H1 | 903.5945 | 903.5945 | 903.5945 |
| **Accrual Duration** | | | |
| Maximum | 1.8 | 1.8 | 1.8 |
| Expected Under H0 | 1.8 | 1.8 | 1.8 |
| Expected Under H1 | 1.7985 | 1.7968 | 1.7989 |
| **Study Duration** | | | |
| Maximum | 4.8 | 4.8 | 4.8 |
| Expected Under H0 | 4.4164 | 4.3608 | 4.3708 |
| Expected Under H1 | 3.3044 | 3.242 | 3.2771 |

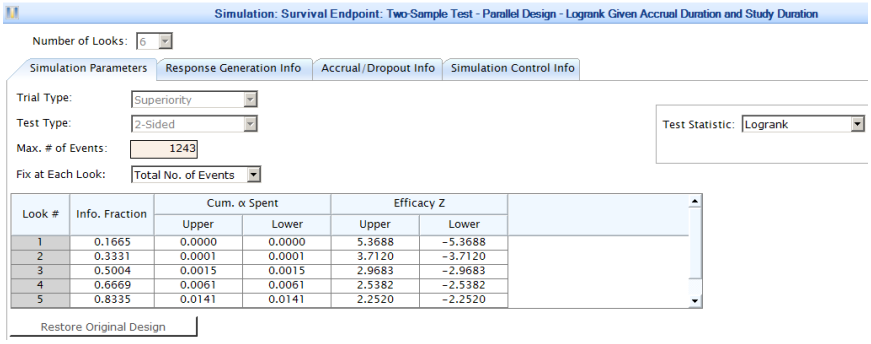
Notice that we now need 1837 subjects to be enrolled to compensate for the overall later enrollment of subjects.

## 13.5 Simulation

- 13.5.1 Simulating Under $H_1$
- 13.5.2 Simulating Under $H_0$

It would be useful to verify the operating characteristics of the various plans created in the previous section by simulation. Select Des3 in the **Library** and click the  icon. You will be

taken to the following simulation worksheet.




### 13.5.1 Simulating Under $H_1$

We will first simulate the trial under the alternative hypothesis $H_1$. In the **Simulation Parameters** tab select `Total No. of Events` to fix at each look - the default option. Select `LogRank` from the drop-down menu next to **Test Statistic**. Other options for a test statistic include the Wilcoxon-Gehan and Harrington-Flemming. Next, click the **Simulate** button to simulate 10000 trials. A new row labeled Sim1 will appear in the **Output Preview** window. Select Sim1 in the **Output Preview** and click the  icon to save it to the **Library**. In the **Library**, double-click Sim1. A portion of the output is displayed below. (The actual

values may differ, depending on the starting seed used).

**Simulation: Survival Endpoint: Two-Sample Test - Parallel Design - Logrank Given Accrual Duration and Study Duration**

| Simulation Parameters | |
|---|---|
| Simulation ID: | Sim1 |
| Design Type: | Superiority |
| Number of Looks: | 6 |
| Test Type: | 2-Sided |
| Fix at Each Look: | Total No. of Events |
| Test Statistic: | Logrank |
| Average Events: | 906.0727 |
| Total Accrual Duration: | 1.8 |
| Avg. Power at Termination: | 0.8999 |
| **Simulation Control Parameters** | |
| Starting Seed: | Clock |
| Number of Simulations: | 10000 |

**Simulation Boundaries and Boundary Crossing Probabilities:**

| Look # | Events | Boundaries: Efficacy: Upper | Boundaries: Efficacy: Lower | Early Stopping For: Upper Efficacy | Early Stopping For: Lower Efficacy | Total Simulations: Count | Total Simulations: % |
|---|---|---|---|---|---|---|---|
| 1 | 207 | 5.3688 | -5.3688 | 0 | 0 | 0 | 0 |
| 2 | 414 | 3.712 | -3.712 | 0 | 348 | 348 | 3.48 |
| 3 | 622 | 2.9683 | -2.9683 | 0 | 2192 | 2192 | 21.92 |
| 4 | 829 | 2.5382 | -2.5382 | 0 | 3058 | 3058 | 30.58 |
| 5 | 1036 | 2.252 | -2.252 | 0 | 2191 | 2191 | 21.91 |
| 6 | 1243 | 2.0448 | -2.0448 | 0 | 1210 | 2211 | 22.11 |
| Total | | | | 0 | 8999 | 10000 | |
| % | | | | 0 | 89.99 | | |

**Average Sample Size, Dropouts and Look Times:**

| Look # | Average Sample Size | Average Events: Control | Average Events: Treatment | Average Dropouts: Control | Average Dropouts: Treatment | Average Look Time | Average Follow up |
|---|---|---|---|---|---|---|---|
| 1 | 1220.3248 | 111.9543 | 95.0458 | 15.1613 | 15.4132 | 1.2544 | 0.4874 |
| 2 | 1796.2525 | 223.0408 | 190.9592 | 30.1314 | 30.9663 | 1.7683 | 0.663 |
| 3 | 1837 | 332.9846 | 289.0154 | 45.0338 | 46.7186 | 2.2524 | 0.9745 |
| 4 | 1837 | 438.4521 | 390.5479 | 59.9051 | 62.4934 | 2.8495 | 1.2994 |
| 5 | 1837 | 540.7971 | 495.2029 | 74.7926 | 78.366 | 3.6357 | 1.6248 |
| 6 | 1837 | 641.3609 | 601.6391 | 89.5952 | 94.1655 | 4.7846 | 1.9496 |
| Average | 1835.6939 | 482.2528 | 423.8199 | 65.1517 | 68.7752 | 3.2823 | 1.4219 |

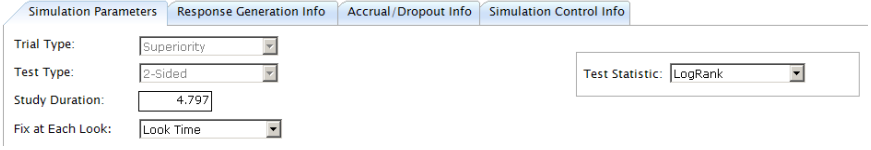
**Response Generation Parameters**

No. of Hazard Pieces: 1
Input Method: Hazard Rates

| Piece # | Starting At | Control | Treatment | Hazard Ratio |
|---|---|---|---|---|
| 1 | 0 | 0.38 | 0.3154 | 0.83 |

The column labeled **Average Look Time** displays the average calendar times at which each interim look was taken. Thus, the first interim look (taken after observing 207 events) occurred after an average elapse of 1.254 years; the second interim look (taken after observing 414 events) occurred after an average elapse of 1.768 years; etc. We see that 8999 of the 10000 simulations crossed the lower stopping boundary, thus confirming (up to Monte Carlo accuracy) that this design has 90% power.

