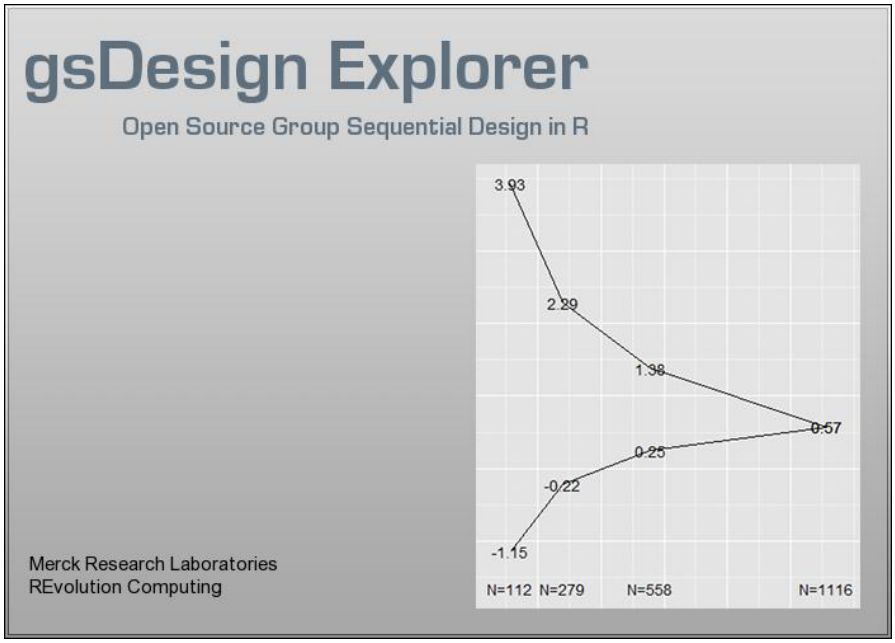


Version 1.0

User's Guide

Keaven M. Anderson, Merck Research Laboratories

William Constantine & Rich Calaway, REvolution Computing

January 2010

Contents

1

Overview

Group sequential methods allow clinical trial design with interim analyses to evaluate efficacy while controlling Type I error and power. Interim analyses offer opportunities for early stopping of trials, if, for example, the new treatment is demonstrably better than the standard treatment or clearly inferior. This can have benefits in many areas, from improved patient outcomes to significant savings in money and time.

Sequential methods in statistics have a fairly long history, but the modern theory of group sequential methods essentially begins with the papers of Pocock (1977) and O'Brien and Fleming (1979). These gave useful criteria for selecting boundaries for critical trials. Slud and Wei (1982) introduced the notion of *error spending*, in which the Type I error α is partitioned into probabilities $\pi_1, \ldots, \pi_k$ which sum to α, and π_k represents the probability of stopping at analysis k to reject the null hypothesis when this hypothesis is true, also called the "error spent at stage k." Lan and DeMets (1983) introduced the notion of *spending functions*, where Type I error is spent as a function of the observed information levels. When combined with a sampling rule aimed at continuing to a pre-specified maximum information level, the spending function approach can satisfy a given power condition quite accurately ((Jennison & Turnbull, 2000).

The gsDesign package in R (Anderson, 2009) uses the spending function approach to create a flexible tool for designing group sequential clinical trials comparing two treatment groups.

Among the supported spending functions are those of Lan and Demets (1983), Hwang, Shih, and DeCani (1990), Kim and DeMets (1987), and Anderson and Clark (2010). The package also supports user-written spending functions. Other important features of the package include the ability to design with non-binding futility rules, the ability to adapt designs based on changes to the timing and number of analyses performed. Group sequential designs considered here are based on asymptotic normal approximations to the joint distribution of interim test statistics. Thus, for "small" sample sizes, you may find it useful to simulate your trial to examine the actual Type I error and power; unfortunately, this facility is not yet built into gsDesign Explorer.

The gsDesign Explorer is a graphical user interface to the primary features of the gsDesign package. With gsDesign Explorer, you can instantly create, compare, and document multiple group sequential designs—you focus on the model parameters and their statistical and clinical implications, not the R syntax. The gsDesign Explorer provides high-quality graphical and textual summaries of group sequential designs that can immediately be incorporated into protocols and presentations and help you choose, justify, and document the best combination of design options for your trial.

This manual describes the operation of the gsDesign Explorer; for a more complete description of the calculations being performed behind the scenes, see the gsDesign manual (Anderson, 2009).

2

A Tutorial Introduction

This chapter introduces you to gsDesign Explorer by means of a brief tutorial that walks you through the following basic operations:

- Starting and stopping gsDesign Explorer
- Specifying Type I error limits and test power
- Specifying timing of interim analyses
- Sample size calculations
- Specifying spending function parameters
- Adding designs
- Saving and exporting designs
- Getting help

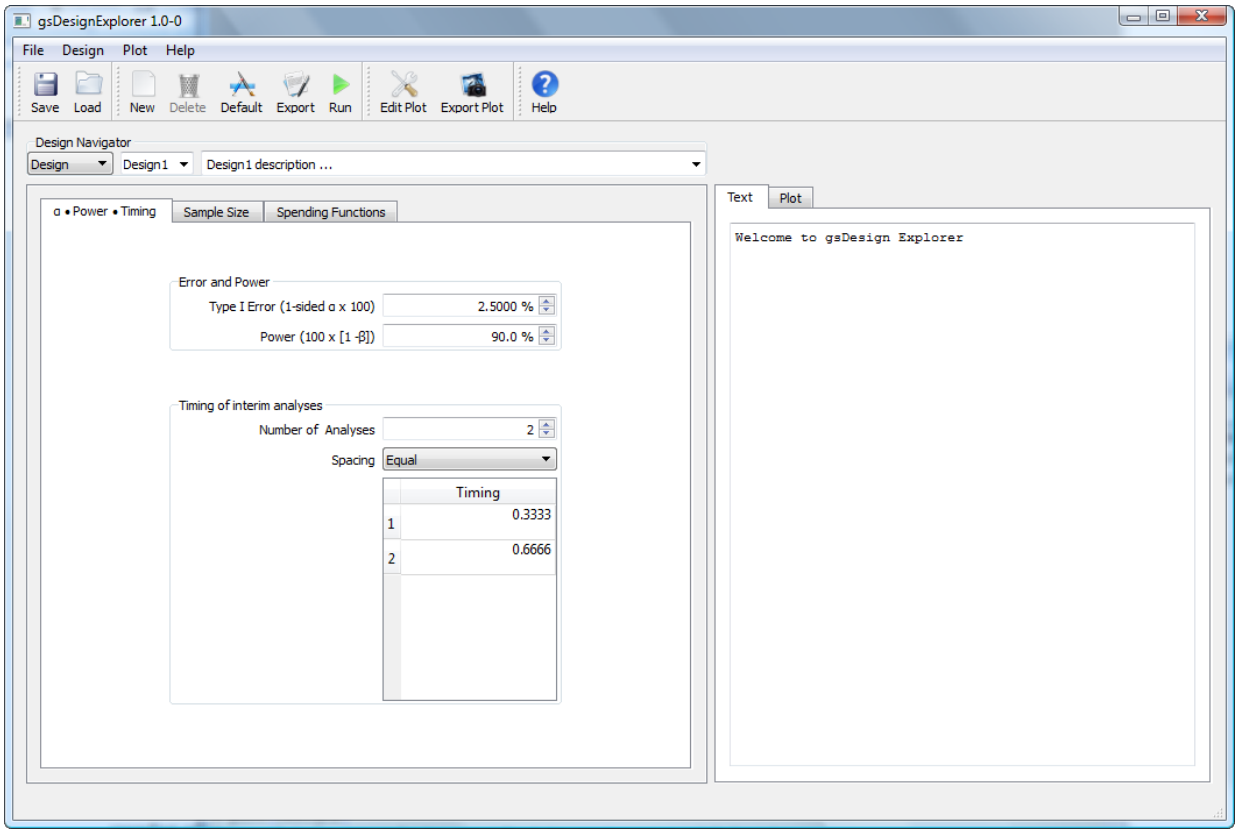
The gsDesign Explorer works with one or multiple designs at a time, and allows adaptation of interim analysis timing during execution of the trial (Lan & Demets, 1983). By allowing multiple designs, easy comparisons of multiple design candidates is simple, which enables you to quickly select the design that best meets your requirements.

Starting and Stopping gsDesign Explorer

To start gsDesign Explorer, type the following at an R prompt:

```
library(gsDesignExplorer)
gsDesignExplorer()
```

The gsDesign Explorer user interface appears as follows on Windows:

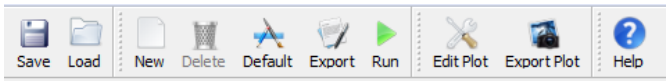


If you are running gsDesign Explorer from the standard R graphical user interface, you should minimize the R window once you've launched gsDesign Explorer so that focus does not change each time you run a design. In any event, the R Console prompt will be unavailable during your gsDesign Explorer session.

To close gsDesign Explorer, click the Close button on the window's title bar.

A Quick Tour of the gsDesign Explorer User Interface

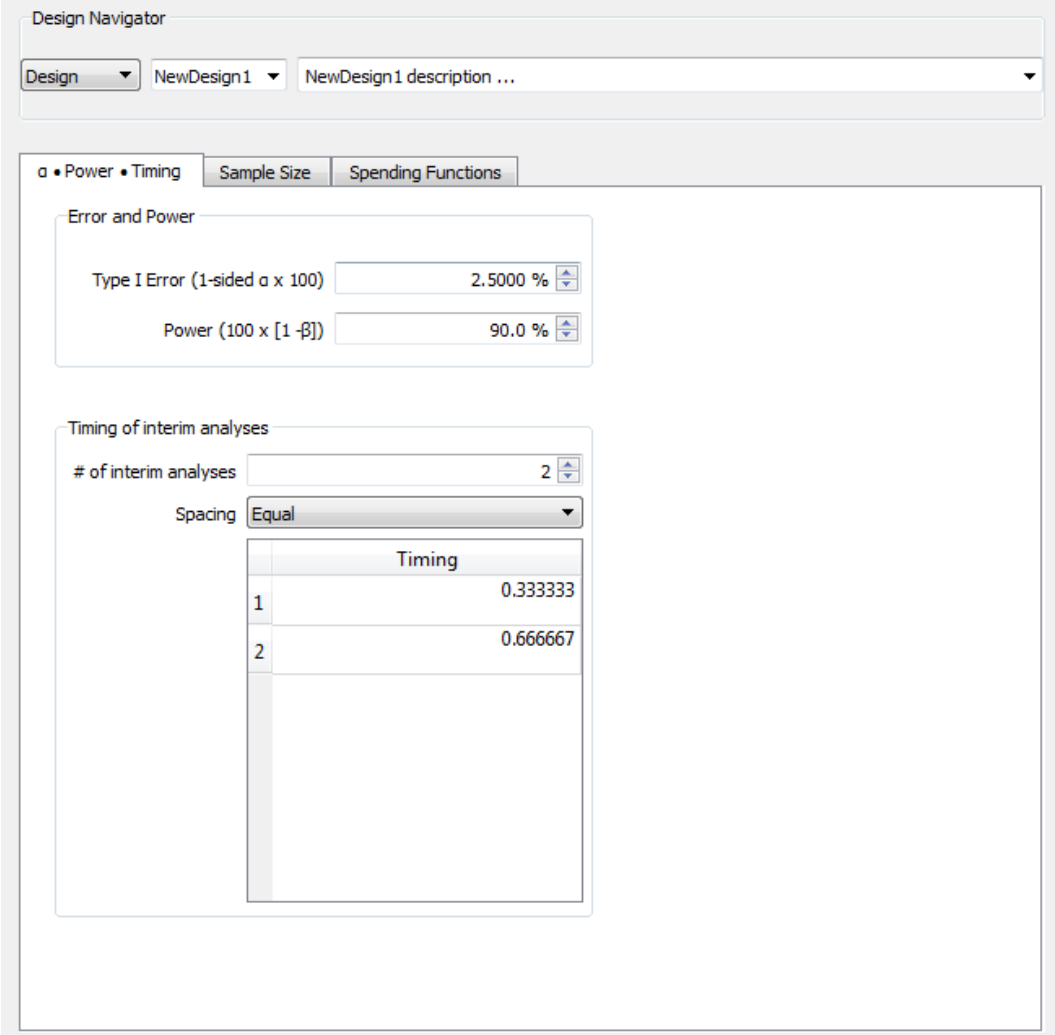
The gsDesign Explorer includes several menus, a toolbar, and two main panels, one for specifying model parameters, and one for showing text or graphical output. The toolbar appears as follows on Windows:



From left to right, the buttons are as follows:

- Save: Save the current set of designs to a file (in gsDesign format, with a .gsd extension).
- Load: Load a gsDesign format (.gsd) file.
- New: Create a new design.
- Delete: Delete the current design
- Default: Set all parameters in the current design to default values.
- Export: Export the current design as an R script.
- Run: Send the current design to R as a call to gsDesign and display results in the Output window.
- Edit Plot/Edit Design: Toggle this button to switch between editing the current design and editing the current plot.
- Export Plot: Save the current plot in any of the following graphic formats: pdf, png, jpg, tif, or bmp.
- Help: Get context-sensitive help on the application. Click on the help button to get the help cursor, and then click on the control for which you want help. (Help in the form of status line updates and tooltips appears when you hover over most gsDesign Explorer controls.)

