Software for design and analysis of group sequential clinical trials with multiple primary endpoints

Department of Biostatistics and Medical Informatics Technical Report 177

Shuangge Ma, M.S.[a], Michael R. Kosorok, Ph.D.[a,b] [1], Thomas D. Cook, Ph.D.[b]

[a] *Department of Statistics, University of Wisconsin, Madison*

[b] *Department of Biostatistics & Medical Informatics, University of Wisconsin, Madison*

Abstract

In many phase III clinical trials, it is desirable to separately assess the treatment effects on more than one primary endpoint. C++ programs are provided for the computation of sample size, critical boundaries and test statistics for group sequential clinical trials with one or two primary endpoints. Our programs cover both the design and analysis phases. The computations are proper for any trial based on normally distributed test statistics, including those in which patients give a continuous response and in survival studies. All computations are based on the methodology proposed by Kosorok et al [1]. This approach resolves the problems of how to stop early when the treatment effect is clear in all endpoints and how to control the many possible error rates for concluding wrong hypotheses. We apply our software to the design and interim analysis of the Copernicus trial, which tests the effect of the beta-blocker carvedilol on the survival of patients with severe heart failure. It is shown that with certain marginal α and β spending functions, we could conclude a statistically significant beneficial treatment effect in both endpoints at the fourth interim analysis.

Keywords. Alpha-spending function; Beta-spending function; Multivariate outcomes; Time-to-event Data.

[1] **Corresponding author: Dr. Michael R. Kosorok, Departments of Statistics and Biostatistics & Medical Informatics, University of Wisconsin-Madison, 1210 West Dayton Street, Madison, WI 53706. Tel: +1-608-262-2537. Email: kosorok@biostat.wisc.edu**

Introduction

The design and analysis of group sequential clinical trials with more than one primary endpoint has been of considerable interest in recent years. In many phase III clinical trials, it is desirable to assess the treatment effect on multiple primary endpoints separately. Consider the Copernicus trial, where the primary endpoints of interest are all cause mortality and the earliest time of all cause mortality or hospitalization. Previous work does not resolve some important issues, including how to stop early when at least one of the endpoints has no treatment effect or how to control the many error rates for concluding wrong hypotheses. Kosorok et al [1] develops general methodology that resolves this issue. The essence of this approach is to set up group sequential critical regions that address the questions of interests for each primary endpoint while allowing the study to stop early if all priority questions have been addressed. Meanwhile, the probability of wrong conclusions is kept low. A global $\alpha-$spending function is used to control the overall type I error and a multiple decision rule to control error rates for concluding incorrect alternative hypotheses. This approach is proposed basically in the context of clinical trials with two arms (control and treatment) and multiple endpoints. For clinical trials with multiple arms, for example a clinical trial with a control group, a low-dose treatment group and a high-dose treatment group, we can apply this method for pairwise comparisons of treatment effect.

User-friendly software is developed to facilitate pharmaceutical use of Kosorok et al's method. The software is coded with Microsoft Visual C++. We suggest a thorough reading of Kosorok et al [1] for better understanding of the algorithms before using this program. Only minor modifications are made to make the algorithm computationally more efficient. This program has been tested under standard Microsoft Windows and Unix systems with C++. The source code, example data files and program manual can be downloaded from:

`http://www.biostat.wisc.edu/~kosorok/`

We apply our program to the design and analysis of the Copernicus trial. The Copernicus trial is a large-scale, prospective, double-blinded, placebo-controlled trial designed to evaluate the effect of the beta-blocker carvedilol on the survival of patients with severe heart failure. We compare our analysis results with those in [2].

The overall strategy and the modifications we made are discussed in the section Overall strategy. We demonstrate how to use our software to design and analyze a sample clinical trial with two primary endpoints in the section Design and analysis of clinical trials with the software. Possible technical problems and potential modifications we may need for trials with more than two primary endpoints are discussed in the section Possible technical problems. The paper concludes with an analysis of the Copernicus trial in the last section.

Overall strategy

The method proposed by Kosorok et al can be applied to clinical trials with virtually any number of primary endpoints. However, considering the prevalence of clinical trials with one or two primary endpoints and the exponentially increased complexity of calculations with more than two endpoints, we only provide the program for the design and interim analyses of clinical trials with one or two endpoints. Only minor modifications are needed to make the software applicable to trials with more endpoints.

We consider two types of decision rules. Consider a clinical trial with one primary endpoint, and let μ_0 denote the true treatment effect. For a clinical trial with normally distributed response, μ_0 is the mean of the response. For simplicity, we assume a positive effect is beneficial. The null hypothesis is $H_0 : \mu_0 = 0$ and the beneficial and detrimental alternatives are $H_+ : \mu_0 \geq \mu_1$ and $H_- : \mu_0 \leq -\mu_2$, respectively, for some specified positive μ_1 and μ_2. The vague alternatives are defined as $H_{(+)} : \mu_0 \geq 0$ for non-detrimental effect, $H_{(-)} : \mu_0 \leq 0$ for non-beneficial effect and $H_* : \mu_0 \in \mathrm{R}$, if the treatment effect is completely unknown. We can now define joint null

and alternative hypotheses for all endpoints simultaneously. A joint hard alternative has no vague components. For example, the joint hard alternative H_{+-} has a beneficial effect in the first endpoint and a detrimental effect on the second endpoint. Any joint alternative with one or more vague component is called a joint soft alternative. For example, $H_{0(-)}$ indicates that the treatment has no effect in the first endpoint and a non-beneficial effect in the second endpoint. Soft alternatives give us flexibility to terminate the trial earlier, when, for example, there is clearly no treatment effect for the primary endpoint and a detrimental effect for the secondary endpoint.

The global type I error is the probability of drawing a conclusion which is not consistent with the global null hypothesis, when the global null is true. The type II error for a given joint alternative hypothesis is the probability of drawing a conclusion which is not consistent with the true alternative, where the true alternative is distinct from the null hypothesis.