We will now run another 10000 simulations, this time fixing the calendar time of each look instead of fixing the number of events. Select Des3 in the **Library** and click the S icon. In the **Simulation Parameters** tab select **`Look Time`** from the drop-down menu next to **Fix at Each Look:**.




When the **`Look Time`** option is selected the locations of the interim looks at which stopping

boundaries are computed are expressed in terms of the calendar time of each interim look instead of the number of events at each interim look.

Next, click the **Simulate** button to simulate 10000 trials. A new row labeled Sim2 will appear in the **Output Preview** window. Select Sim2 in the **Output Preview** and click the  icon to save it to the **Library**. In the **Library**, double-click Sim2. A portion of the output is displayed below. (The actual values may differ, depending on the starting seed used).

**Simulation: Survival Endpoint: Two-Sample Test - Parallel Design - Logrank Given Accrual Duration and Study Duration**

| Simulation Parameters | |
|---|---|
| Simulation ID: | Sim2 |
| Design Type: | Superiority |
| Number of Looks: | 6 |
| Test Type: | 2-Sided |
| Fix at Each Look: | Look Time |
| Test Statistic: | Logrank |
| Average Events: | 902.3466 |
| Total Accrual Duration: | 1.8 |
| Avg. Power at Termination: | 0.905 |
| **Simulation Control Parameters** | |
| Starting Seed: | Clock |
| Number of Simulations: | 10000 |

**Simulation Boundaries and Boundary Crossing Probabilities:**

| Look # | Look Time | Boundaries Efficacy Upper | Boundaries Efficacy Lower | Early Stopping For Upper Efficacy | Early Stopping For Lower Efficacy | Total Simulations Count | Total Simulations % |
|---|---|---|---|---|---|---|---|
| 1 | 1.2556 | 5.3688 | -5.3688 | 0 | 0 | 0 | 0 |
| 2 | 1.7692 | 3.712 | -3.712 | 0 | 355 | 355 | 3.55 |
| 3 | 2.254 | 2.9683 | -2.9683 | 0 | 2251 | 2251 | 22.51 |
| 4 | 2.8522 | 2.5382 | -2.5382 | 0 | 3046 | 3046 | 30.46 |
| 5 | 3.6404 | 2.252 | -2.252 | 0 | 2195 | 2195 | 21.95 |
| 6 | 4.7968 | 2.0448 | -2.0448 | 0 | 1203 | 2153 | 21.53 |
| Total | | | | 0 | 9050 | 10000 | |
| % | | | | 0 | 90.5 | | |

**Average Sample Size, Dropouts and Look Times:**

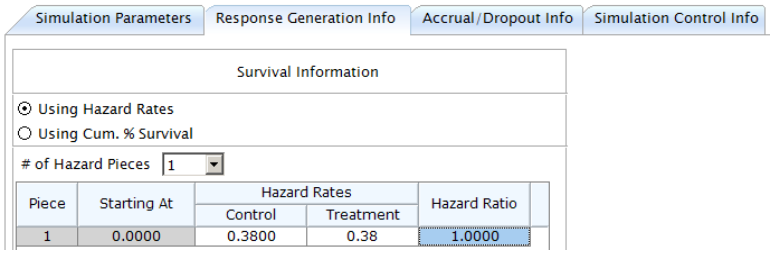
| Look # | Average Sample Size | Average Events Control | Average Events Treatment | Average Dropouts Control | Average Dropouts Treatment | Average Look Time | Average Follow up |
|---|---|---|---|---|---|---|---|
| 1 | 1221.3261 | 111.9336 | 95.0605 | 15.1319 | 15.3144 | 1.2556 | 0.4878 |
| 2 | 1801.922 | 223.0065 | 190.8106 | 30.1229 | 30.8533 | 1.7692 | 0.6616 |
| 3 | 1837 | 333.0185 | 288.8085 | 45.0139 | 46.6648 | 2.254 | 0.9756 |
| 4 | 1837 | 438.3702 | 390.5917 | 59.9874 | 62.4464 | 2.8522 | 1.3011 |
| 5 | 1837 | 540.877 | 495.3077 | 75.0476 | 78.2597 | 3.6404 | 1.6274 |
| 6 | 1837 | 641.784 | 601.6935 | 89.8941 | 94.4988 | 4.7968 | 1.9538 |
| Average | 1835.7669 | 480.4424 | 421.9042 | 64.8572 | 68.5277 | 3.2708 | 1.4174 |

We see that the first interim look is taken, on average, after 1.256 years during which time an average of 206.99 events are observed. In our simulations we have achieved 90.5% power, thus confirming (up to Monte Carlo accuracy) that the study has 90% power.

### *13.5.2 Simulating Under $H_0$*

To simulate under the null hypothesis we must go to the **Response Generation Info** tab in

the simulation worksheet. In this tab change the hazard rate for the treatment arm to 0.38.




This change implies that we will be simulating under the null hypothesis. Next, click on the **Simulation Parameters** tab and make sure that the `Total No. of Events` is fixed at each look. Next, click the **Simulate** button to simulate 10000 trials. A portion of the results are displayed below.