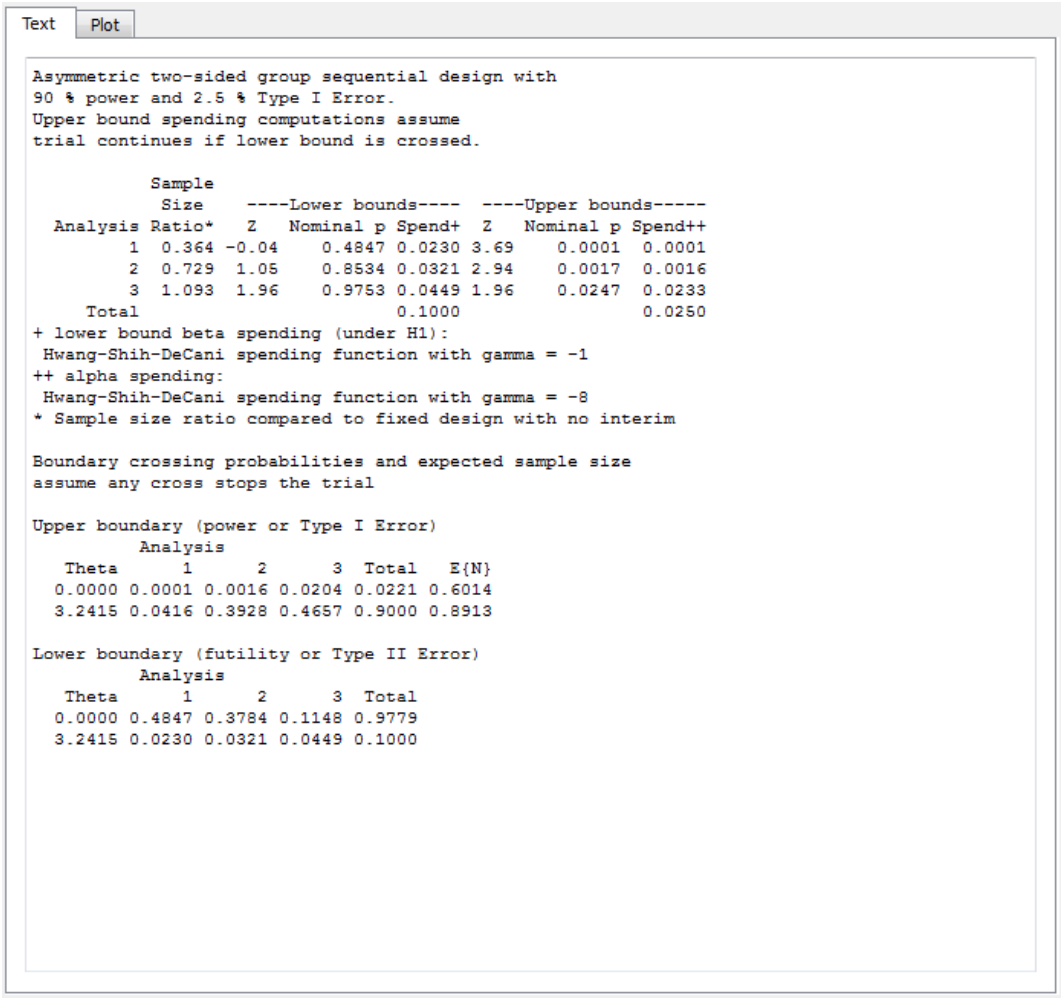
The central component of the gsDesign Explorer is the left pane, the Design Navigator. Here this pane is opened to the α-Power-Timing tab:



Using the Design Navigator, you can specify all the parameters necessary to create a group sequential design. The α-Power-Timing tab allows you to specify α, the 1-sided Type I error, and the *power* of the design, $1 - \beta$, where β is the Type II error. You also use this tab to specify the number and timing of interim analyses.

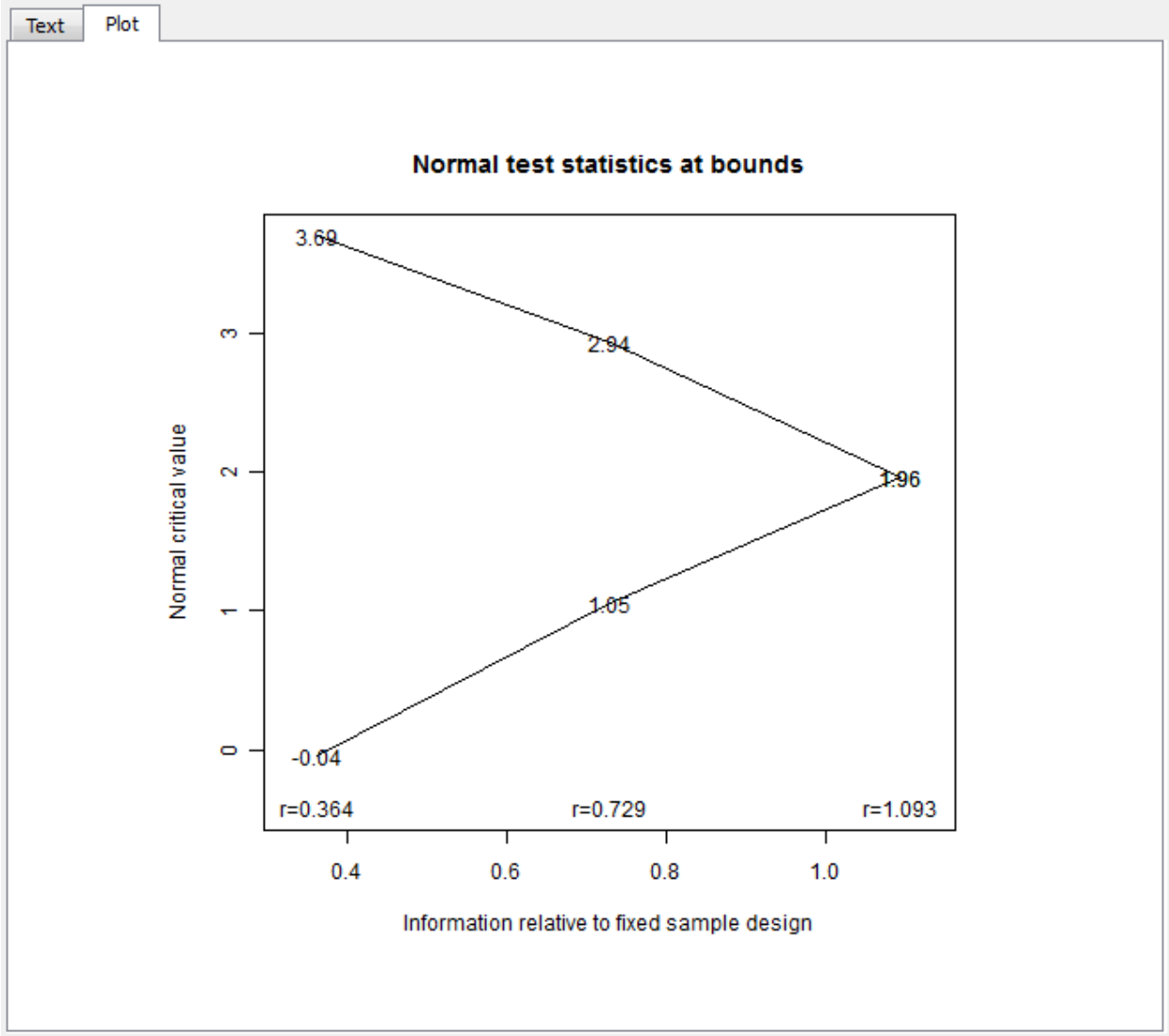
The Sample Size tab allows you to specify or calculate the required sample size for a fixed (non-group sequential) design with the same Type I and II error as the group sequential design you wish to create. This is used to transform a generic group sequential design with the same properties to one using the outcome specifications indicated. The Spending Functions tab allows you to specify various parameters for specifying spending functions and their associated parameters.

On the right side of the interface is the Output pane. Click Run upon opening the interface and you will see the Text tab of the Output pane populated as follows:



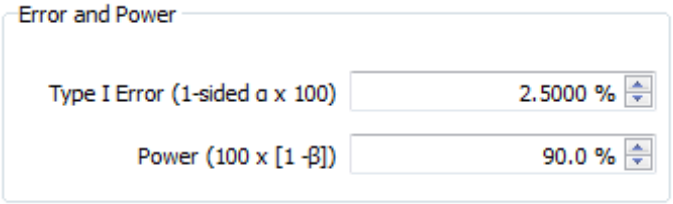
This is a generic design with upper and lower Hwang-Shih-DeCani spending functions with the indicated parameters. All designs with the same spending function parameters, relative timing of analyses, and error specifications have the same boundaries, with only sample size changing based on specification of outcome distributions in the two treatment groups.

Click the Plot tab to see (by default) a Boundaries plot for the design:



Specifying Type I Error

In clinical trials, one of the most important functions is to control Type I error in a manner acceptable to regulatory authorities. In gsDesign, Type I error is always one-sided, and is specified in gsDesign Explorer using the α-Power-Timing tab of the Design Navigator:



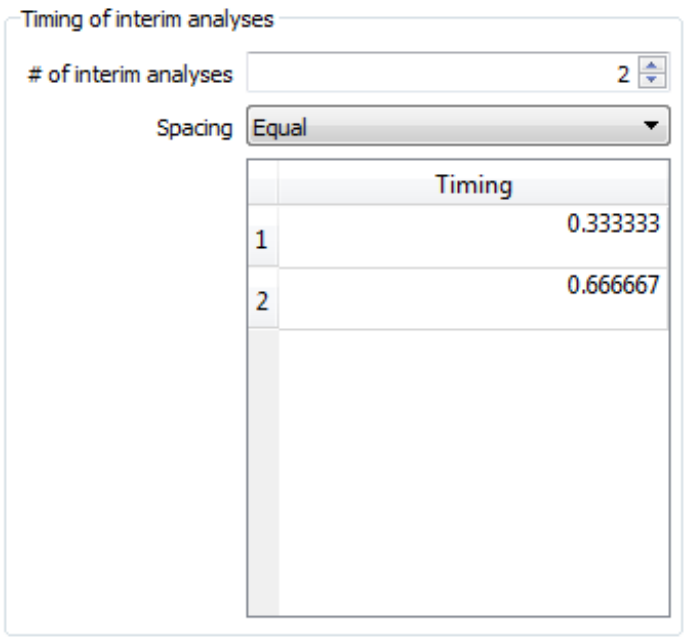
The default is 2.5 percent. The spinners allow you to modify the current value in increments of .1 percent; you can also enter values directly. For this tutorial, we accept the default.

Specifying Test Power

Directly beneath the Type I Error specification, you can specify the *power* of the test, which determines the allowable Type II error. (Power $= 1 - \beta$.) The default power is 90.0%. The spinners allow you to modify the power in increments of .5 percent, or you can enter values directly. For this tutorial, we set the power to 80.0%.

Specifying Timing of Interim Analyses

You also use the α-Power-Timing tab of the Design Navigator to specify the number and relative timing of interim analyses.

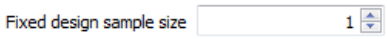


There is always a final analysis at the end of the trial denoted by a relative timing of 1. For interim analyses, you can specify either equally spaced analyses (the default) or unequally spaced analyses. The values specified in the Timing control correspond to the proportion of statistical information (for example, sample size or events) available for the data analyzed at each interim analysis. For example, the default of 2 equally spaced interim analyses corresponds to analysis performed after 33.3% and 66.7% of the way through the trial.

Calculating Sample Sizes

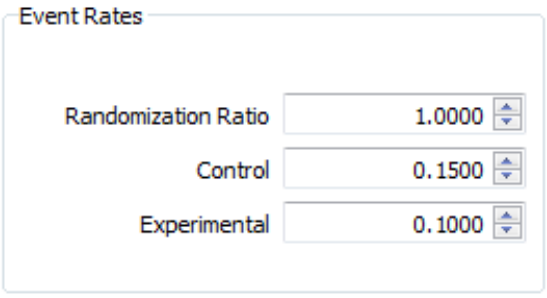
Use the Sample Size tab of the Design Navigator to specify or calculate the required sample size for a fixed design with no interim analyses and the same Type I error and power you specified for your group sequential design. This fixed design is then used to generate the sample size for a group sequential design, as described by, for example, Jennison and Turnbull (2000).

The User Input tab allows you to enter a sample size, or use the spinners to adjust a previously specified sample size:



You can use any sample size program (e.g., nQuery for many types of endpoints) to get a sample size for this tab. Setting the value to 1 results in a "generic" design that may be used with any sampling situation. Sample size ratios for timing of interim analyses relative to the fixed sample size are then generated for a group sequential design. For example, with sample size ratios of .54 and 1.08 and fixed design size of 100, the interim and final analysis would include 54 and 108 subjects, respectively.