Two types of responses are considered: normally distributed responses and time-to-event responses (survival type). Consider a clinical trial with J interim analyses and one primary endpoint. We assume the response is normally distributed. Let $X_{jl},\ j = 1, ..., J,\ l = 1, ..., n_j$ be independent observations, where n_j is the number of observations at the j^{th} interim analysis. Then the standardized test statistic we use at the j^{th} look is $Z_j = \sum_{i=1}^{j} \sum_{l=1}^{n_i} X_{il} / \sqrt{\sum_{i=1}^{j} n_i}$. Then the distribution of Z_j is completely specified. In the execution stage, the joint distribution of Z_j can be estimated from the data. For the survival type data, we use the logrank test statistic as the response. The asymptotic normality and independent increment structure of the logrank test statistic are used. In the design stage, once we get the non-centrality parameter at each look and interim analysis information time, the joint distributions of the logrank test statistics are completely specified. In the analysis stage, the covariance structure of the logrank statistics can be estimated as in Lin [3].

The overall strategy for both design and analysis phases is to construct marginal critical boundaries satisfying certain α and β spending functions and then to adjust these boundaries so that global type I and type II errors can be controlled. For a detailed review of standard α and β

spending methods, see Lan and DeMets [4] and Chang et al [5].

Design of clinical trials

Consider a clinical trial with two primary endpoints and J interim analyses. Denote by μ_{k1} and $-\mu_{k2}$ the minimal clinically meaningful beneficial and detrimental effects for endpoint k, where $k = 1, 2$ and all μ are known and positive. Two types of critical boundaries can be built: symmetric boundaries and asymmetric boundaries. For a detailed discussion of asymmetric boundaries, see DeMets and Ware [6]. For the current version of our software, we consider clinical trials with symmetric boundaries only.

There exist two types of decision rules: symmetric decision rules and non-symmetric decision rules. We now consider building critical boundaries for clinical trials with symmetric decision rules, where we only accept decisions with no vague components. Consider endpoint k at the j^{th} interim analysis. For the symmetric boundaries, let U_{jk}, L_{jk} and Z_{jk} denote the marginal upper critical boundary, lower critical boundary and test statistic at the j^{th} interim analysis for the k^{th} endpoint, respectively. Denote marginal α spending and β spending, for endpoint k at the j^{th} interim analysis, by α_{jk} and β_{jk}. The marginal stopping region S_{jk} is $(-\infty, -U_{jk}]$, $[-L_{jk}, L_{jk}]$ and $[U_{jk}, +\infty)$, where we conclude H_-, H_0 and H_+, respectively. The rest is the continuation region C_{jk}. We compute U_{jk} by $Pr[|Z_{jk}| > U_{jk}; Z_{ik} \in C_{ik}, i = 1...j-1|H_0] = \alpha_{jk}$ and L_{jk} by $Pr[|Z_{jk}| < L_{jk}; Z_{ik} \in C_{ik}, i = 1...j-1|H_+] = \beta_{jk}$ If $L_{jk} < 0$, then L_{jk} is set equal to zero; if $L_{jk} > U_{jk}$, then L_{jk} is set to be U_{jk}. We adjust the sample size so that the marginal critical boundaries close exactly at the last look. The maximum of the sample sizes calculated marginally from all endpoints is taken as the initial guess of the sample size. In the design phase, we assume sample sizes in the control and treatment groups are equal. Denote by D_{jk} marginal decision regions at the j^{th} interim analysis for the endpoint k. We then construct the multivariate boundaries, $D_{j1} \times D_{j2}$. If the target global type I error rate is not met, a constant c is multiplied by the

multivariate boundaries. We adjust the constant c until the global target type I error is satisfied. Since the adjustment of the sample size affects the global beta spending only, once this constant c is found, it is kept fixed. The sample size is then increased (or decreased) accordingly to ensure that the worst type II error for all alternatives is equal to the targeted global type II error. Plots of bivariate critical regions at the j^{th} look and final look for symmetric decision rules are shown in Figure 1.

We can also have clinical trials with non-symmetric decision rules, where we may accept decisions with possibly vague components. In this case, critical boundaries can be built using similar techniques. For a detailed discussion, see Kosorok et al [1]. Plots of bivariate critical regions at the j^{th} look and final look for a non-symmetric decision rule are shown in Figure 2.

Analysis of clinical trials

Suppose we initiate a clinical trial with sample size N and a pre-specified global α spending function. At each interim analysis, the standardized statistics for the two endpoints and a covariance estimator of the joint covariance of these statistics and all statistics from previous interim analyses are computed. For each endpoint, the noncentrality parameter based on target effect size is also estimated. The marginal critical boundaries at the j^{th} interim analysis are computed to satisfy certain prespecified marginal spending functions α_k and β_k, based on the estimated covariance and non-centrality parameters, and previous marginal boundaries. The same adjustments on boundaries are made as discussed for the design phase, if necessary. For the j^{th} interim analysis, we multiply the critical boundaries with a constant c_j, so that the global alpha spending function is satisfied at this current interim analysis. Comparing the test statistics at each interim analysis with the critical boundaries, we can decide to continue or terminate the trial.

In the design phase, we may only want to control type II error for alternatives we are interested in. For example, consider a clinical trial with two endpoints and a symmetric decision rule. We

may wish to control type II error for all hard alternatives except for the inconsistent hypotheses H_{+-} and H_{-+}.

Design and analysis of clinical trials with the software

In this section, we give an example of how to design and analyze a clinical trial with two primary endpoints. Consider a hypothetical EXAMPLE clinical trial with two endpoints. The first endpoint has a normally distributed response, while the second endpoint has a time-to-event response. This design is inspired by clinical trials like COMPANION, which is a multicenter clinical study evaluating treatment of chronic heart failure. The primary endpoints of interests for the COMPANION trial are exercise performance and quality of life (normal type response) and all cause mortality (time-to-event type data). Details of the COMPANION trial can be found from Guidant Corporation's webpage

`http://www.guidant.com/news/0/web_release/nr_000056.shtml`

and Bristow et al [7].