**Simulation Boundaries and Boundary Crossing Probabilities:**

| Look # | Events | Boundaries Efficacy | | Early Stopping For | | Total Simulations | |
|---|---|---|---|---|---|---|---|
| | | Upper | Lower | Upper Efficacy | Lower Efficacy | Count | % |
| 1 | 207 | 5.3688 | -5.3688 | 0 | 0 | 0 | 0 |
| 2 | 414 | 3.712 | -3.712 | 0 | 2 | 2 | 0.02 |
| 3 | 622 | 2.9683 | -2.9683 | 11 | 17 | 28 | 0.28 |
| 4 | 829 | 2.5382 | -2.5382 | 52 | 40 | 92 | 0.92 |
| 5 | 1036 | 2.252 | -2.252 | 83 | 75 | 158 | 1.58 |
| 6 | 1243 | 2.0448 | -2.0448 | 110 | 110 | 9720 | 97.2 |
| Total | | | | 256 | 244 | 10000 | |
| % | | | | 2.56 | 2.44 | | |

**Average Sample Size, Dropouts and Look Times:**

| Look # | Average Sample Size | Average Events | | Average Dropouts | | Average Look Time | Average Follow up |
|---|---|---|---|---|---|---|---|
| | | Control | Treatment | Control | Treatment | | |
| 1 | 1168.3164 | 103.604 | 103.396 | 13.9964 | 13.976 | 1.2082 | 0.4656 |
| 2 | 1726.9583 | 207.0993 | 206.9007 | 28.0179 | 27.9528 | 1.7026 | 0.6302 |
| 3 | 1837 | 311.1532 | 310.8468 | 42.0658 | 41.9334 | 2.1465 | 0.8907 |
| 4 | 1837 | 414.7712 | 414.2288 | 56.0412 | 55.8752 | 2.6867 | 1.1875 |
| 5 | 1837 | 518.3153 | 517.6847 | 69.9973 | 69.84 | 3.391 | 1.4839 |
| 6 | 1837 | 621.8174 | 621.1826 | 83.998 | 83.7572 | 4.4083 | 1.7803 |
| Average | 1836.9872 | 617.2857 | 616.7303 | 83.4076 | 83.1433 | 4.3693 | 1.7673 |

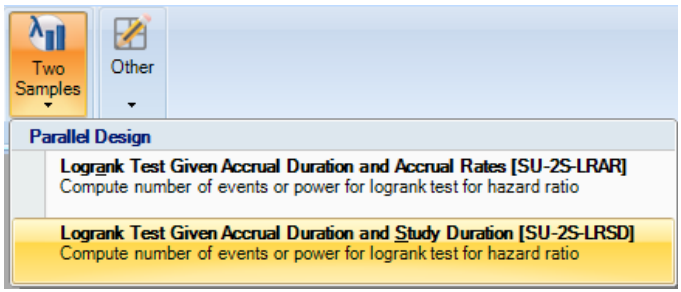
Out of 10000 simulated trials 244 crossed the lower stopping boundary and 256 crossed the upper stopping boundary thus confirming (up to Monte Carlo accuracy) that the type-1 error

is preserved for this design.

## 13.6 User Defined R Function

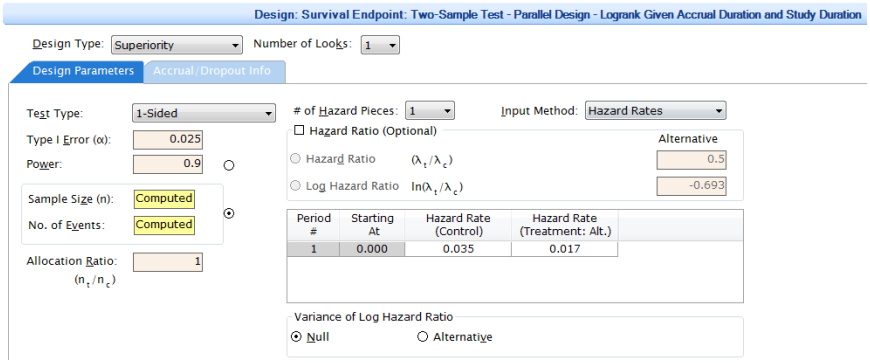
East allows you to customize simulations by inserting user-defined R functions for one or more of the following tasks: generate response, compute test statistic, randomize subjects, generate arrival times, and generate dropout information. The R functionality for arrivals and dropouts will be available only if you have entered such information at the design stage. Although the R functions are also available for all normal and binomial endpoints, we will illustrate this functionality for a time-to-event endpoint. Specifically, we will use an R function to generate Weibull survival responses.

Start East afresh. On the **Design** tab, click **Survival: Two Samples** and then **Logrank Test Given Accrual Duration and Study Duration**.

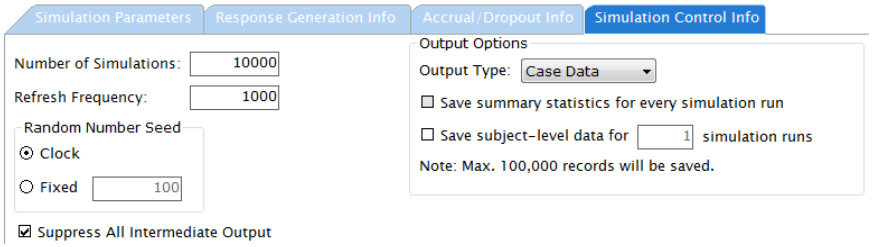



Choose the design parameters as shown below. In particular, select a one sided test with

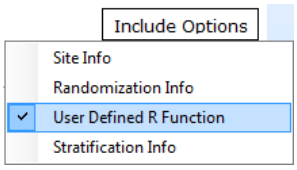
type-1 error of $\alpha = 0.025$.




Click **Compute** and save this design (Des1) to the **Library**. Right-click Des1 in the **Library** and click **Simulate**. In the **Simulation Control Info** tab, check the box for **`Suppress All Intermediate Input`**. Type 10000 for **Number of Simulations** and select **`Clock`** for **Random Number Seed**.

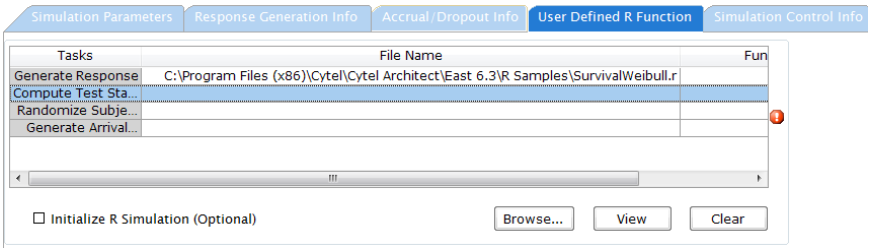



In the top right-hand corner for the input window, click **Include Options**, and then click **User Defined R Function**.