The Binomial tab allows developing sample size for comparing the absolute difference in the rate of binary outcomes by specifying event rates for the control and experimental test groups, and the ratio of sample sizes in the two groups, from which a total sample size is calculated when you press Run. In the default case shown here, the Randomization Ratio is 1, the Control event rate is 0.15, and the Experimental event rate is 0.10, indicating that the expected result is that the treatment will reduce the probability of a primary endpoint by 1/3:

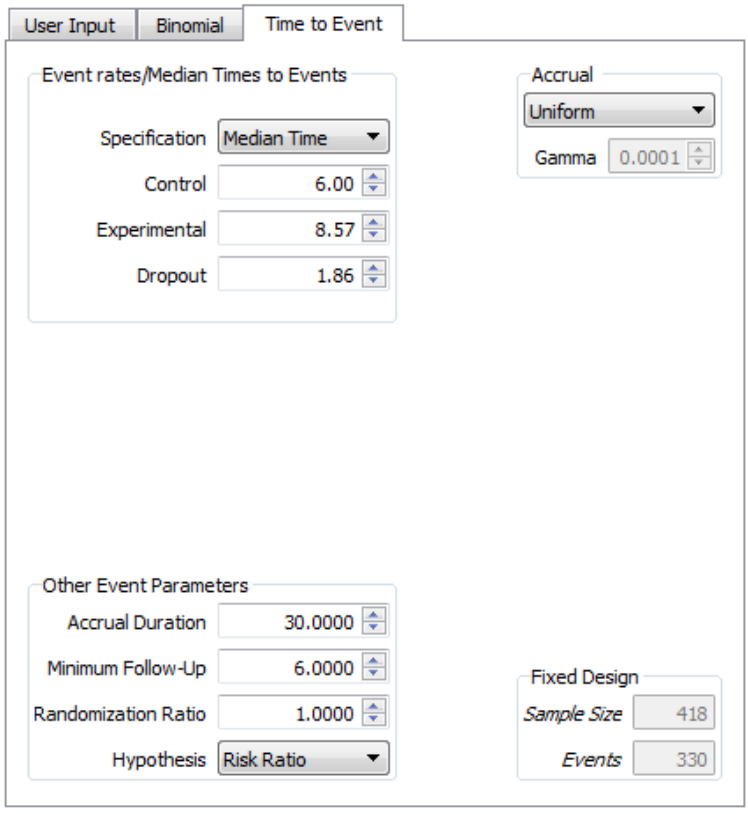


The Binomial tab also includes support for non-inferiority designs using the methods of Farrington and Manning (1990).

The Time to Event tab allows you to calculate the sample size and required number of events for a trial with a time-to-event outcome, using the method of Lachin and Foulkes (1986) and assuming the following distributions are known:

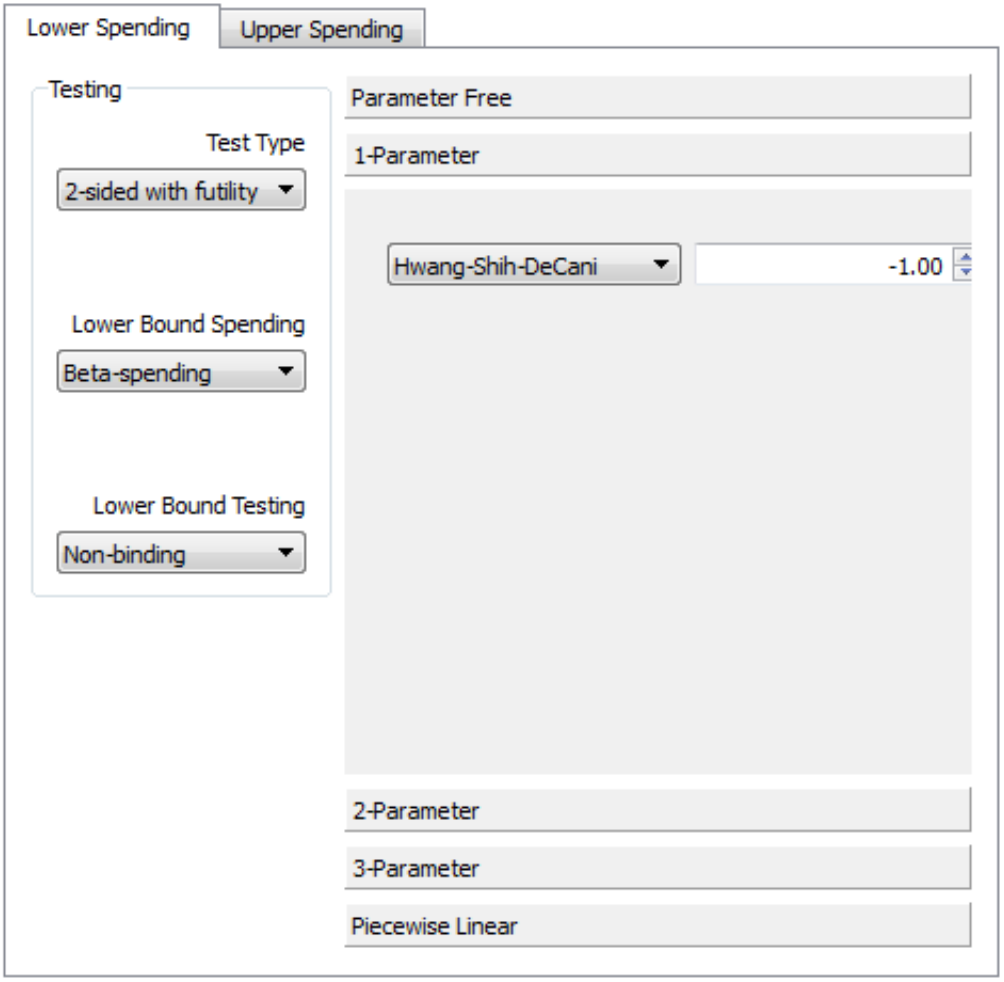
- The median time or exponential event rate for the time to a primary endpoint in each treatment group
- The median time or exponential event rate for the time until dropout
- Enrollment rate and duration
- Minimum follow-up time.

You can specify the first two quantities as either event rates or as a median time-to-event. For example, suppose you have a cancer trial in which the primary endpoint is the time from randomization until the first of disease progression or death (progression free survival or PFS), and that patients on the standard treatment are assumed to have a median PFS of 6 months. The treatment is assumed to reduce the hazard rate for PFS by 30%. Patients are assumed to drop out at a rate of 5% per year. In this case, enrollment is assumed to be uniform over 30 months with a minimum follow-up of six months. This model is reflected in the tab shown; when you press Run, the sample size and events required are calculated for you:



Specifying Spending Function Parameters

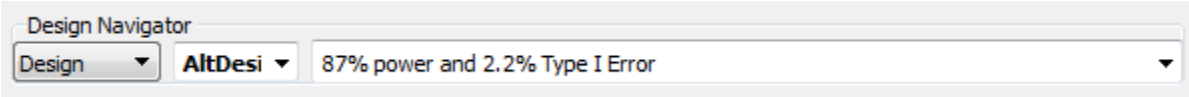
The Spending Functions tab of the Design Navigator offers you a wide choice of spending functions for the lower and upper bounds, as well as a range of test types. Certain test types imply specific spending functions. The default spending function for the lower bound is the Hwang-Shih-DeCani spending function with parameter -1. The default spending function for the upper bound is the Hwang-Shih-DeCani spending function with parameter -8. The spending functions available in gsDesign allow a great deal of flexibility in trial design and execution. The default Lower Spending tab is shown here; we will spend much more time on this tab (and its Upper Spending counterpart) in a later chapter.



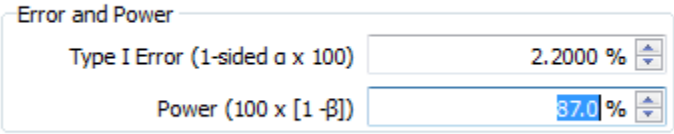
Adding Designs

One of the most powerful features of gsDesign Explorer is the ability to create multiple designs in a single interactive session, moving back and forth from one design to another to see how various modifications affect the sample size requirements and boundary characteristics.

To add a design, click New on the toolbar; this copies parameters from the design you are currently working with to a new design. By default, designs have names of the form *DesignN*, but you can give them any name you like. When you type a name, you see the text you enter in **bold**; this indicates the text is being edited. You must press Enter before leaving the control, or the input will not be recorded. You should also add a brief description for the design. (Again, you must press Enter before leaving the control, or your input will not be recorded.) Here, we give the name AltDesign and the Description "87% power and 2.2% Type I Error":



To make our description accurate, we modify the default design to specify 87% power and 2.2% Type I Error:



Saving and Exporting Designs

You can save designs for future use in gsDesign Explorer, or export them as R scripts that can be further customized or incorporated into other summaries. When you save, all current designs are written to a file with extension .gsd that can be loaded into gsDesign Explorer at a later time. The gsd format is strictly for use with gsDesign Explorer; if you want to save a design for later use with gsDesign in R, use Export instead. When you export a design, only the design currently in view is written as an R script. (Additional export formats may be available as well.)

To save your designs:

1. Either click the Save button on the toolbar, or choose Save from the File menu.
2. If the design has not previously been saved, specify a file name for the saved file.

To export a design:

1. Either click the Export button on the toolbar, or choose Export from the File menu.
2. Specify a file name for the exported R script file saving the currently viewed design.

Getting Help

Help is available for most controls in the gsDesign Explorer. To obtain help on a particular control, click the Help button in the toolbar, and then click the control. There are also hints in the lower left hand corner of the screen, and tooltips that pop up over each control when the mouse pointer hovers over the control.

This manual is also available through the gsDesign Explorer interface. To open the manual, click Manual in the Help menu.

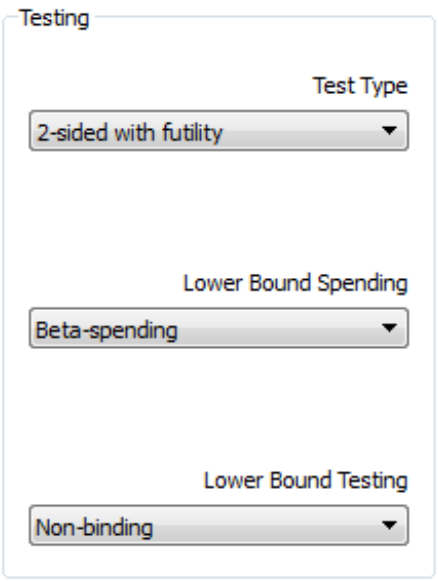
3

Specifying Spending Functions

The Spending Function tab is the heart of the gsDesign Explorer; by exploring the effect of different spending functions and spending function parameters on the design, you can gain valuable understanding of the trade-offs the various designs can bring in terms of maximum or expected sample size and boundary characteristics such as boundary crossing probabilities, approximate effect sizes at bounds, and conditional power at bounds.

Specifying the Test Type

You specify the test type using the Testing control on the Lower Spending tab of the Spending Functions tab:



Six test types are supported:

- 1-sided: If you only wish to test for a treatment difference in a single direction away from the null hypothesis, choose 1-sided from the Test Type control. In the one-sided case, only upper spending parameters are specified, so the remaining lower spending options are disabled. This option corresponds to specifying `test.type=1` in a call to the `gsDesign` function.
- 2-sided symmetric: If you wish to test for either treatment group being better than the other using the same criteria in either direction, choose 2-sided from the Test Type control. This would be used to compare, for instance, two approved treatments in order to decide which might be better. In the two-sided symmetric test, the parameters used for the upper boundary are also used for the lower boundary, so again the remaining lower spending options are disabled. This option corresponds to specifying `test.type=2` in a call to the `gsDesign` function.
- The other four types of lower bounds are for asymmetric testing and would commonly be used when testing a new treatment versus a standard control where there is no need to show the control is better than the new treatment. The lower bound is generally considered a futility bound for such studies, where a trial is discontinued early if the new treatment is unlikely to be shown to be superior to control based on interim results. This can also be used for non-inferiority studies to stop early if the new

treatment is unlikely to be shown to be non-inferior to control based on interim results. The most frequent choice for these options is to use a non-binding lower bound with β-spending, which is the default.

- By choosing a binding lower bound you assume the trial *must* be stopped if a lower bound is crossed. This option can be used to reduce the maximum sample size compared to options with a non-binding bound where the trial may continue and Type I error still be controlled if an upper bound is crossed following the crossing of a lower bound at an earlier interim analysis. Generally, a binding futility bound is frowned upon by regulators and this option reduces the flexibility of the Data Monitoring Committee for a study to base decisions to stop or continue on all available data, including other endpoints.
- The other option available for the lower bound is whether you wish to control interim lower boundary crossing probabilities under the alternative hypothesis (β-spending) or the null hypothesis (H0 spending). Typically β-spending is chosen here.