We plan 5 interim analyses for this EXAMPLE trial. Our objective is to design a two-sided group sequential study with equally spaced information and global type I error 0.05 and type II error 0.1. We assume the response for the primary endpoint is distributed as $N(0,1)$ under H_0 and $N(0.33,1)$ under H_1. We assume the first endpoint has the following α and β spending functions: $\alpha(t) = 0.05(1-\exp(4t))/(1-\exp(4))$ and $\beta(t) = 0.1(1-\exp(4t))/(1-\exp(4))$, respectively, where t is the information spending. We assume that the true event time for the secondary endpoint has an exponential distribution with hazard rate $\lambda_0 = 0.1733$ and $\lambda_1 = 0.1118$ (hazard rates measured in months), under H_0 and H_1 respectively. The planned recruiting time is $T_0 = 28$ months with a $\tau = 6$ months follow-up time. α and β spending functions for the second endpoint are $\alpha(t) = 0.05t$ and $\beta(t) = 0.1t$, respectively. Detailed information of the design can be found in Table 1.

Design of EXAMPLE trial

We illustrate how to input information to compute the minimal sample size and corresponding critical boundaries with our software. In the design phase, we assume the two endpoints are independent. Under this assumption, both the sample size and the global alpha spending function tend to be conservative. Since the α spending at the first look (interim analysis) is generally small, to avoid the possible loss of accuracy due to extreme values in simulations, the critical boundaries for the first look are calculated exactly by Algorithm AS 241 [8]. We assume equal sample sizes in the control and treatment groups. User input is enscribed by a box (□).

Example 1: Design of EXAMPLE Trial

```
Page 1:
Design of Group Sequential Clinical Trials with Multiple Endpoints
Version 1.0
Section I: Calculations of Sample Size and Critical Boundaries
CopyRight(R) Shuangge Ma and Michael Kosorok 2002
Department of Biostatistics and Medical Informatics
University of Wisconsin-Madison

Page 2:
Section I -- Design of Group Sequential Clinical Trials
Minimal Sample Size and Corresponding Critical Boundaries Calculation:
Core Steps:
  1. Sample size and Critical boundary calculation for each endpoint.
  2. Multiple decision rules are used to control overall alpha and beta
     spending.
  3. Key assumptions: independent endpoints; equal sample sizes in
     control and treatment groups.

Page 3:
Step 1 -- Calculation of each endpoint.
  How many endpoints do you have (1 or 2) ?
  Please input the number of endpoints:  2

Page 4:
Step 1 -- Calculation of each endpoint
  Number of endpoints: 2
  ----------------------
  Case Selection:
    Case 1. Two Normal Endpoints.
    Case 2. Two Survival Endpoints.
```

```
    Case 3. A Normal Endpoint(1st) and a Survival Endpoint(2nd).
    Case 4. A Survival Endpoint(1st) and a Normal Endpoint(2nd).
  Please Select (1, 2, 3 or 4):  3

Page 5:
Step 1 -- Calculation of each endpoint
  Number of endpoints: 2
  ----------------------
  What's the decision rule, hard or soft?
  Choose either "h" or "s": h

Page 6:
Select the hard alternatives you want to control for type II error:
(1) H-0, (2) H+0, (3) H0+, (4) H0-, (5) H-+, (6) H++, (7) H--, (8) H+-
Please input here: 1 2 3 4 6 7

Page 7:
Step 1 -- Calculation of each endpoint
Case 3: Two Endpoints: 1st Normal & 2nd Survival
  1) For the first normal endpoint:
  The number of looks this trial has:  5
  Number of Monto carlo replications:  50000
  Under Null Hypothesis:               Mean = 0.00  Variance = 1.00
  Under Althernative Hypothesis:       Mean = 0.33  Variance = 1.00
  The alpha for this endpoint is:      0.05
  The beta for this endpoint is:       0.10
  The spending number is:              1

Page 8:
Step 1 -- Calculation of each endpoint
Case 3: Two Endpoints: 1st Normal & 2nd Survival
  1) For the first normal endpoint:
  --------------------------------
  Information Spending:
  relative information(1):    1      information(1):   0.2
  relative information(2):    2      information(2):   0.4
  relative information(3):    3      information(3):   0.6
  relative information(4):    4      information(4):   0.8
  relative information(5):    4      information(5):   1.0

Page 9:
Step 1 -- Calculation of each endpoint
Case 3: Two Endpoints: 1st Normal & 2nd Survival
  2) For the second survival endpoint:
  The number of looks this trial has:              5
  The number of Monto Carlo replications:          50000
  The type I error (alpha) for the endpoint is:    0.05
  The type II error (beta) for the endpoint is:    0.1
```

```
Page 10:
Step 1 -- Calculation of each endpoint
Case 3: Two Endpoints: 1st Normal & 2nd Survival
  2) For the second survival endpoint:
  ------------------------------------
  Information Spending:
  relative information(1):    1
  relative information(2):    2
  relative information(3):    3
  relative information(4):    4
  relative information(5):    5

Page 11:
Step 1 -- Calculation of each endpoint
Case 3: Two Endpoints: 1st Normal & 2nd Survival
  2) For the second survival endpoint:
  The alphaSpend for this endpoint is:      2
  The betaSpend for this endpoint is:       2
  The lambdaNull for this endpoint is:      0.1733
  The lambdaAlter for this endpoint is:     0.1118
  The entry for this trial is:              28
  The followup for this trial is:           6
  The lambdadrop for this trial is:         0

Page 12:
Step 1 -- Calculation of each endpoint
Case 3: Two Endpoints: 1st Normal & 2nd Survival
  The correlation is:                       0
  The global alpha is:                      0.05
  The global beta is:                       0.1
  -------------------------------------------
  Click "Calculate".
```

In our software, α and β spending functions are denoted with integers. For example, α spending function 1 refers to the α spending function $\alpha(t) = \alpha(1-\exp(4t))/(1-\exp(4))$. For a detailed list of spending functions, see the software webpage. The calculated sample sizes and corresponding critical boundaries for hard (soft) decisions are also shown in Table 1. The plots of critical boundaries with information spending can be visualized from the interface. The boundary plot for symmetric decision rule is shown in Figure 3. For non-symmetric decision rule, the plot looks quite similar, with only numerical differences. In our software, we only show the positive half of the boundary.