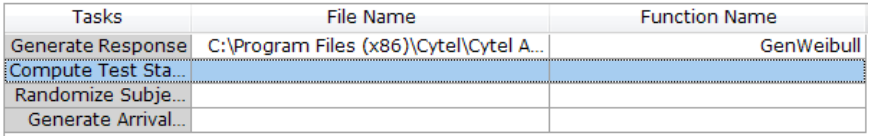


For now, leave the box **Initialize R simulation (optional)** unchecked. This optional task can be useful for loading required libraries, setting seeds for simulations, and initializing global variables.

Select the row for **Generate Response**, click **Browse...**, and navigate to the folder containing your R file. Select the file and click **Open**. The path should now be displayed under **File Name**.

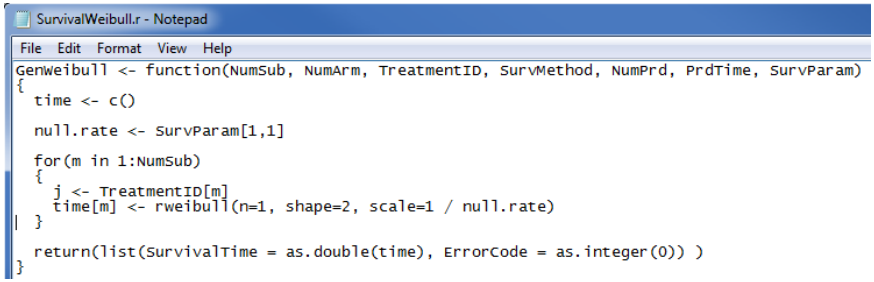



Click **View** to open a notepad application to view your R file. In this example, I am generating survival responses for both control and treatment arms from a Weibull with shape parameter $= 1$ (i.e. exponential), with the same hazard rate in both arms.

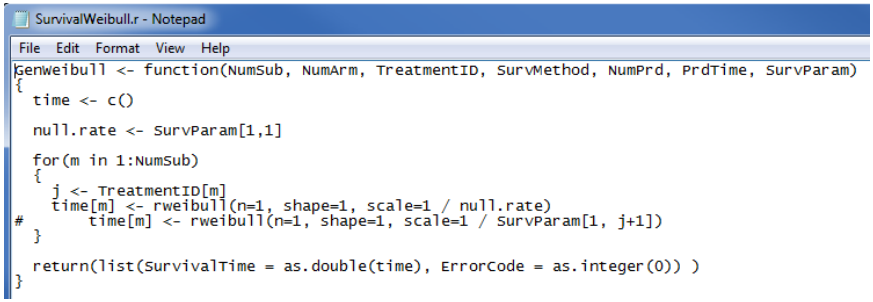



Copy the function name (in this case *GenWeibull*) and paste it into the cell for **Function Name**.

Save and close the R file, and click **Simulate**.




Return to the tab for **User Defined R Function**, select the **Generate Response** row, and click **View**. In the R function, change the shape parameter $= 2$, to generate responses from a Weibull distribution with increasing hazards. Save and close the R file, and click **Simulate**.




Select both simulations (Sim1 and Sim2) from the **Output Preview**, and on the toolbar, click

 to display in the **Output Summary**.

| | Sim1 | Sim2 |
|---|---|---|
| Mnemonic | SU-2S-LRSD | SU-2S-LRSD |
| **Test Parameters** | | |
| Design Type | Superiority | Superiority |
| Test Type | 1-Sided | 1-Sided |
| Test Statistic | Logrank | Logrank |
| Power | 0.026 | 0.027 |
| No. of Looks | 1 | 1 |
| **Model Parameters** | | |
| No. of Hazard Pieces | 1 | 1 |
| **Accrual & Dropout Parameters** | | |
| Followup Duration | Until End of Study | Until End of Study |
| No. of Accrual Periods | 1 | 1 |
| **Sample Size** | | |
| Maximum | 182 | 182 |
| **Events** | | |
| Maximum | 88 | 88 |
| **Simulation Results (Overall)** | | |
| Average Study Duration | 34.637 | 30.681 |
| Average Sample Size | 182 | 182 |
| Average Events | 88 | 88 |

Notice that the type-1 error appears to be controlled in both cases. When we simulated from the exponential (Sim1), the average study duration (30.7 months) was close to what was calculated at Des1 for the expected study duration under the null. However, when we simulated from the Weibull with decreasing hazards (Sim2), the average study duration increased to 34.6 months.

Appendix **??** contains detailed specifications for the required inputs and outputs of R functions for each task and endpoint. The ability to use custom R functions for many simulation tasks allows considerable flexibility in performing sensitivity analyses and assessment of key operating characteristics.

# 14 Non Inferiority Trials Given Accrual Duration and Study Duration

This chapter will illustrate through a worked example how to design and simulate a two-sample non inferiority trial with a time to event trial endpoint, when the accrual duration and study duration are fixed.

## 14.1 Calculating a Sample Size

For this design, East obtains the maximum number of events $D_{max}$ from the maximum information $I_{max}$, as described in Appendix sections **??** and **??**. To calculate the sample size, we first equate the expected number of events $d(S_a + S_f)$ (as calculated in Appendix **??** which depends on the accrual duration ($S_a$) and the duration of follow-up ($S_f$) to the maximum number of events $D_{max}$.

$$d(S_a + S_f) = D_{max} \tag{14.1}$$

In this type of design the accrual duration $S_a$ and the study duration $S_a + S_f$ are given as input. East iterates between sample sizes, increasing onwards from a minimum value of $D_{max}$, enrolled over a duration of $S_a$ until $D_{max}$ events are found to occur within a study duration of $S_a + S_f$. The result is the unique sample size required to obtain the proper power for the study.

## 14.2 The Non Inferiority Margin

The first step in designing a non-inferiority trial is to establish a suitable non inferiority margin. This is typically done by performing a meta-analysis on past clinical trials of the active control versus placebo. Regulatory agencies then require the sponsor of the clinical trial to

demonstrate that a fixed percentage of the active control effect (usually 50%) is retained by the new treatment. A further complication arises because the active control effect can only be estimated with error. We illustrate below with an example provided by reviewers at the FDA.