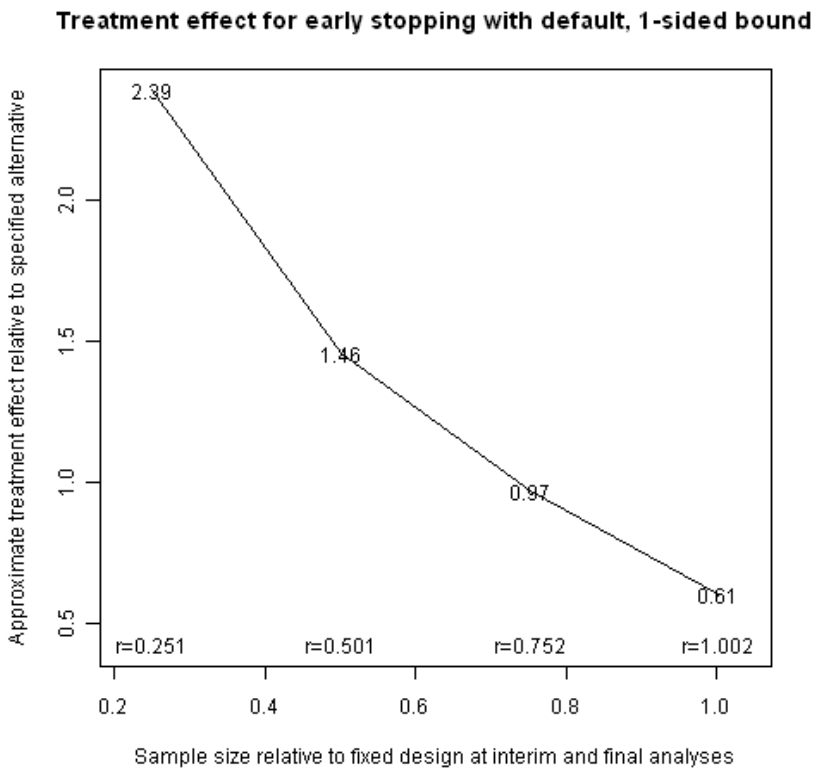
Specifying an Upper Spending Function

Different upper bounds are appropriate for different clinical situations. For a completely new treatment, stopping a trial early for less than a very convincing treatment difference would likely be a mistake; this would suggest using a very conservative rule for early stopping. More aggressive stopping rules for efficacy may be acceptable and even desirable in the case of a trial planned to confirm the results of an initial trial or a trial planned for a treatment already shown to work in a related indication. For the 1-sided and two-sided symmetric test types, the choice of upper spending function completely specifies spending. You use the Upper Spending tab of the Spending Functions tab to specify upper spending, which in combination with the interim timing specification determines early stopping boundaries convincing enough to have met trial objectives without continuing to the planned maximum sample size. A variety of spending functions are available, in five groups. The first four groups are organized according to the number of parameters required to specify the design; the fifth group, labeled Piecewise Linear to indicate that the upper boundary is specified as a sequence of points, actually includes just one function. Functions available in the remaining groups are as follows:

- Parameter Free: Two options are available: O'Brien-Fleming and Pocock. Developed by Lan and DeMets (1983), these spending functions can be used to approximate O'Brien-Fleming (1979) and Pocock (1977) designs, respectively.
- 1-Parameter: Three options are available: Hwang-Shih-DeCani (1990), Power (Kim & DeMets, 1987), and Exponential (Anderson & Clark, 2010).

- 2-Parameter (Anderson & Clark, 2010): Six options are available: Logistic, Beta Distribution, Cauchy, Extreme Value, Extreme Value (2), and Normal. For all options except the Beta distribution, you can specify the parameters in two ways: as a set of two points (the "Points" specification) or as a slope and intercept (the "Slope/Intercept" specification). For the Beta distribution, only the "Points" specification is available.
- 3-Parameter (Anderson & Clark, 2010): This group includes only one option, a t-distribution spending function that can be specified using the "Points" or "Slope/Intercept" methods just discussed. In addition, for each parameterization, the degrees of freedom for the t-distribution used is an option available to further customize the shape of the spending function.

The default upper spending function is a Hwang-Shih-DeCani spending function, which is a one-parameter spending function. The default value for the single parameter is -8; setting the parameter to -4 yields an O'Brien-Fleming-like upper bound, while setting it to 1 yields a Pocock-like upper bound. The value of -8 was chosen based on an assumption that early stopping is generally only desirable with a very convincing early result. This is demonstrated in the figure below generated by gsDesign Explorer. The figure assumes three equally spaced interim analyses. Treatment effects relative to that for which the trial is powered are plotted in the vertical direction. In order to stop after 25% of the maximum planned sample size, 2.39 times the planned effect must be observed. Even after 75% of the observations, essentially the full effect size (.97 out of 1) for which the trial was powered must be observed for early stopping. Note that the maximum sample size inflation relative to a fixed design of 1.002 on the x-axis indicates that sample size for this design is increased by only .2% compared to a fixed design. That is, if a fixed design required 1000 patients to achieve the desired Type I error and power, this group sequential design would require only 1002 patients.



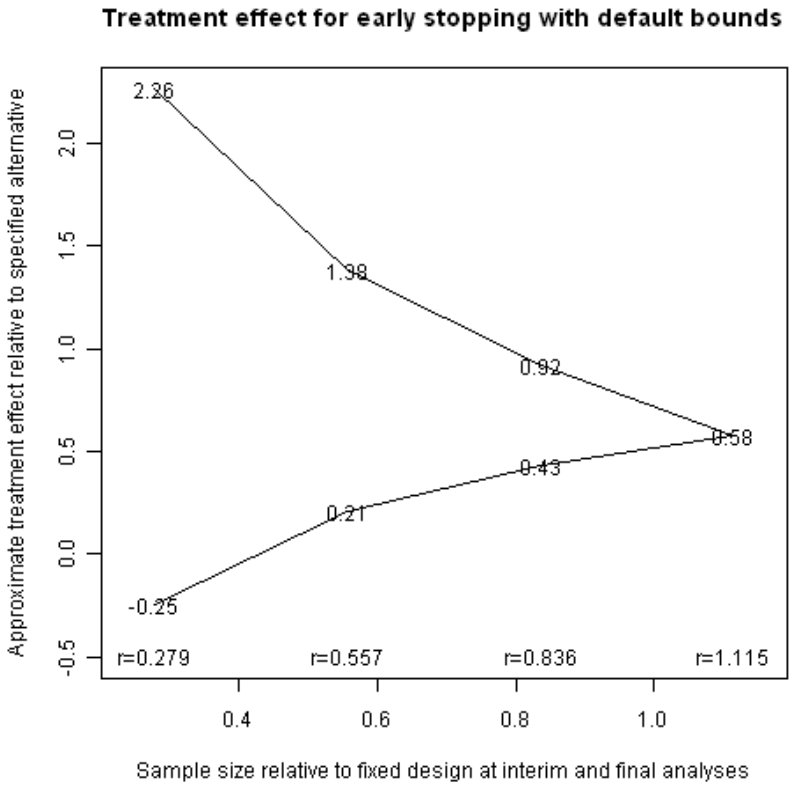
Other ways to obtain an O'Brien-Fleming-like upper bound are to use the 1-Parameter Exponential option and specify a parameter 0.75, or the 1-Parameter Power option with parameter 3.

Other ways to obtain a Pocock-like upper bound are to use the 1-Parameter Power option with parameter .8 or the 1-Parameter Exponential option with parameter .18 (this looks quite good with different numbers of interim analyses). Making such comparisons will be quite easy when we demonstrate the plot options later and toggle between designs.

Specifying a Lower Spending Function

If you have chosen one of the two-sided, asymmetric test types, you also need to specify a lower spending function. All of the spending functions available for the upper boundary are available for the lower boundary as well, but the default parameters differ. The default lower spending function is the Hwang-Shih-DeCani spending function with parameter -1.00. This provides a not-too-aggressive and not-too-conservative approach to early futility stopping. Approximate treatment effects for boundary crossing for the design with the default spending upper (HSD -8) and lower (HSD -1) functions and three interim analyses is shown below. Note that this design requires 11.5% sample size inflation relative to a fixed design. Choice of the futility bound and interim timing is critical, and is probably best when customized to suit the

clinical and regulatory scenario that your trial presents. For the scenario below, it may be desirable to alter the timing and number of interim analyses as well as the lower spending function to make appropriate futility stopping rules.



4

Visualizing the Design

We have already seen z-value and treatment effect plots summarizing design characteristics for group sequential designs. These plots appear in the gsDesign Explorer Plot tab of the output pane. By default a plot of the z-values at boundaries is shown. The Plot tab can be used to create several plots related to the design. In this chapter, we show the seven types of plots available and explore some of the options available for modifying the basic displays. We encourage you to use gsDesign Explorer while reading this chapter and suggest several options to try in addition to those displayed here.

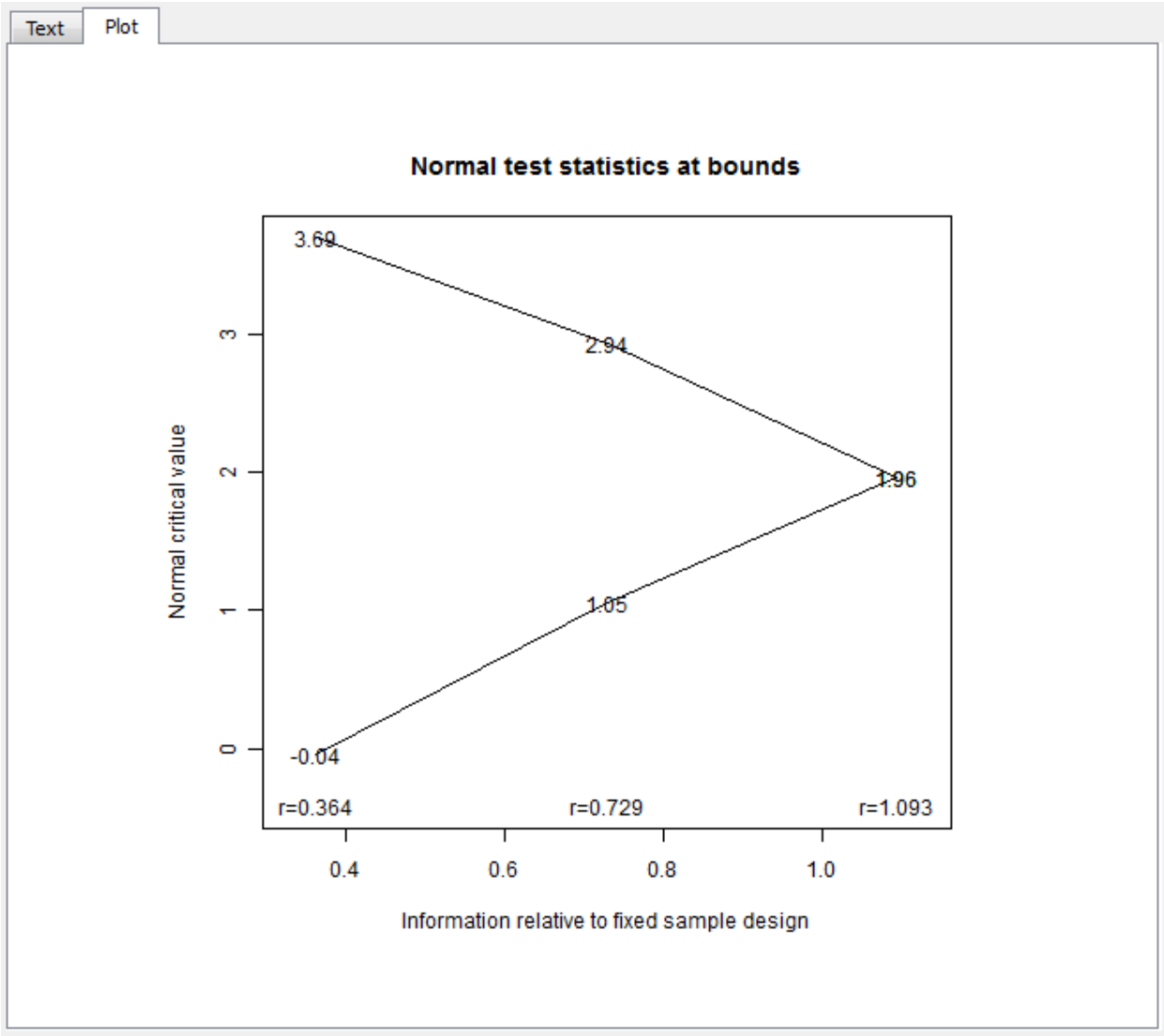
We demonstrate versions of the plots with the base graphics package here. On an experimental basis, high-quality plots using the ggplot2 package (Wickham, 2009) are also available. As ggplot2 is evolving, this may prove problematic, but you are encouraged to try this option. At present, gsDesign Explorer works with version 0.8.3 of ggplot2, but not with version 0.8.5. Thus, careful attention to your installation is important.

The following seven plots are available:

- z-value at boundaries
- power
- treatment effect
- conditional power
- spending function
- expected sample size
- B-values

The Boundaries Plot

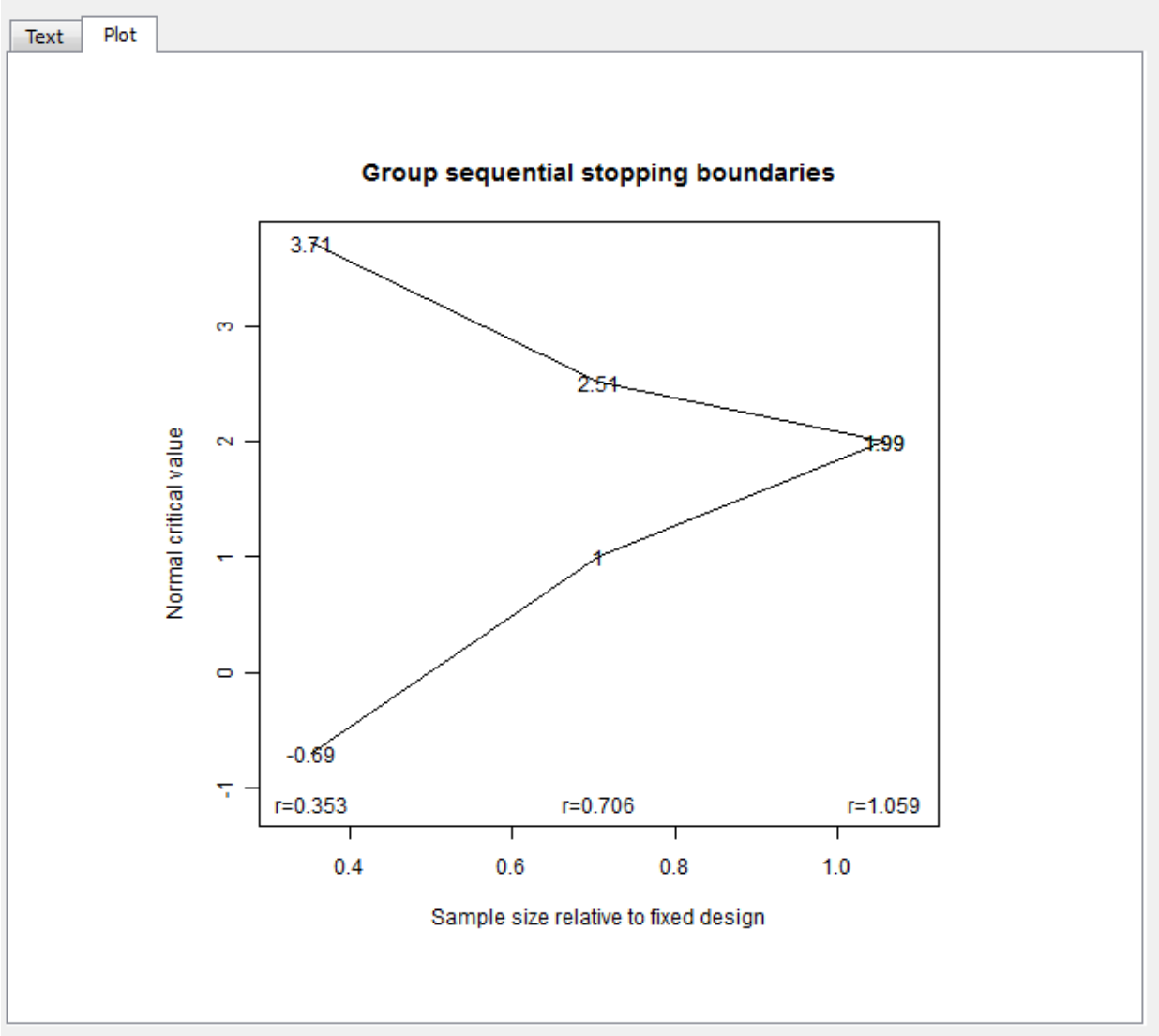
The default plot type shows the z-values at boundaries. In the default design, with the Hwang-Shih-DeCani spending function with parameter -8.00 as the upper spending function and the Hwang-Shih-DeCani spending function with parameter -1.00 as the lower spending function, the z-values at boundaries plot looks like the following:



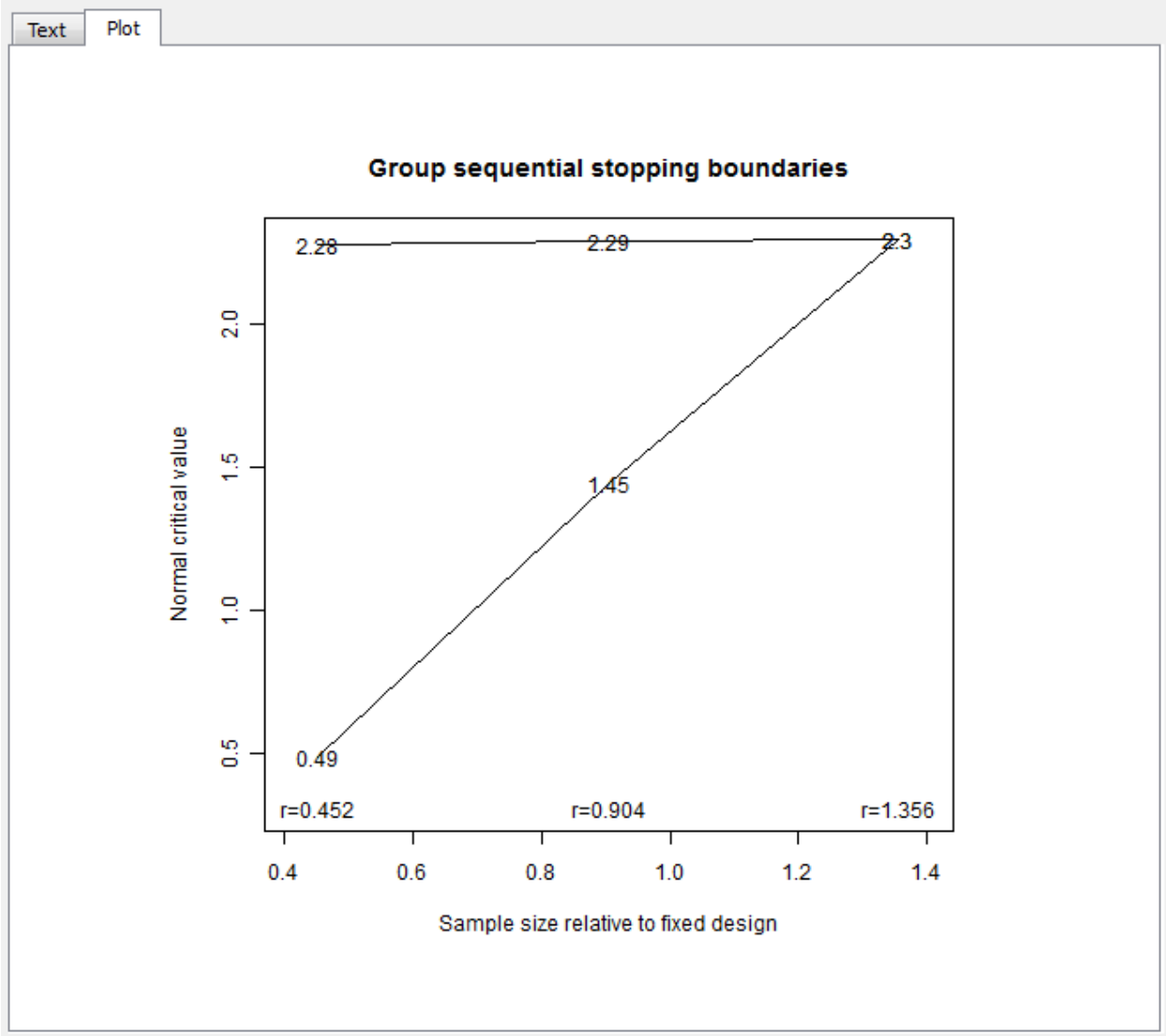
Note that there is no “p-value plot” and the nominal p-values at stopping bounds are available in the Text tab of the output pane. Click on the Text tab to see the p-values at bounds for the default design.

26 The Boundaries Plot

A z-value plot using the Lan-DeMets approximation to the O'Brien-Fleming bound as both the upper and lower spending function is shown below. This is different from the usual O'Brien-Fleming design which has symmetric bounds; try the 2-sided symmetric option to see a plot of those bounds. Being able to switch between such options and quickly observe their implications is a primary design feature of the gsDesign Explorer.

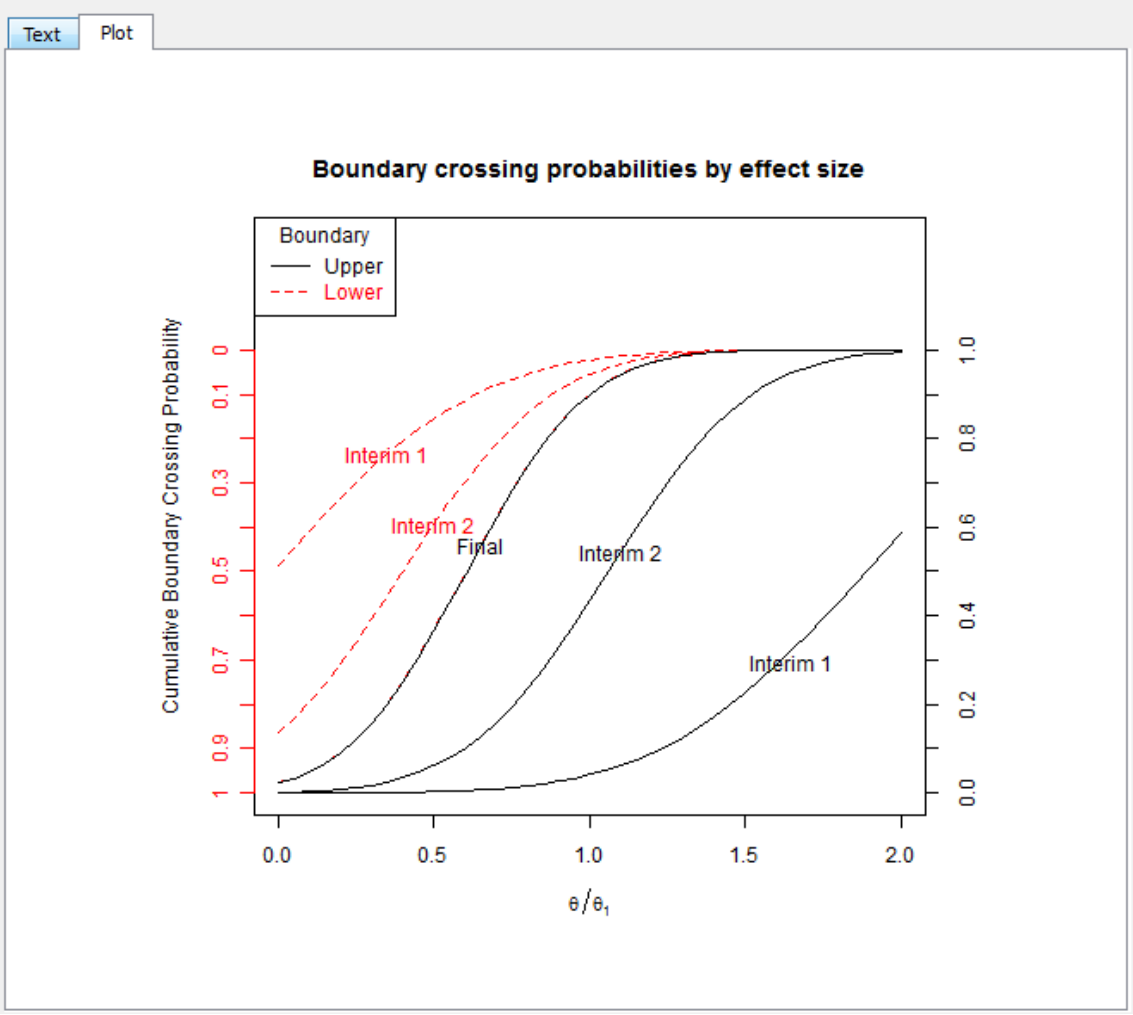


Using the Lan-DeMets approximation to the Pocock bound as both upper and lower spending function yields the z-value at boundaries plot shown below. Again, to see the prototypical 2-sided approximation, select 2-sided symmetric as the Test Type in the Lower Spending tab and re-run the plot; you will see the upper bounds do not change and the lower bounds become 2.28, -2.29, and -.3 at the first through third analyses, respectively. If you choose asymmetric testing with a binding lower bound, you will see the bounds below become a bit tighter and the sample size inflation at the end of the trial is reduced to a factor of 1.29. Try various options to see their implications. If you like a particular design, create a new design which will automatically copy that design's options; this saves the design of interest while allowing you to explore further.



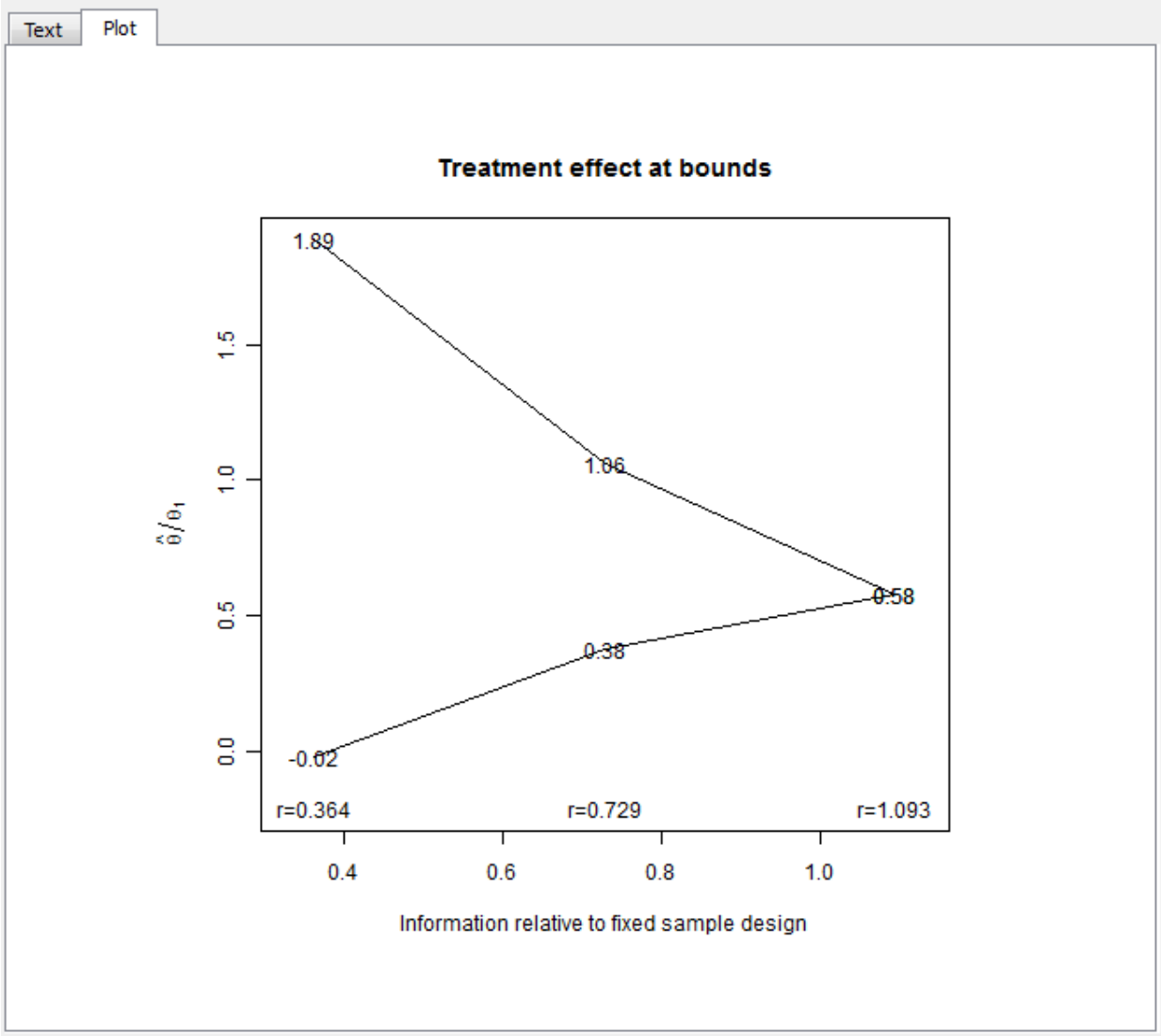
Power Plots

A group sequential power plot shows the cumulative probabilities of stopping at each analysis for different treatment effects. The solid lines show the upper boundary probabilities at each analysis, while the dashed lines show the lower boundary probabilities. The scale for the upper boundary crossing probabilities is on the right-hand side while the scale for the lower boundary crossing probabilities is on the left. Here we denote the treatment effect for which the trial is powered by θ_1 while the true underlying treatment effect is denoted by θ. The power and Type II error can be read from the value 1 on the x-axis on the "Final" curve. The power (.9) is on the right-hand axis while the Type II error (.1) is on the left. Where $\theta/\theta_1 = 0$, we can read the Type I error of .025 on the right y-axis. Cumulative upper and boundary crossing probabilities through each analysis can be read off the displayed curves. The "Final" curve provides total upper boundary (right y-axis) and lower boundary (left y-axis) crossing probabilities for the trial for each value of θ/θ_1, thus representing a power curve for the design.