Interim analysis of EXAMPLE trial

Experimental data from clinical trials stored in a certain format can be analyzed with our software. As in the design stage, we can analyze data from clinical trials with one (Normal or time-to-event type) or two endpoints (Normal-Normal, Normal-survival, survival-Normal, survival-survival). We demonstrate how to use our software to analyze data with the EXAMPLE data. We make the same assumptions as in the design phase.

Example 2: Analysis of EXAMPLE Trial

```
Page 1:
Data Analysis of
Group Sequential Clinical Trials with Multiple Endpoints
Windows Version 1.0
CopyRight(R) Shuangge Ma and Michael Kosorok 2002
Department of Biostatistics and Medical Informatics
University of Wisconsin -- Madison

Page 2:
Section II -- Data Analysis of Group Sequential Clinical Trials
Step 1: Calculation of Marginal Critical Boundaries Satisfying Certain
        Alpha(beta) Spending Functions
Step 2: Use Global Alpha Spending Function to Control Overall Type I
        Error.
Step 3: Compare Test Statistics with Critical Boundaries
Software Homepage: http://www.biostat.wisc.edu/ kosorok/

Page 3:
Section II -- Data Analysis of Group Sequential Clinical Trials
How many endpoints does this trial have (1 or 2)?
Please input the number of endpoints: 2

Page 4:
Section II -- Data Analysis of Group Sequential Clinical Trials
Number of endpoints: 2
----------------------
Case Selection ?
  Case 1. Two Normal Endpoints.
  Case 2. Two Survival Endpoints.
  Case 3. A Normal Endpoint(first) and a Survival Endpoint(second).
  Case 4. A Survival Endpoint(first) and a Normal Endpoint(second).
Please select (1, 2, 3 or 4):         3
```

```
Page 5:
Section II -- Data Analysis of Group Sequential Clinical Trials
Number of endpoints: 2
-----------------------
Choose decision rule: hard or soft ?
Choose "h" or "s":                    h

Page 6:
Section II -- Data Analysis of Group Sequential Clinical Trials
Case 3: Two Endpoints: first Normal & second Survival
For the first Normal Endpoint:
------------------------------
Please Load Information Files. Two Files Needed.
Sample Information Files Available at:
http://www.biostat.wisc.edu/ kosorok/
Please input the name of the first information files:    new1a.txt

Page 7:
Section II -- Data Analysis of Group Sequential Clinical Trials
Case 3: Two Endpoints: first Normal & second Survival
The following information received.
-----------------------------------
Order =               1                       Alpha =          0.05
Number of Looks =     5                       Beta =           0.1
Number of Monto Carlo replications = 20000    Global Alpha =   0.05
Spending Function =   1                       Global Beta =    0.1
Sample size control = 120                     mean (Null) =    0
Sample size treat =   150                     mean (Alter) =   0.15

Page 8:
Section II -- Data Analysis of Group Sequential Clinical Trials
Case 3: Two Endpoints: first Normal & second Survival
For the first Normal Endpoint:
------------------------------
Please input the name of the second file:               new1b.txt

Page 9:
Section II -- Data Analysis of Group Sequential Clinical Trials
Case 3: Two Endpoints: first Normal & second Survival
-----------------------------------------------------
            Information      Alpha            Beta
Look 1:          1           0.0011           0.0022
Look 2:          2           0.0025           0.0050
Look 3:          3           0.0057           0.0114
Look 4:          4           0.0126           0.0252
Look 5:          5           0.0281           0.0562

Page 10:
```

```
Section II -- Data Analysis of Group Sequential Clinical Trials
Case 3: Two Endpoints: first Normal & second Survival
For the first Normal Endpoint:
------------------------------
Current Look Number:
Please input here: 3

Page 11:
Section II -- Data Analysis of Group Sequential Clinical Trials
Case 3: Two Endpoints: first Normal & second Survival
For the first Normal Endpoint:
------------------------------
Please Load Data File.
Please input the file name here:    newtest.txt

Page 12:
Section II -- Data Analysis of Group Sequential Clinical Trials
Case 3: Two Endpoints: first Normal & second Survival
For the first Normal Endpoint:
------------------------------
Data loaded successfully.

Page 13:
Section II -- Data Analysis of Group Sequential Clinical Trials
Case 3: Two Endpoints: first Normal & second Survival
For the second Survival Endpoint:
---------------------------------
Please Load Information Files. Two Files Needed.
Sample Information Files Available at:
http://www.biostat.wisc.edu/ kosorok/
Please input the name of the first information files: survival2a.txt

Page 14:
Section II -- Data Analysis of Group Sequential Clinical Trials
Case 3: Two Endpoints: first Normal & second Survival
For the second Survival Endpoint:
---------------------------------
The following information received.
Order =               2                       Alpha =          0.05
Number of Looks =     5                       Beta =           0.1
Number of Monto Carlo replications = 20000    Global Alpha =   0.05
Lambda (Null) =       0.1                     Global Beta =    0.2
Lambda (Alter) =      0.164
Lambda (drop) =       0.005
Sample size control = 120
Sample size treat =   150

Page 15:
```

```
Section II -- Data Analysis of Group Sequential Clinical Trials
Case 3: Two Endpoints: first Normal & second Survival
For the second Survival Endpoint:
---------------------------------
Please input the name of the second information file:   survival2b.txt

Page 16:
Section II -- Data Analysis of Group Sequential Clinical Trials
Case 3: Two Endpoints: first Normal & second Survival
For the second Survival Endpoint:
---------------------------------
               Global Alpha    Alpha       Beta    Look time
  Look 1:          0.02        0.01        0.02        7
  Look 2:          0.02        0.01        0.02        14
  Look 3:          0.02        0.01        0.02        21
  Look 4:          0.02        0.01        0.02        28
  Look 5:          0.02        0.01        0.02        34

Page 17:
Section II -- Data Analysis of Group Sequential Clinical Trials
Case 3: Two Endpoints: first Normal & second Survival
For the second Survival Endpoint:
---------------------------------
Current Look Number:
Please input here: 3

Page 18:
Section II -- Data Analysis of Group Sequential Clinical Trials
Case 3: Two Endpoints: first Normal & second Survival
For the second Survival Endpoint:
---------------------------------
Please Load Data File.
Please input the file name here:      send333.txt

Page 19:
Section II -- Data Analysis of Group Sequential Clinical Trials
Case 3: Two Endpoints: first Normal & second Survival
For the second Survival Endpoint:
---------------------------------
Data loaded successfully.

Click "Next" to run the program.

Page 20:
Section II -- Data Analysis of Group Sequential Clinical Trials
Case 3: Two Endpoints: first Normal & second Survival
Critical Boundaries: First endpoint              Second endpoint
                 ------------------------    ------------------------
```

```
                lower    upper     testZ    lower    upper    testZ
Look 1:         0.0027   1.6502   -0.7219 | 0.0176   1.3175   1.1636
Look 2:         0.0122   1.6936   -0.6651 | 0.0731   1.4424   0.3465
Look 3:         0.1176   2.0106    0.2249 | 0.3259   1.7833   0.4958
```