Rothman et al. (2003) have discussed a clinical trial to establish the non inferiority of the test drug Xeloda (treatment $t$) relative to the active control (treatment $c$) consisting of 5 fluorouracil with leucovarin (5FU+LV) for metastatic colorectal cancer. In order to establish a suitable non inferiority margin for this trial it is necessary to first establish the effect of 5FU+LV relative to the reference therapy of 5FU alone (treatment $p$, here regarded as placebo). To establish this effect the FDA conducted a ten study random effects meta analysis (FDA Medical Statistical review for Xeloda, NDA 20 896, April 2001) of randomized comparisons of 5-FU alone versus 5-FU+LV. Letting $\lambda_t, \lambda_c$ and $\lambda_p$ denote the constant hazard rates for the new treatment, the active control and the placebo, respectively, the FDA meta analysis established that

$$\ln\,(\widehat{\lambda_p/\lambda_c}) = 0.234$$

with standard error

$$\mathsf{se}[\ln\,(\widehat{\lambda_p/\lambda_c})] = 0.075\ .$$

Thus with $100\gamma$% confidence the active control effect lies inside the interval

$$[0.234 - 0.075\Phi^{-1}(\frac{1+\gamma}{2}), 0.234 + 0.075\Phi^{1}(\frac{1+\gamma}{2})] \tag{14.2}$$

The new study is required to demonstrate that some fraction (usually 50%) of the active control effect is retained. Rothman et al. (2003) state that the claim of non inferiority for the new treatment relative to the active control can be demonstrated if the upper limit of a two sided $100(1-\alpha)$% confidence interval for $\ln(\lambda_t/\lambda_c)$ is less than a pre specified fraction of the lower limit of a two sided $100\gamma$% confidence interval for the active control effect established by the meta-analysis. This is known as the "two confidence intervals procedure". Specifically in order to claim non inferiority in the current trial it is necessary to show that

$$\ln\,(\widehat{\lambda_t/\lambda_c}) + \Phi^{-1}(1-\alpha/2)\mathsf{se}[\ln\,(\widehat{\lambda_t/\lambda_c})] < (1-f_0)\{\ln\,(\widehat{\lambda_p/\lambda_c}) - \Phi^{-1}(\frac{1+\gamma}{2})\mathsf{se}[\ln\,(\widehat{\lambda_p/\lambda_c})]\}\ . \tag{14.3}$$

We may re-write the non inferiority condition (14.3) in terms of a one-sided Wald test of the form

$$\frac{\ln(\widehat{\lambda_t/\lambda_c}) - \delta_0}{\text{se}[\ln(\widehat{\lambda_t/\lambda_c})]} < \Phi^{-1}(1-\alpha/2)\ , \tag{14.4}$$

where

$$\delta_0 = (1-f_0)\{\ln(\widehat{\lambda_p/\lambda_c}) - \Phi^{-1}(\frac{1+\gamma}{2})\text{se}[\ln(\widehat{\lambda_p/\lambda_c})]\} \tag{14.5}$$

is the non inferiority margin.

The choice $f_0 = 1$ implies that the entire active control effect must be retained in the new trial and amounts to running a superiority trial. At the other end of the spectrum, the choice $f_0 = 0$ implies that none of the active control effect need be retained; i.e., the new treatment is only required to demonstrate effectiveness relative to placebo. The usual choice is $f_0 = 0.5$, implying that the new treatment is required to retain at least 50% of the active control effect. The usual choice for $\alpha$ is $\alpha = 0.05$. A conservative choice for the coefficient $\gamma$ is $\gamma = (1-\alpha) = 0.95$. Rothman et al. (2003) refer to this method of establishing the non inferiority margin as the "two 95 percent two sided confidence interval procedure" or the "95-95 rule". In general this approach leads to rather tight margins unless the active control effect is substantial. Rothman et al. (2003) have also proposed more lenient margins that vary with the amount of power desired. Fleming (2007), however, argues for the stricter 95-95 rule on the grounds that it offers greater protection against an ineffective medical compound being approved in the event that the results of the previous trials used to establish the active control effect are of questionable relevance to the current setting. Accordingly we evaluate (14.5) with $\gamma = 0.95$, $f_0 = 0.5$, $\ln(\widehat{\lambda_p/\lambda_c}) = 0.234$ and $\text{se}[\ln(\widehat{\lambda_p/\lambda_c})] = 0.075$ thereby obtaining the non inferiority margin to be $\delta_0 = 0.044$ for the log hazard ratio and $\exp(0.044) = 1.045$ for the hazard ratio.

## 14.3 Design of Metastatic Colorectal Cancer Trial

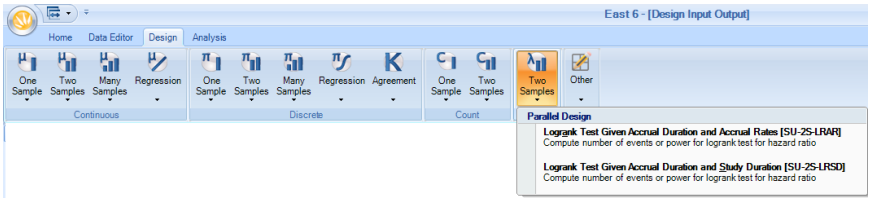
In this section we will use East to design a single-look non inferiority trial comparing the test drug Xeloda (treament $t$) to the active control 5FU+LV (treatment $c$) for the treatment of metastatic colorectal cancer. On the basis of a meta analysis of ten previous studies of the active control versus placebo (Rothman et. al. 2003), a non inferiority margin of 1.045 for

$\lambda_t/\lambda_c$ has been established. Thus we are interested in testing the null hypothesis of inferiority $H_0$: $\lambda_t/\lambda_c \geq 1.045$ versus the one-sided alternative hypothesis that $\lambda_t/\lambda_c < 1.045$. Suppose the trial is planned to enroll for 30 months and finish within 70 months of the last patient enrolled.

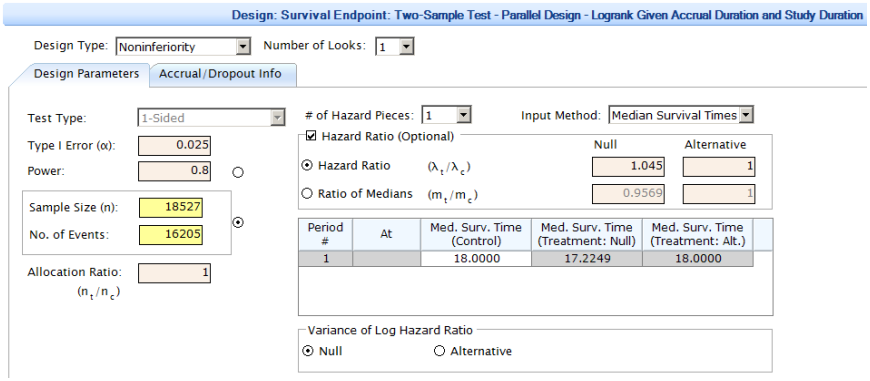
### 14.3.1 Single-Look Design

We will use East to create an initial single-look design having 80% power to detect the alternative hypothesis $H_1$: $\lambda_t/\lambda_c = 1$ with a one sided level-0.025 non-inferiority test.