Treatment Effect Plot

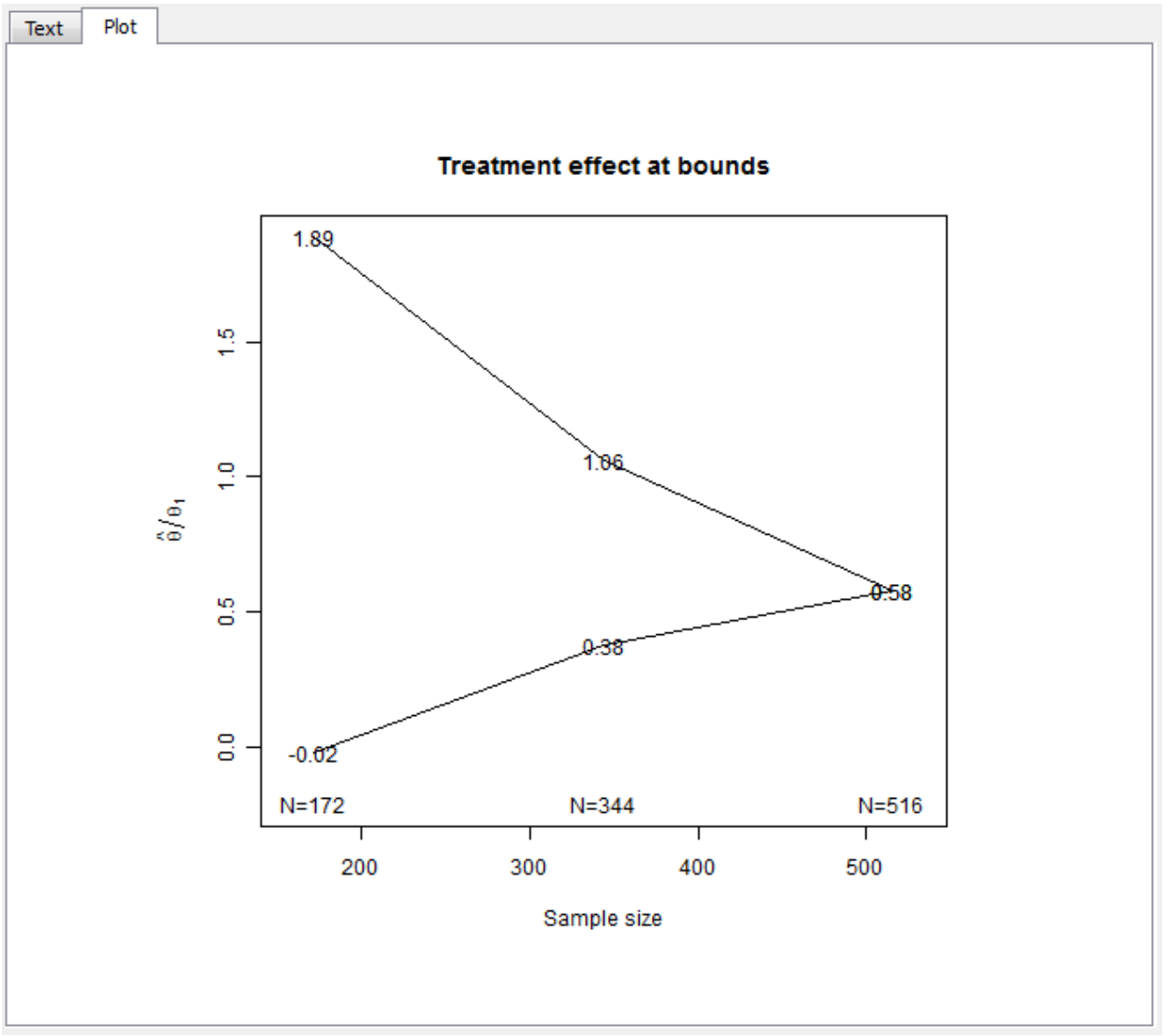
The treatment effect plot is analogous to the boundaries plot except that the *y*-axis approximates the treatment effect θ/θ_1, as defined above for the power plot, instead of the normal critical values. We have seen this plot previously when discussing spending function selection. The treatment effect plot for the default design is shown below. Note that at the first interim, an effect size that 1.89 times θ_1 is required to declare a positive result, while only .58 times θ_1 is required at the end of the trial.



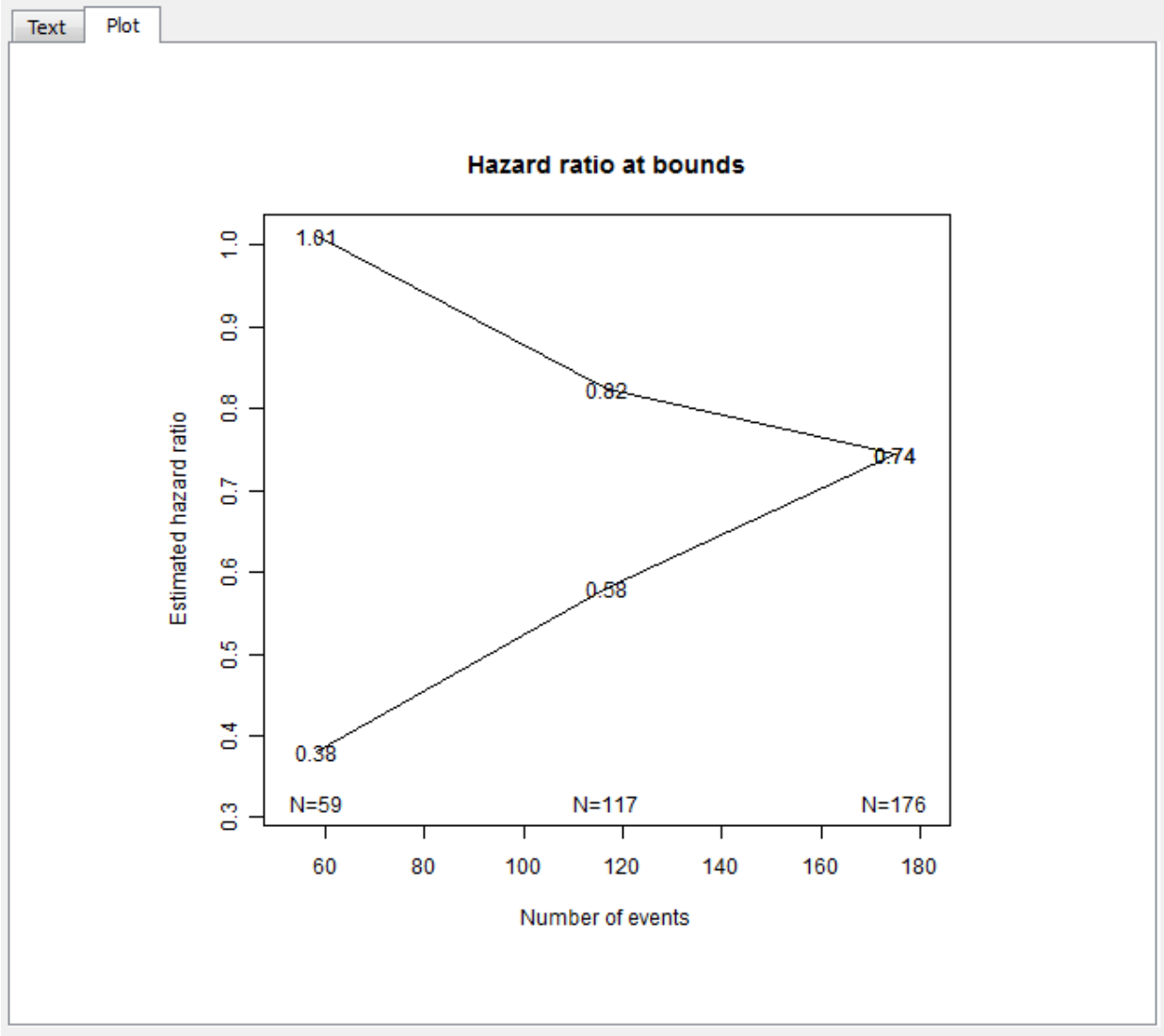
The treatment effect plots vary depends on which Sample Size tab is currently being studied. The plot above is the plot shown for the default model, which reflects the user-specified Sample Size tab.

30 Treatment Effect Plot

If you switch to the Binomial tab and press run, you obtain the following plot:

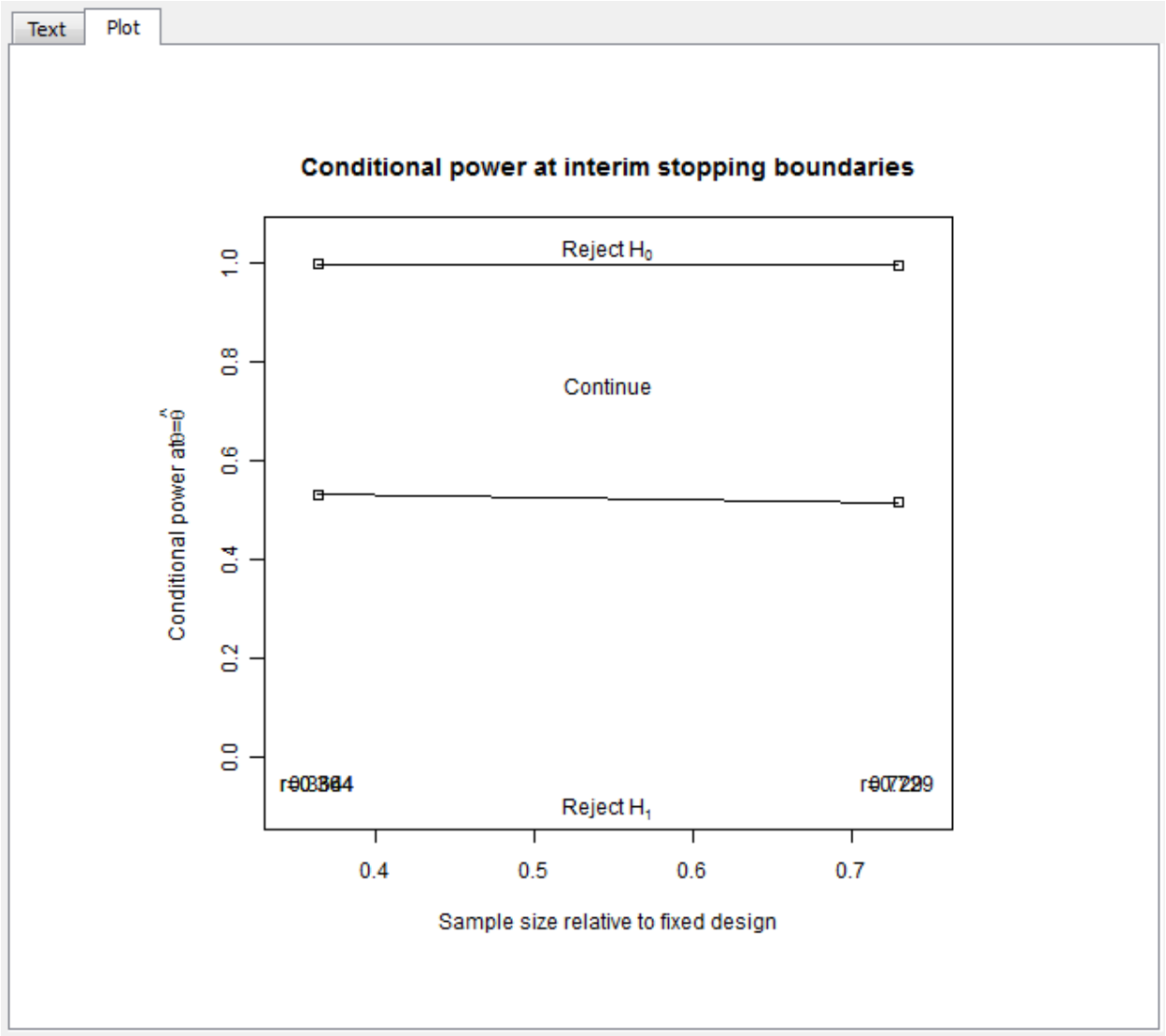


If you click the Time to Event tab and click Run, you obtain the following plot:



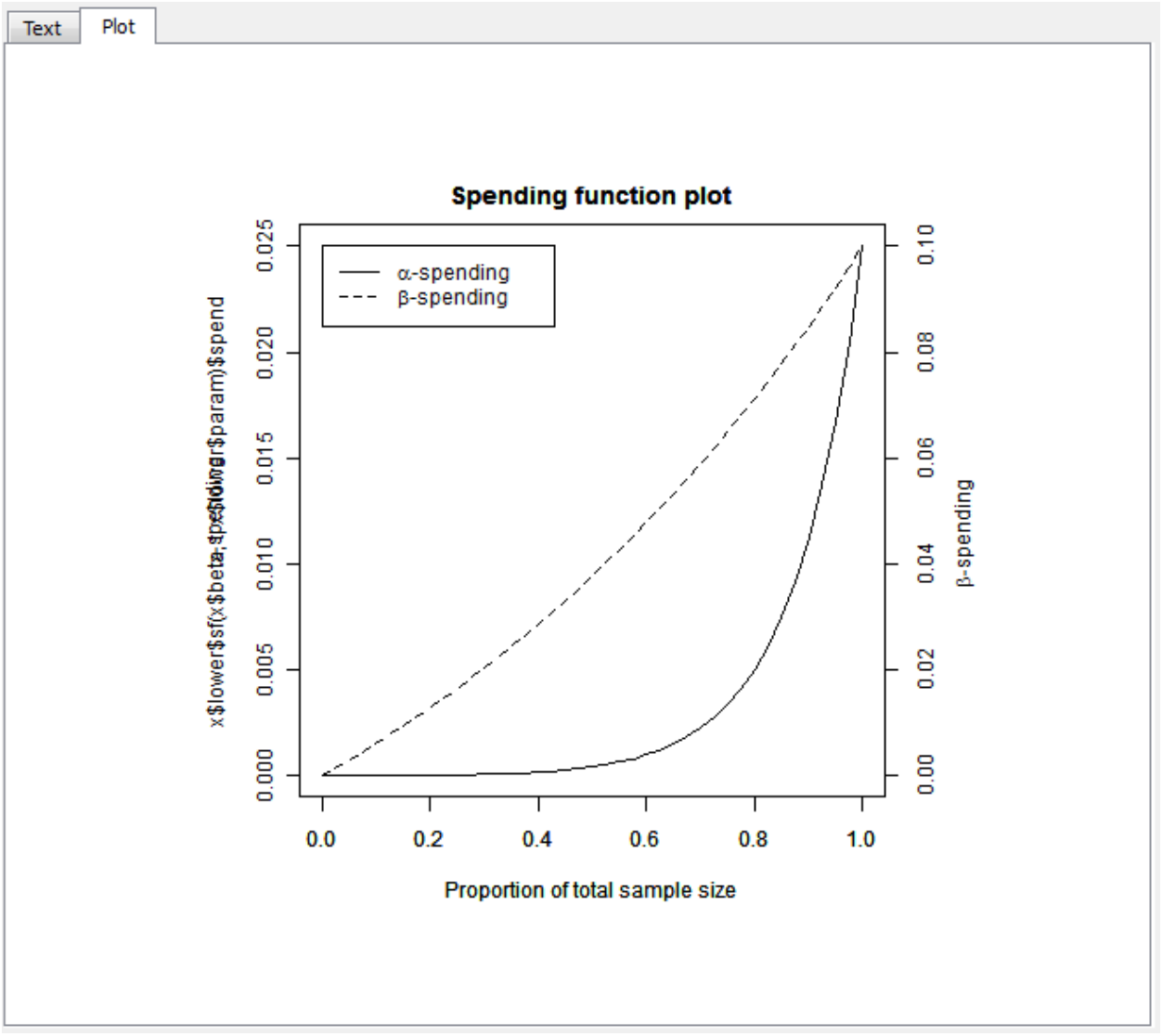
Conditional Power Plots

A conditional power plot graphs the conditional power at $\theta = 0$ versus the sample size relative to a fixed design. The conditional power plot for the default design is shown here:



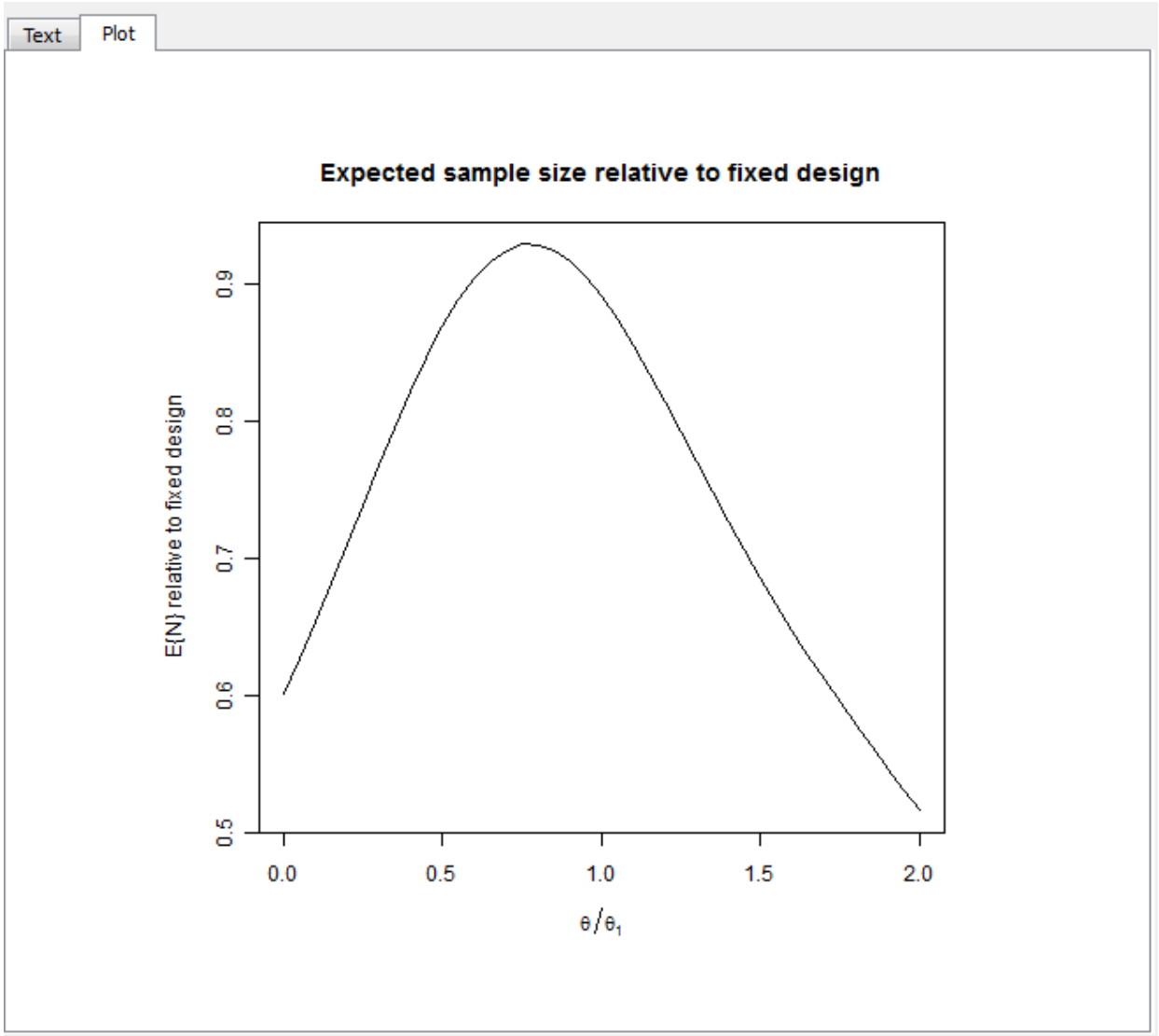
Spending Function Plots

Spending function plots show α- and β-spending plotted against the proportion of the total sample size. While shown on a common plot using base graphics, the α and β plots have different *y*-axes. The α *y*-axis is shown on the left, the β *y*-axis on the right. The spending function plot for the default design is shown below. There is little α-spending early reflecting our desire not to inflate the final sample size and to have conservative stopping rules for efficacy. There is more β-spending earlier, indicating a less conservative approach to early stopping. These facts were also reflected in the z-value and treatment effect plots shown earlier.



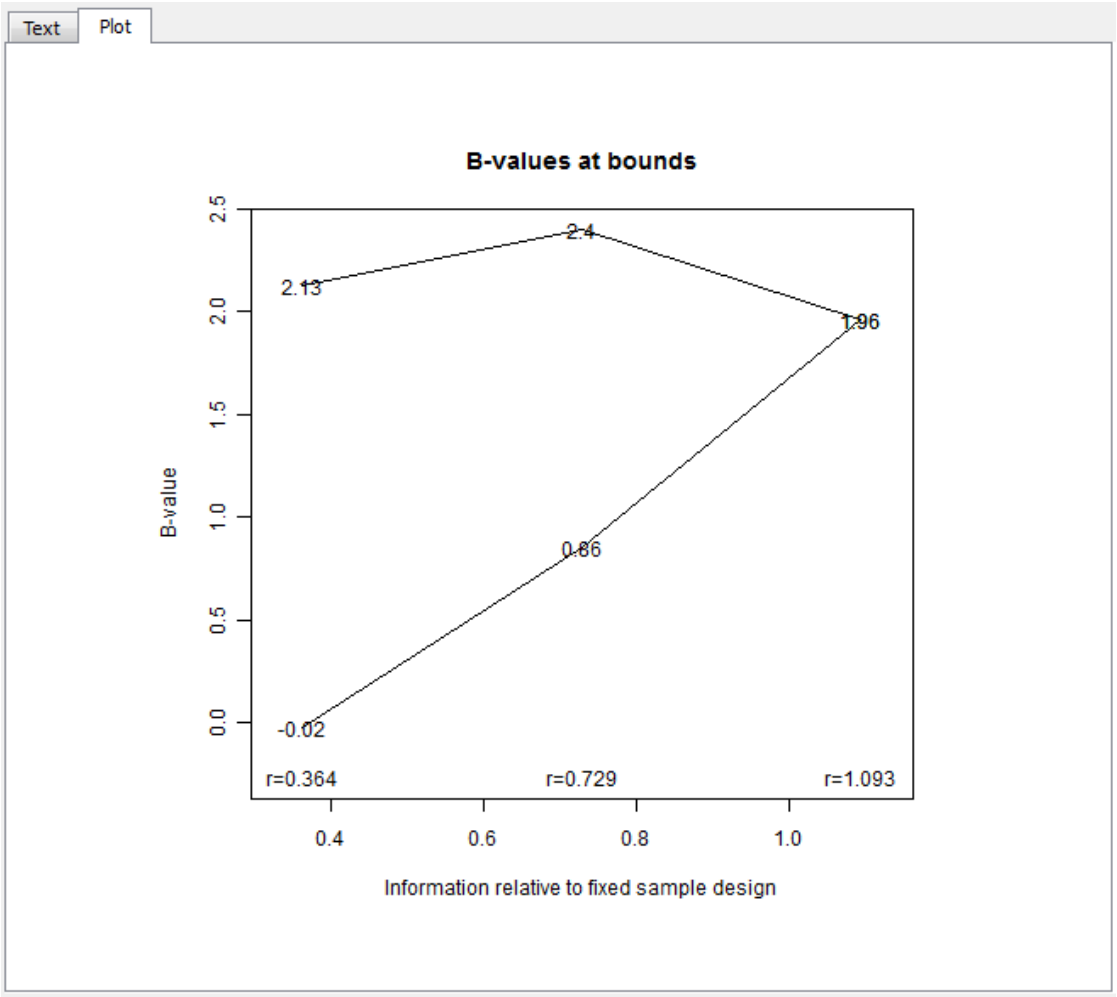
Expected Sample Size Plots

Expected sample size plots graph the sample size as a percentage of the fixed-design sample size versus the treatment effect θ/θ_1. In the default design, the maximum expected sample size is required when the treatment effect is about $.65\theta_1$, an effect size where it is difficult to choose between the null hypothesis $\theta = 0$ versus the alternative $\theta = \theta_1$. When the true effect size is substantially smaller or greater, the chance of crossing a bound and stopping the trial at an interim analysis substantially lowers the expected sample size at the analysis at which the trial terminates. Note that this calculation does *not* incorporate enrollment that may occur while an interim analysis is being prepared. If such enrollment is substantial, this plot may not reflect actual expectations.