The information spending, global α spending, marginal critical boundaries for each interim analysis and test statistics are shown in Table 2. The numbers given in Table 2 correspond to absolute values of critical boundaries. The plots of the marginal critical boundaries and test statistics for the 3rd look are shown in Figure 4 (symmetric decision rule) and Figure 5 (non-symmetric decision rule). Comparing the test statistics and critical boundaries, we conclude that we should continue this trial.

Possible technical problems

The software has been tested under standard Unix environment in the University of Wisconsin-Madison biostatistics computing systems. Minor modifications of the data loading function may be needed for other Unix systems. Only files in the correct format can be analyzed with our software.

Analysis of Copernicus trial

From previous clinical trials, it is well known that Beta-blocking agents reduce the risk of hospitalization and death in patients with mild-to-moderate heart failure. The primary aim of the Copernicus trial was to find out their effects in severe heart failure. The Copernicus trial was a large-scale, prospective, randomized, double-blind, placebo-controlled trial of the effect of the beta-blocker carvedilol on the survival of patients with severe heart failure. The primary outcomes of interests are:

- All cause mortality.
- The earliest time of all cause mortality and hospitalization.

The trial results were first reported in [2].

Design of Copernicus trial

In this section, we re-examine the design of the Copernicus trial with our program. The sample size is estimated on the basis of the following assumptions: the one-year mortality in the placebo group will be 28 percent; the risk of death will be decreased by 20 percent as a result of treatment with carvedilol. Previous study shows it is reasonable to assume the event times are exponentially distributed with hazard rates $\lambda_{10} = 0.0009$ (all hazard rates measured in days) for the placebo group and $\lambda_{11} = 0.00072$ for the treatment group. Based on previous clinical trial results, the earliest time to all cause mortality or hospitalization is assumed to be exponentially distributed with hazard rates $\lambda_{20} = 0.002$ and $\lambda_{21} = 0.0012$ for the control and treatment groups respectively. The planned enrollment time is 900 days followed by 85 days followup. We assume no drop out for the design stage. Only simple modifications are needed if we take dropout into consideration. A uniformly distributed enrollment is also assumed. As discussed previously, the two endpoints are assumed to be independent. Eight interim analyses are planned with equally spaced information spending for the first endpoint. The information spending for the second endpoint is estimated under the above assumptions. The first endpoint is tested marginally at the two-sided cumulative 0.05 level, with the target type II error 0.1. For the second endpoint, the marginal α and β are both 0.05. For both endpoints, we assume α and β spending functions take the form $\alpha(t) = \alpha t$ and $\beta(t) = \beta t$, respectively. We adjust the sample size and critical regions so that our global type I error is 0.05 and the worst case type II error is also 0.05.

The minimal sample size needed and corresponding critical boundaries for symmetric and non-symmetric decisions are shown in Table 3. Information spending for the second endpoint is also estimated and shown in Table 3. It can be seen that we will need at least 1221 patients in each group for the symmetric decision rules and 1051 patients in each group if we can accept the non-symmetric decision rules, which means a 14 percent decrease in minimal required sample size. With symmetric decision rules, the expected number of death is 820 and the expected number of death

plus hospitalization is 1291.

Interim analysis of Copernicus data

The trial evaluated 2289 patients who had symptoms of heart failure at rest or on minimal exertion, who were clinically euvolemic, and who had an ejection fraction of less than 25 percent. In a double blinded fashion, 1133 patients were randomly assigned to the placebo group and 1156 patients to the treatment with carvedilol. Randomization began on October 28, 1997 and was stopped early on March 20, 2000 on the recommendation of the data and safety monitoring board. This recommendation was based on the finding of a significant beneficial effect of carvedilol. The mean duration of follow-up is 10.4 months. During this time, no patients were lost to follow-up with regard to mortality, as discussed in [2].