To begin click **Survival: Two Samples** on the **Design** tab and then click **Parallel Design: Logrank Test Given Accrual Duration and Study Duration** as shown below.



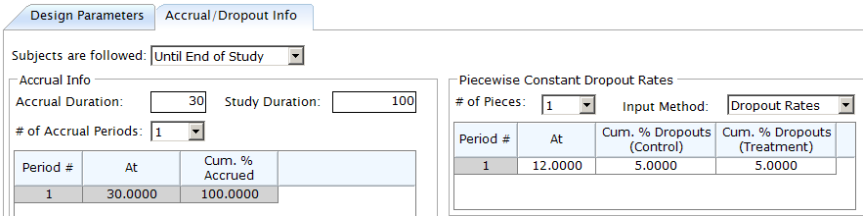

A new screen will appear. Enter the appropriate design parameters into the dialog box as shown below.




The box labeled **Variance of Log Hazard Ratio** specifies whether the calculation of the

required number of events is to be based on the variance estimate of the log hazard ratio under the null hypothesis or the alternative hypothesis. The default choice in East is **Null**. Most textbooks recommend this choice as well (see, for example Collett, 1994, equation (2.21) specialized to no ties). It will usually not be necessary to change this default. For a technical discussion of this issue refer to Appendix **??**, Section **??**.

Next click on the **Accrual/Dropout Info** tab. Here we will specify the accrual information and dropout rates. Set the accrual duration to 30 months and the study duration to 100 months in the **Accrual Info** box. Also, suppose that there are 5% drop-outs per year in each arm. Enter these values as shown below.




Click on **Compute** to complete the design. The design is shown as a row in the **Output Preview** located in the lower pane of this window. You can select this design by clicking anywhere along the row in the **Output Preview**. With Des1 selected, click the  icon to display the details of this design in the upper pane, which are shown below. You may also wish to save this design. Select Des1 in the **Output Preview** window and click the  to save

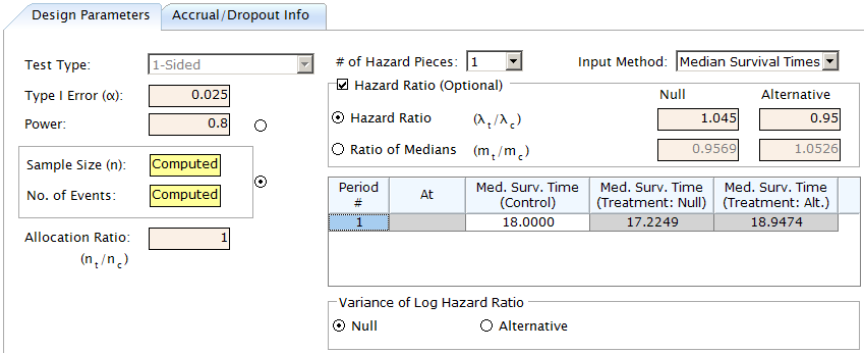
this design to Workbook1 in the **Library**.

| | Wbk1:Des1 |
|---|---|
| Mnemonic | SU-2S-LRSD |
| **Test Parameters** | |
| Design Type | Noninferiority |
| No. of Looks | 1 |
| Test Type | 1-Sided |
| Specified α | 0.025 |
| Power | 0.8 |
| **Model Parameters** | |
| Hazard Ratio (Null) | 1.045 |
| Hazard Ratio (Alt.) | 1 |
| Var (Log HR) | Null |
| Allocation Ratio (nt/nc) | 1 |
| **Accrual & Dropout Parameters** | |
| Subjects are Followed | Until End of Study |
| No. of Accrual Periods | 1 |
| No. of Dropout Pieces | 1 |
| **Sample Size** | |
| Maximum | 18527 |
| Expected Under H0 | 18527 |
| Expected Under H1 | 18527 |
| **Events** | |
| Maximum | 16205 |
| Expected Under H0 | 16205 |
| Expected Under H1 | 16205 |
| **Accrual Duration** | |
| Maximum | 30 |
| Expected Under H0 | 30 |
| Expected Under H1 | 30 |
| **Study Duration** | |
| Maximum | 100 |
| Expected Under H0 | 96.7434 |
| Expected Under H1 | 99.9562 |

It is immediately evident that Des1 is untenable. It requires 16,205 events to be fully powered and 18,527 subjects to obtain those events within the course of the study. The problem lies with trying to power the trial to detect a hazard ratio of 1 under the alternative hypothesis. Suppose instead that the investigators actually believe that the treatment is slightly superior to the active control, but the difference is too small to be detected in a superiority trial. In that

case a non-inferiority design powered at a hazard ratio less than 1 (0.95, say) would be a better option because such a trial would require fewer events.

To see this create a new design by selecting Des1 in the **Library**, and clicking the  icon on the **Library** toolbar. Then edit this design by specifying a hazard ratio of 0.95 under the alternative hypothesis as shown below.