B-Values Plots

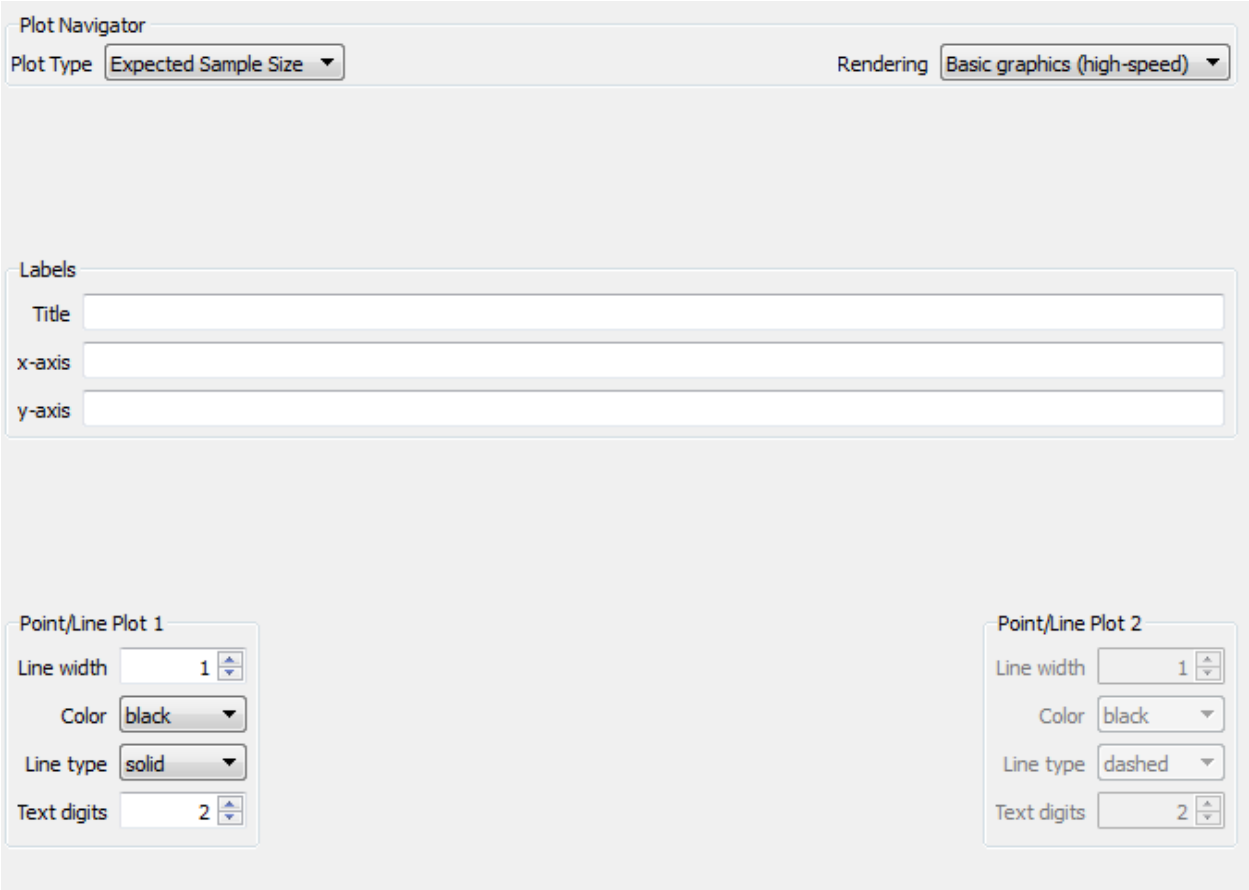
B-values are z-values multiplied by $\sqrt{t}$, where $\sqrt{t}$ is the proportion of the final planned sample size required at each analysis. The expected value of a B-value at an analysis is $\theta \times t$, where θ represents the standardized treatment effect (Proschan, Lan, & Wittes, 2006). The value of θ under the alternative hypothesis is denoted by θ_1 in this manual and in default plot labeling in the gsDesign package; this is at variance with the variable name `delta` returned in the R objects created by the `gsDesign` function. B-values behave randomly about the trend line $\theta \times t$ during the course of the trial. Projections for future B-values can be made by projecting forward with a line through the origin and the current observed interim B-value (current trend) or by using a slope from the current interim point that is equal to θ_1, the assumed treatment effect under the alternative hypothesis; note that in this case the slope needs to be adjusted to account for the x-axis value of r or N (that is, multiply the slope by r/t or t/N). The B-values plot for the default design is shown here:



Editing Plots

You can edit many aspects of the plots produced in gsDesign Explorer using the option Edit Plot on the Toolbar. When you are in Edit Plot mode, you cannot change the Design mode (that is, move from Design to Analysis or vice-versa).

When you select Edit Plot, the Plot Navigator shows the controls seen here:



Beneath the options for changing plot type and rendering mode, there are text boxes for editing the plot's various labels, including the main title, the *x*-axis label, and the *y*-axis label. In the section labeled Point/Line Plot 1, you can specify the line width, color, line type, and the number of digits shown in text output. The section labeled Point/Line Plot 2 is enabled when two or more plots are displayed at once, as in Power plots and Spending Function plots. Again, this gives you control over line width, color, line type, and displayed digits.

When entering labels, you can add math characters such as operators and Greek letters by enclosing them in the control sequence mt(). For example, to create a main title "Plot of α Spending", you would type "Plot of mt(alpha) spending" in the Title text box.

A slightly more challenging example is to create the following expression:

$$\bar{x} = \sum_{i=1}^{n} x_i$$

You can do this with the following call to `mt` (note the == to obtain the equal sign):

```
mt(bar(x)==sum(x[i], i==1,n))
```

The available expressions are described in the R help topic “Mathematical Annotation in R”; type ?plotmath at an R prompt to view this topic. (As mentioned earlier, the R Console is unavailable while you run gsDesign Explorer, so you may want to launch this help topic before starting gsDesign Explorer if you think you’ll be modifying graphics labels.)

The Rendering control enables you to control whether the graphs are rendered at high-speed, using R’s basic plotting engine, or at high-quality, using Hadley Wickham’s ggplot2 package (Wickham, 2009); as note previously, the ggplot2 option should be considered experimental at the time of this writing. The default in gsDesign Explorer is to use the basic graphics; the current gsDesign default is to use ggplot2 graphics.

5

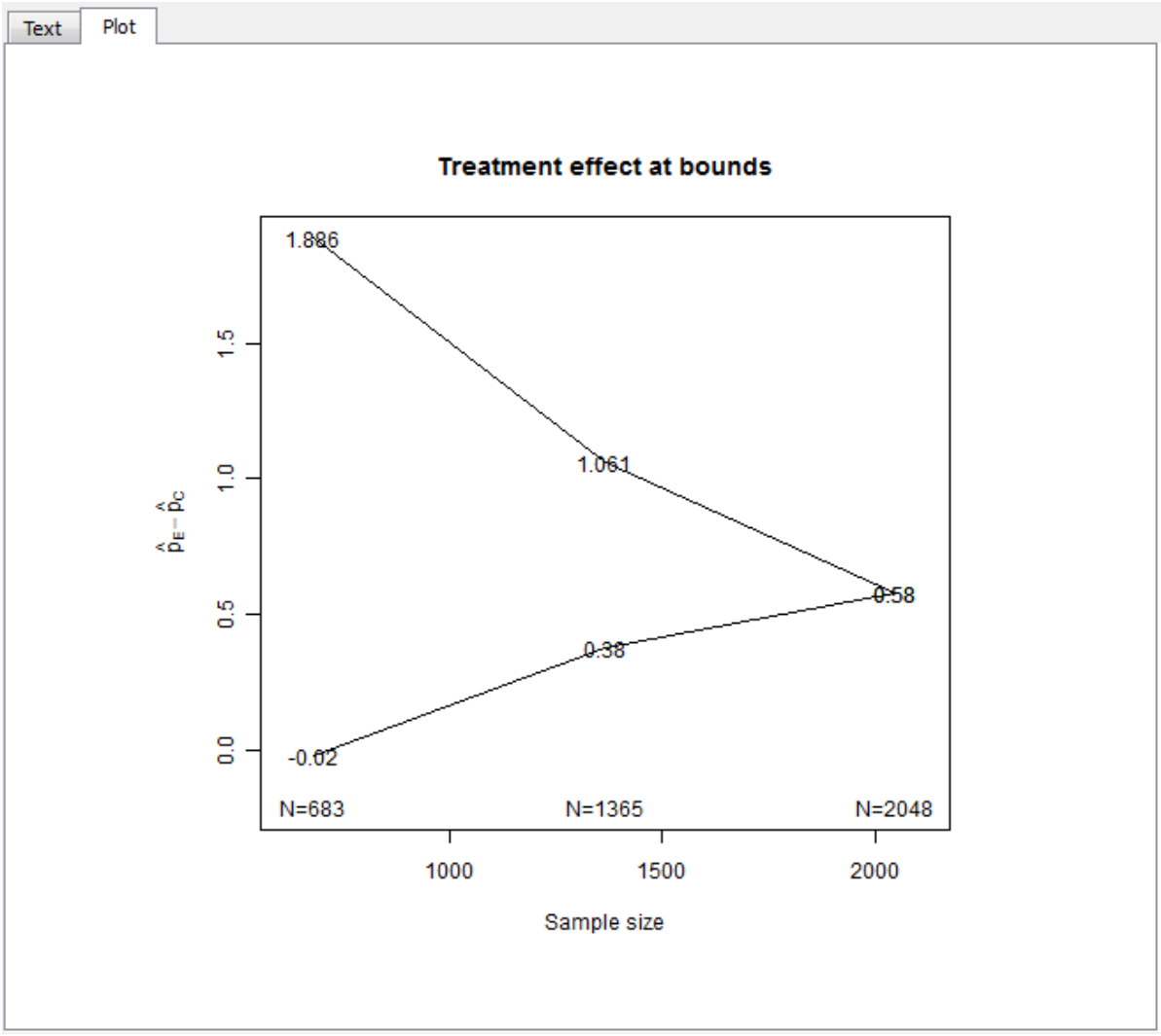
Example: Non-inferiority testing

This is an “advanced” topic that not all will wish to review. However, we demonstrate a rather simple implementation of a non-inferiority group sequential design for a binomial outcome. Consider a trial examining a new drug that is more convenient to administer than an approved control. There is no expectation of a substantially improved response with a new drug. While the new drug may be a little better or worse than control, there is some suggestion that the new drug may not be as efficacious as control. Rather than powering the trial to show non-inferiority when the new drug is slightly worse than control, the strategy is taken to stop the trial early for futility if there is a “substantial” trend towards the new drug being inferior.

Non-inferiority designs for binary outcomes have been implemented in gsDesign Explorer and we document their use here. Assume the control drug has provided a (binomial) response rate of 67.7% in a past trial and regulators have agreed with a non-inferiority margin of 7%. Let the underlying event rate in the control and experimental groups be denoted by p_C and p_E, respectively. Let $\delta_0 = .07$ represent the non-inferiority margin of 7%. This means our null hypothesis is $H_0: p_E - p_C + .07 = 0$ and the alternative is $H_1: p_E - p_C = 0$. While there is no desire to stop the test early to establish non-inferiority, we include a “high” efficacy bar for an early stop for a positive efficacy result.

When setting this up, we select the Sample Size tab and within that, the Binomial tab. We enter event rates of 0.677 for both control and experimental, and then specify “Non-inferiority/sup with margin” in the Testing control. Finally, we enter .07 as the non-inferiority margin Delta. Click Run to generate the design. We have used the standard 2.5% Type I error and 90% power.

Next, we show a treatment effect plot. This indicates that after the first interim analysis with 688 subjects, we may stop and declare non-inferiority if experimental therapy has about a .039 higher success rate. At the second interim, no difference is sufficient (control .005 better or less), while at the final analysis we may declare non-inferiority with a .028 rate difference favoring control. Futility to demonstrate non-inferiority is declared if the comparison favors control by a margin of about .079 or more at the first interim and .046 or more at the second interim. These differences require careful evaluation in choosing the appropriate spending functions.



Bibliography

Anderson, K. M. (2009). *gsDesign: An R Package for Designing Group Sequential Clinical Trials.* Version 2.0 Manual. Merck Research Laboratories. http://cran.r-project.org/web/packages/gsDesign/.

Anderson, K. M., & Clark, J. B. (2010). Fitting Spending Functions. *Statistics In Medicine . To appear.*

Farrington, C., & Manning, G. (1990). Test statistics and sample size formulae for comparative binomial trials with null hypothesis of non-zero risk difference or non-unity relative risk. *Statistics in Medicine* (9), 1447-1454.

Hwang, I. K., Shih, W. J., & DeCani, J. S. (1990). Group sequential designs using a family of type I error probability spending functions. *Statist. Med. , 9*, 1439-1445.

Jennison, C., & Turnbull, B. W. (2000). *Group Sequential Methods with Applications to Clinical Trials.* Boca Raton, FL: Chapman & Hall/CRC.

Kim, K. M., & DeMets, D. L. (1987). Design and analysis of group sequential tests based on Type I error spending rate functions. *Biometrika* (74), 149-157.

Lachin, J. M., & Foulkes, M. A. (1986). Evaluation of Sample Size and Power for Analyses of Survival with Allowance for Nonuniform Patient Entry, Losses to Follow-up, Noncompliance, and Stratification. *Biometrics* (42), 507-519.

Lan, K. K., & Demets, D. L. (1983). Discrete sequential boundaries for clinical trials. *Biometrika* (70), 659-663.

O'Brien, P. C., & Fleming, T. R. (1979). A multiple testing procedure for clinical trials. *Biometrics* (40), 549-556.

Pocock, S. J. (1977). Group sequential methods in the design and analysis of clinical trials. *Biometrika* (64), 191-199.

Proschan, M., Lan, K., & Wittes, J. (2006). *Statistical Monitoring of Cliinical Trials. A Unified Approach.* New York: Springer.

Slud, E. V., & Wei, L.-J. (1982). Two-sample repeated significance tests based on the modified Wilcoxon statistic. *J. Amer. Statist. Assoc.* (77), 862-868.

Wickham, H. (2009). *ggplot2: elegant graphics for data analysis.* New York: Spring.