We apply our program to the analysis of this trial. There were four interim analyses at day 379, 496, 671 and 875, if we take October 28, 1997 as day one. Information is taken to be proportional to the total number of events happened by the time of interim analysis for each endpoint. As shown in Table 4, marginal information spending for the first endpoint is 0.060, 0.107, 0.215 and 0.390. Marginal information spending for the second endpoint is 0.120, 0.220, 0.413 and 0.720. Global information spending is taken as the information spending for the first endpoint. Details of α and β spending can also be found in Table 4. We use our software to analyze the Copernicus data. The critical boundaries are shown in Table 4. Comparing the critical boundaries with the test statistics, we can draw the conclusion that we can terminate the trial at the fourth interim analysis with the conclusion of significantly beneficial effects for both endpoints. The statistical conclusions are the same as in [2]. The conclusion of significantly beneficial effects is supported by estimates from the data directly. Under an exponential assumption, the hazard rate estimates are $\hat{\lambda}_{10} = 0.00054$ and $\hat{\lambda}_{11} = 0.00035$ for the control and treatment group, respectively, of the first endpoint. For the second primary endpoint, we have $\hat{\lambda}_{20} = 0.002$ and $\hat{\lambda}_{21} = 0.0015$. So we have shown that in

Copernicus trial case, our method is at least as effective as the classical methods used in [2].

Acknowledgments

The first two authors were partially supported by grant CA75142 from the National Cancer Institute. The authors wish to thank Dr. Ellen B. Roecker for help with the Copernicus data analysis.

References

[1] Kosorok, M.R., Shi Y.J., & DeMets, D.L. Design and analysis of group sequential clinical trials with multiple primary endpoints. Biometrics, In press.

[2] Milton Packer, Andrew J.S. Coats, Michael B. Fowler, et al. Effect of carvedilol on survival in severe chronic heart failure. The New England Journal of Medicine 2001; 344: 1651-1658.

[3] Lin D.Y. Nonparametric sequential testing in clinical trials with incomplete multivariate observations. Biometrika 1991, 78 123-31.

[4] Lan K.K.G. & DeMets, D.L. Discrete sequential boundaries for clinical trials. Biometrika 1983, 70, 659-663.

[5] Chang, S.G., Hwang, I.K. & Shih, W.J. Group sequential designs using both type I and type II error probability spending function. Communication in statistics, Part A-Theory and Methods, 27, 1323-1339.

[6] DeMets, D.L. & Ware, J.H. Asymmetric group sequential boundaries for monitoring clinical trials. Biometrika 1982, 69 661-663.

[7] Bristow M.R., Feldman, A.M. & Saxon, L.A. "Heart failure management using implantable devices for ventricular resynchronization: Comparison of medical therapy, pacing, and defibrillation in chronic heart failure (COMPANION) trial," Journal of Cardiac Failure 2000:6:3; 276-285.

[8] Wichura, M.J. Algorithm AS 241: The percentage points of the Normal distribution. Applied Statistics 1988; 37: 477-484.

Table 1. *Sample Size Calculation and Critical Boundaries for Design of EXAMPLE Trial*

Stage (j)	1	2	3	4	5
Information Spending	0.2	0.4	0.6	0.8	1.0
Endpoint 1					
α Spending	0.0011	0.0025	0.0057	0.0126	0.0281
β Spending	0.0022	0.0050	0.0114	0.0252	0.0562
Endpoint 2					
α Spending	0.01	0.01	0.01	0.01	0.01
β Spending	0.02	0.02	0.02	0.02	0.02
Symmetric Decision Rule: Sample Size 114					
Endpoint 1					
Upper Bound	3.3273	3.1435	2.9664	2.5784	2.1928
Lower Bound	0.0094	0.0686	0.3423	1.2184	2.1928
Endpoint 2					
Upper Bound	2.5376	2.5376	2.5376	2.4324	2.3218
Lower Bound	0.1425	0.3804	0.9498	1.5973	2.3218
Nonsymmetric Decision Rule: Sample Size 108					
Endpoint 1					
Upper Bound	3.3972	3.2096	3.0287	2.6326	2.2389
Lower Bound	0.0096	0.0700	0.3495	1.2440	2.2389
Endpoint 2					
Upper Bound	2.5910	2.5910	2.5910	2.5495	2.3793
Lower Bound	0.1616	0.4039	1.0025	1.6596	2.3793

Table 2. *Critical Boundaries for EXAMPLE Data at the Third Interim Analysis*

Stage (j)	1	2	3	4	5
Information Spending	0.2	0.4	0.6		
Global α Spending	0.02	0.02	0.02		
Symmetric Decision Rule					
Endpoint 1					
Upper Bound	1.6502	1.6936	2.0106		
Lower Bound	0.0027	0.0122	0.1176		
Endpoint 2					
Upper Bound	1.3175	1.4424	1.7833		
Lower Bound	0.0176	0.0731	0.3259		
Nonsymmetric Decision Rule					
Endpoint 1					
Upper Bound	2.5562	2.4799	2.5064		
Lower Bound	0.0041	0.0179	0.1467		
Endpoint 2					
Upper Bound	2.0408	2.1120	2.2230		
Lower Bound	0.0272	0.1071	0.4063		
Test Statistics					
Endpoint 1	-0.7219	-0.6651	0.2249		
Endpoint 2	1.1636	0.3465	0.4958		

Table 3. *Sample Size Calculation and Critical Boundaries for Design of Copernicus Trial*