Click the **Compute** button to generate output for Des2. With Des2 selected in the **Output Preview**, click the  icon to save Des2 to the **Library**. In the **Library**, select the rows for Des1 and Des2, by holding the Ctrl key, and then click the  icon. The upper pane will

display the details of the two designs side-by-side:

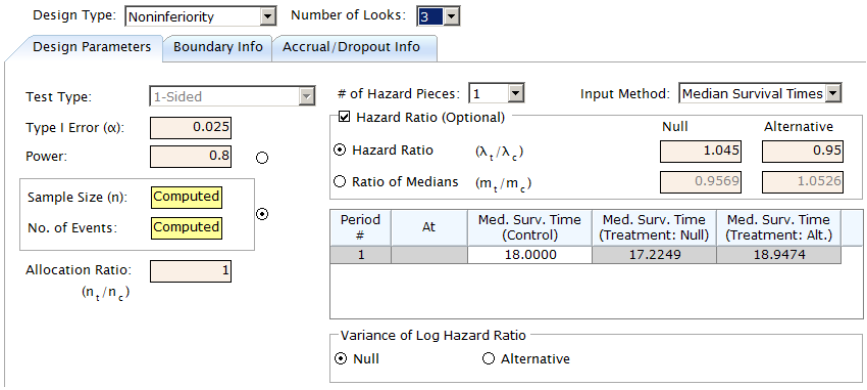
| | Wbk1:Des1 | Wbk1:Des2 |
|---|---|---|
| Mnemonic | SU-2S-LRSD | SU-2S-LRSD |
| **Test Parameters** | | |
| Design Type | Noninferiority | Noninferiority |
| No. of Looks | 1 | 1 |
| Test Type | 1-Sided | 1-Sided |
| Specified α | 0.025 | 0.025 |
| Power | 0.8 | 0.8001 |
| **Model Parameters** | | |
| Hazard Ratio (Null) | 1.045 | 1.045 |
| Hazard Ratio (Alt.) | 1 | 0.95 |
| Var (Log HR) | Null | Null |
| Allocation Ratio (nt/nc) | 1 | 1 |
| **Accrual & Dropout Parameters** | | |
| Subjects are Followed | Until End of Study | Until End of Study |
| No. of Accrual Periods | 1 | 1 |
| No. of Dropout Pieces | 1 | 1 |
| **Sample Size** | | |
| Maximum | 18527 | 3973 |
| Expected Under H0 | 18527 | 3973 |
| Expected Under H1 | 18527 | 3973 |
| **Events** | | |
| Maximum | 16205 | 3457 |
| Expected Under H0 | 16205 | 3457 |
| Expected Under H1 | 16205 | 3457 |
| **Accrual Duration** | | |
| Maximum | 30 | 30 |
| Expected Under H0 | 30 | 30 |
| Expected Under H1 | 30 | 30 |
| **Study Duration** | | |
| Maximum | 100 | 100 |
| Expected Under H0 | 96.7434 | 93.2179 |
| Expected Under H1 | 99.9562 | 99.8704 |

Des2 is clearly easier to implement than Des1. It requires only 3,457 events to be fully powered. This can be achieved with only 3,973 patients enrolled in the study.

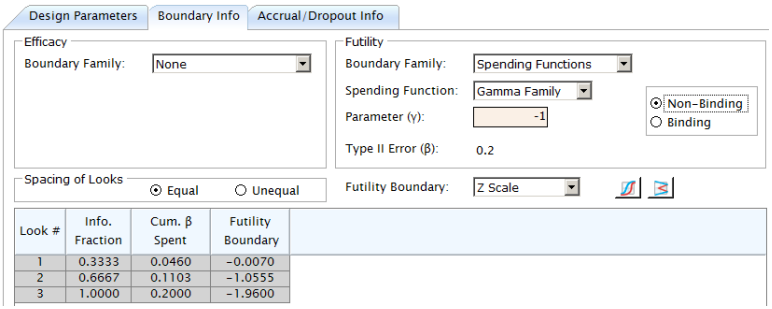
### *14.3.2 Early Stopping for Futility*

Under the null hypothesis, Des2, with 3,457 events, has an expected study duration of 93.2 months. This is a very long time commitment for a trial that is unlikely to be successful. Therefore it would be a good idea to introduce a futility boundary for possible early stopping. Since we wish to be fairly aggressive about early stopping for futility we will generate the futility boundary from the Gamma$(-1)$ $\beta$ spending function. On the other hand since there no interest in early stopping for efficacy we will not use an efficacy boundary.

Create a new design by selecting Des2 in the **Library**, and clicking the icon on the **Library** toolbar. Change the number of looks from 1 to 3 as shown below.




Next, click on the **Boundary Info** tab. Enter the parameters as shown below. Be sure to select the **Non Binding** option. This choice gives us the flexibility to continue the trial even if a futility boundary has been crossed. Data monitoring committees usually want this flexibility; for example, to follow a secondary endpoint.



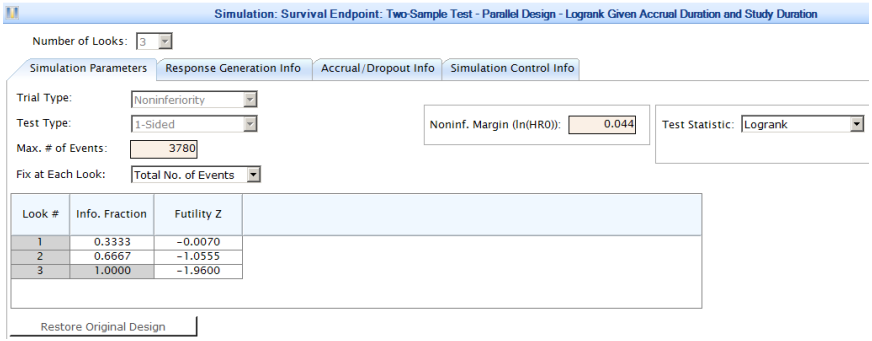

Click the **Compute** button to generate output for Des3. With Des3 selected in the **Output Preview**, click the icon to save Des3 to the **Library**. In the **Library**, select the rows for Des1, Des2, and Des3 by holding the Ctrl key, and then click the icon. The upper pane

will display the details of the three designs side-by-side:

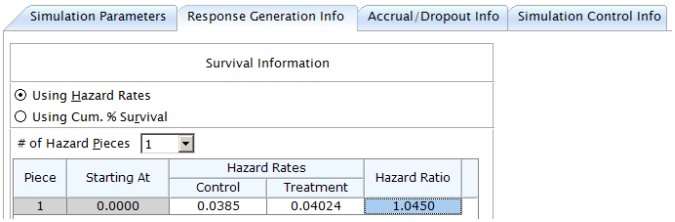
| | Wbk1:Des1 | Wbk1:Des2 | Wbk1:Des3 |
|---|---|---|---|
| Mnemonic | SU-2S-LRSD | SU-2S-LRSD | SU-2S-LRSD |
| **Test Parameters** | | | |
| Design Type | Noninferiority | Noninferiority | Noninferiority |
| No. of Looks | 1 | 1 | 3 |
| Test Type | 1-Sided | 1-Sided | 1-Sided |
| Specified α | 0.025 | 0.025 | 0.025 |
| Attained α | | | 0.0219 |
| Power | 0.8 | 0.8001 | 0.8 |
| **Model Parameters** | | | |
| Hazard Ratio (Null) | 1.045 | 1.045 | 1.045 |
| Hazard Ratio (Alt.) | 1 | 0.95 | 0.95 |
| Var (Log HR) | Null | Null | Null |
| Allocation Ratio (nt/nc) | 1 | 1 | 1 |
| **Boundary Parameters** | | | |
| Futility Boundary | | | Gm (-1) (NB) |
| Spacing of Looks | | | Equal |
| **Accrual & Dropout Parameters** | | | |
| Subjects are Followed | Until End of Study | Until End of Study | Until End of Study |
| No. of Accrual Periods | 1 | 1 | 1 |
| No. of Dropout Pieces | 1 | 1 | 1 |
| **Sample Size** | | | |
| Maximum | 18527 | 3973 | 4344 |
| Expected Under H0 | 18527 | 3973 | 3965.0318 |
| Expected Under H1 | 18527 | 3973 | 4312.7208 |
| **Events** | | | |
| Maximum | 16205 | 3457 | 3780 |
| Expected Under H0 | 16205 | 3457 | 2056.3268 |
| Expected Under H1 | 16205 | 3457 | 3583.036 |
| **Accrual Duration** | | | |
| Maximum | 30 | 30 | 30 |
| Expected Under H0 | 30 | 30 | 27.3828 |
| Expected Under H1 | 30 | 30 | 29.784 |
| **Study Duration** | | | |
| Maximum | 100 | 100 | 100 |
| Expected Under H0 | 96.7434 | 93.2179 | 39.6138 |
| Expected Under H1 | 99.9562 | 99.8704 | 92.7138 |

Observe that while the sample size has been inflated to 4,344 subjects compared to Des2, the expected study duration under $H_0$ has been cut down to 39.6 months and the expected sample size under $H_0$ is 3,965. It would also be useful to simulate Des3 under a variety of scenarios for the hazard ratio. Select Des3 in the **Library** and click the  icon. You will be

taken to the following simulation worksheet.




We wish to simulate this trial under the null hypothesis that the hazard ratio is $\exp(0.044) = 1.045$. To do this go to the **Response Generation Info** tab in the simulation worksheet. In this tab change the control and treatment hazard rates as shown below.




Next, click the **Simulate** button to simulate 10000 trials. A new row labeled Sim1 will appear in the **Output Preview** window. Select Sim1 in the **Output Preview** and click the  icon to save it to the **Library**. In the **Library**, double-click Sim1. A portion of the output is displayed

below. (The actual values may differ, depending on the starting seed used).

**Simulation Boundaries and Boundary Crossing Probabilities:**

| Look # | Events | Boundaries | Early Stopping For | Unable To Reject H1 | Total Simulations | |
|---|---|---|---|---|---|---|
| | | Futility | | | | |
| | | Upper | Futility | | Count | % |
| 1 | 1260 | -0.007 | 5137 | | 5137 | 51.37 |
| 2 | 2520 | -1.0555 | 3542 | | 3542 | 35.42 |
| 3 | 3780 | -1.96 | 1095 | 226 | 1321 | 13.21 |
| Total | | | 9774 | 226 | 10000 | |
| % | | | 97.74 | 2.26 | | |

**Average Sample Size, Dropouts and Look Times:**

| Look # | Average Sample Size | Average Events | | Average Dropouts | | Average Look Time | Average Follow up |
|---|---|---|---|---|---|---|---|
| | | Control | Treatment | Control | Treatment | | |
| 1 | 3590.4581 | 619.7884 | 640.2116 | 68.9008 | 67.9243 | 24.7941 | 8.9151 |
| 2 | 4344 | 1250.1131 | 1269.8869 | 137.2229 | 136.4547 | 40.2259 | 14.742 |
| 3 | 4344 | 1887.2975 | 1892.7025 | 204.6775 | 205.084 | 93.2951 | 22.1092 |
| Average | 3956.4388 | 1003.1299 | 1036.0541 | 111.4037 | 110.0054 | 39.308 | 12.7206 |

**Response Generation Parameters**

No. of Hazard Pieces: 1

Input Method: Hazard Rates

| Piece # | Starting At | Control | Treatment | Hazard Ratio |
|---|---|---|---|---|
| 1 | 0 | 0.039 | 0.04 | 1.045 |

Note that 226 out of the 10000 simulations were unable to reject the alternative hypothesis, thus confirming (up to Monte Carlo accuracy) that this design achieves a type-1 error of $2.5\%$. Also, observe that 51.37% of these trials have crossed the futility boundary at the very first interim look after only 24.794 months of study duration.

# References

Abad-Santos F et al. (2005). Assessment of sex differences in pharmacokinetics and pharmacodynamics of almodipine in a bioequivalence study. *Pharmacologicial Research*, 51, 445-452.

Agresti A (2002). *Categorical Data Analysis*. (2nd Ed). John Wiley & Sons, New York.

Agresti A, Min Y. (2001). On small-sample confidence intervals for parameters in discrete distributions. *Biometrics* 57: 963-971.

Andersen EB (1990).*The Statistical Analysis of Categorical Data*. Springer-Verlag, Berlin-Heidelberg.

Anderson K (2002). Evaluating sponsor responsibilities for interim analysis with DMC's. Presented at the Clinical Trials Data Monitoring Committees meeting, Philadelphia Barnett International Conference Group, Philadelphia.

Anderson S, Hauck WW (1990). Consideration of individual bioequivalence. *J. Pharmacokin. Biopharm*, 18, 259-273.

Andrews DF and Herzberg AM (1985). *Data*. Springer-Verlag, New York.

Armitage P (1955). Test for linear trend in proportions and frequencies. *Biometrics*, 11: 375-386

Armitage P (1957). Restricted sequential procedures. *Biometrika*, 44, 9-56.

Armitage P (1975). *Sequential Medical Trials*. Blackwell Scientific Publications, Oxford.

Armitage P, McPherson CK and Rowe BC (1969). Repeated significance tests on accumulating

data. *J. R. Statist. Soc. A*, 132, 232-44.

# Index

## N



## O



## P

## R



## S

## Index

## U



## W