Stage (j)	1	2	3	4	5	6	7	8
Endpoint 1								
Information	0.125	0.25	0.375	0.5	0.625	0.75	0.875	1
α Spending	0.00625	0.00625	0.00625	0.00625	0.00625	0.00625	0.00625	0.00625
β Spending	0.0125	0.0125	0.0125	0.0125	0.0125	0.0125	0.0125	0.0125
Endpoint 2								
Information	0.125	0.252	0.373	0.489	0.609	0.703	0.805	1
α Spending	0.0065	0.00635	0.00605	0.0058	0.006	0.0047	0.0051	0.00975
β Spending	0.00625	0.00635	0.00605	0.0058	0.006	0.0047	0.051	0.00975
Symmetric Decision Rule: Sample Size 1221								
Endpoint 1								
Upper Bound	2.7010	2.7010	2.7010	2.7010	2.6073	2.5799	2.4176	2.4176
Lower Bound	0.0395	0.0891	0.2529	0.5616	1.1058	1.5321	1.9140	2.4176
Endpoint 2								
Upper Bound	2.7010	2.7010	2.7010	2.6065	2.6065	2.4503	2.4422	2.2254
Lower Bound	0.0259	0.0486	0.1783	0.5338	0.9550	1.3663	1.7486	2.2254
Nonsymmetric Decision Rule: Sample Size 1051								
Endpoint 1								
Upper Bound	2.8465	2.8465	2.8465	2.8126	2.8115	2.7259	2.5594	2.3568
Lower Bound	0.0398	0.0829	0.2074	0.5544	1.0080	1.4436	1.8729	2.3568
Endpoint 2								
Upper Bound	2.8465	2.8294	2.8294	2.8294	2.7681	2.5375	2.5375	2.2787
Lower Bound	0.0282	0.0590	0.2255	0.5983	1.0567	1.4864	1.9355	2.2787

Table 4. *Critical Boundaries at Each Interim Analysis for Copernicus Data*

Stage (j)	1	2	3	4	5	6	7	8
Interim Analysis Date	379	496	671	875				
Endpoint 1								
Information Spending	0.060	0.107	0.215	0.390				
α Spending	0.003	0.00235	0.0054	0.00875				
β Spending	0.006	0.0047	0.0108	0.0175				
Endpoint 2								
Information Spending	0.120	0.220	0.413	0.720				
α Spending	0.006	0.005	0.00965	0.01535				
β Spending	0.006	0.005	0.00965	0.01535				
Global α Spending	0.003	0.00235	0.0054	0.00875				
Symmetric Decision Rule								
Endpoint 1								
Upper Bound	2.0954	2.8418	2.7311	2.6663				
Lower Bound	0.0081	0.0059	0.0357	0.3533				
Endpoint 2								
Upper Bound	1.9400	2.5610	2.4074	2.4775				
Lower Bound	0.0749	0.9861	2.4074	2.4775				
Test Statistics								
Endpoint 1	0.9200	1.4492	2.4287	3.4481				
Endpoint 2	1.0165	1.4913	2.8287	2.9410				

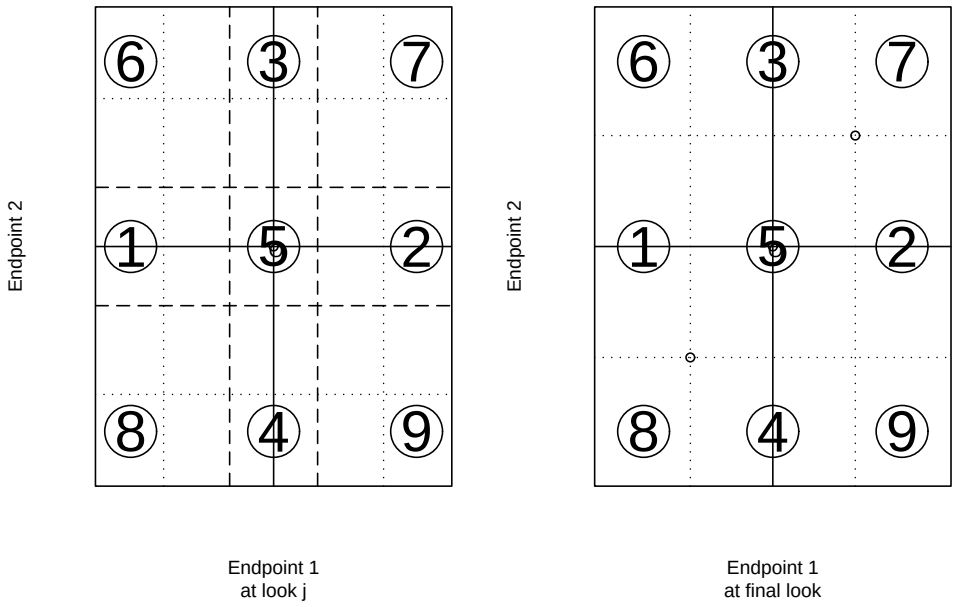

Figure 1: Bivariate critical regions at the j^{th} look and final look for symmetric decision rule. Regions 1–9 are stopping regions which conclude H_{-0}, H_{+0}, H_{0+}, H_{0-}, H_{00}, H_{-+}, H_{++}, H_{--} and H_{+-} respectively. Blank areas are continuation regions. Dashed lines correspond to lower boundaries and dotted lines correspond to upper boundaries.

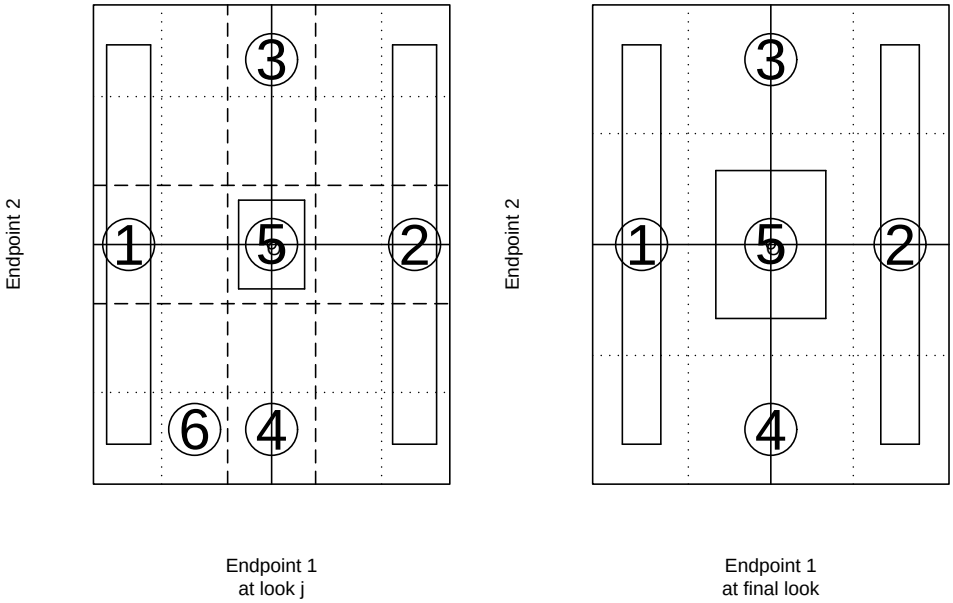

Figure 2: Bivariate critical regions at the j^{th} look and final look for non-symmetric decision rule. Regions 1–6 are stopping regions which conclude H_{-*}, H_{+*}, H_{0+}, H_{0-}, H_{00} and $H_{(-)-}$ respectively. Blank areas are continuation regions. Dashed lines correspond to lower boundaries and dotted lines correspond to upper boundaries.

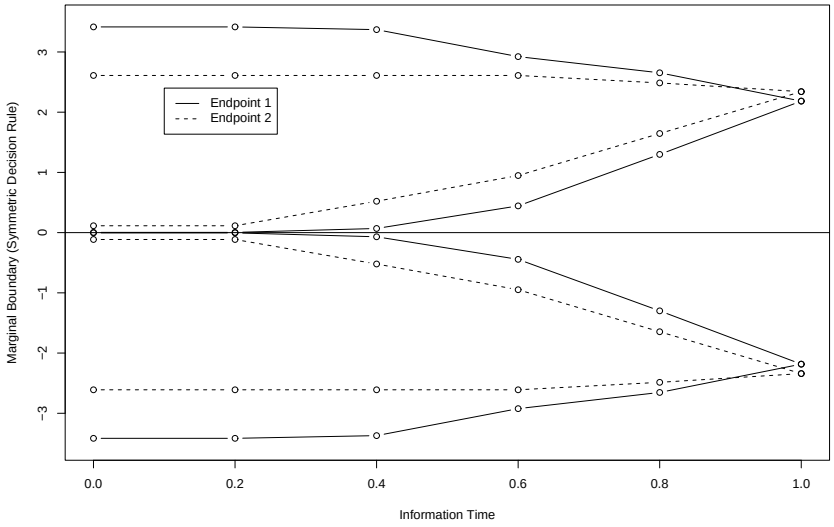

Figure 3: Marginal Critical Boundaries for Design of EXAMPLE Trial (Symmetric Decision Rule).

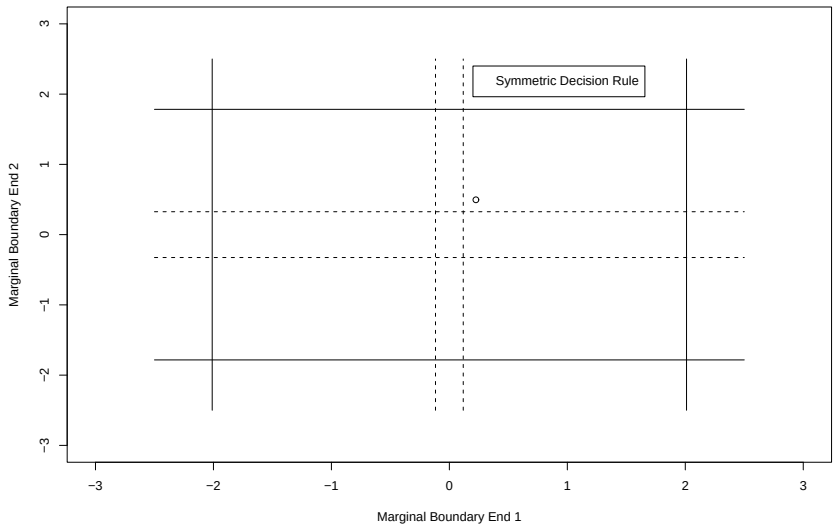

Figure 4: Marginal Critical Boundaries at the 3rd look for Data Analysis of EXAMPLE Trial (Symmetric Decision Rule). Dashed lines are lower boundaries and solid lines are upper boundaries.

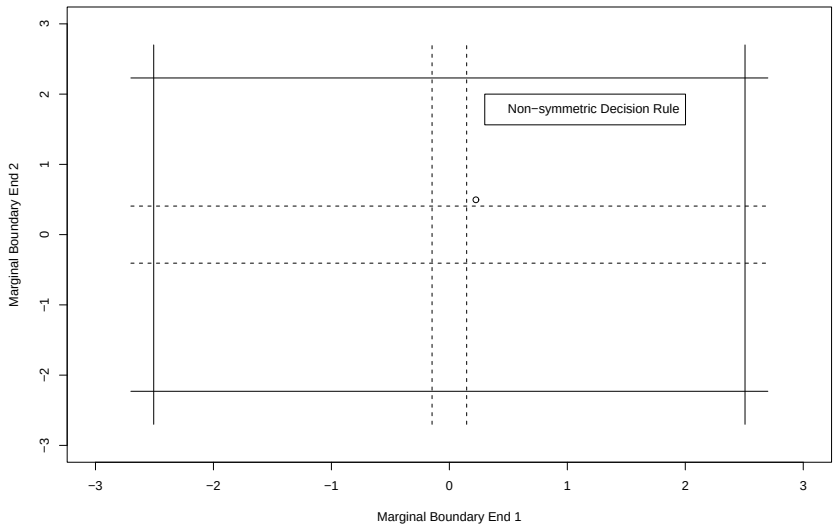

Figure 5: Marginal Critical Boundaries at the 3rd look for Data Analysis of EXAMPLE Trial (Non-symmetric Decision Rule). Dashed lines are lower boundaries and solid lines are upper boundaries.