**University of Alberta**

# Fully sequential monitoring of longitudinal trials using sequential ranks, with applications to an orthodontics study

by

**Paul Joseph Bogowicz**

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of

Master of Science

in

Statistics

Department of Mathematical and Statistical Sciences


Edmonton, Alberta
Fall 2009

# Examining Committee

Edit Gombay, Mathematical and Statistical Sciences

Giseon Heo, Dentistry

Mike MacGregor, Computing Science

Byron Schmuland, Mathematical and Statistical Sciences

This thesis is dedicated to my parents, for their unconditional love and support.

# Abstract

This thesis explores the application of fully sequential methods for the analysis of longitudinal clinical trial data. A new nonparametric approach will be developed, using sequential ranks, for the comparison of several treatment groups. Sequential ranking is an alternative to ranking by the usual method. Although sequential ranks are more likely to suffer from information loss than regular ranks, they are preferred here for their independence.

We will develop three alternative monitoring procedures. The first two will be large-sample, continuous analogues of the Pocock and O'Brien-Fleming group sequential monitoring procedures. The third procedure, a small sample version, will make use of the sign function, and will be grounded in the theory of simple random walks.

The performance of the three monitoring procedures will be assessed via a Monte Carlo simulation study. In particular, we will compare power and average stopping time for various treatment differences, different numbers of treatment groups, and different response distributions. The procedure will then be applied to data arising from an orthodontic clinical trial.

# Acknowledgements

My grateful thanks go to my supervisor, Dr. Edit Gombay, without whom this thesis would not have been possible. She has been delightful to work with, given her vast experience with and extensive knowledge of sequential methods, and of statistics in general. Indeed, her previous contributions to the field form the basis of this thesis.

I also owe many thanks to my co-supervisor Dr. Giseon Heo. She has inspired and encouraged me to be ambitious in my research and in my studies. I am so grateful to her for first introducing me to the field of sequential analysis, and to the ideas of evidence-based research. Dr. Heo continues to offer me her support, advice, and insight.

Finally, I would like to thank Dr. Carlos Flores-Mir, for his wisdom and advice; Dr. Manuel Lagravère, a student in the Orthodontic Graduate Program, Department of Dentistry, for kindly providing the data set that was used in this thesis; Ms. Jennifer Gamble, for sharing this journey with me, and for never tiring of my questions.

Paul

# Contents

# List of Tables

# List of Figures

# Chapter 1

# Introduction

## 1.1 Longitudinal data analysis

In what follows, we consider the historical progression of sequential methods for longitudinal data. Thus it is appropriate to consider first the basic approaches to longitudinal data analysis.

Longitudinal, or panel, data involves the measurement of some quantity on each unit, repeatedly over time. Such data arise frequently in medical and dental applications. Prior to any discussion of analytical methods, it is essential that notation be established.

**Definition 1.1.1** *Let $y_{i\alpha}$ be the $i^{th}$ measurement on the $\alpha^{th}$ unit, for $i = 1, \ldots, p_\alpha$ and $\alpha = 1, \ldots, n$. Each measurement on each unit has a corresponding vector of explanatory variables, $\mathbf{x}_{i\alpha}$, where $\mathbf{x}_{i\alpha}$ is of length $m$.*

To ease the notation, we set $\mathbf{y}_\alpha = (y_{1\alpha}, y_{2\alpha}, \ldots, y_{p_\alpha\alpha})^T$. Finally, $\mathbf{y} = (\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n)^T$ is a $(p_1 + \ldots + p_n) \times 1$ vector, and $\mathbf{X} = (\mathbf{x}_{11}^T, \ldots, \mathbf{x}_{p_1 1}^T, \mathbf{x}_{12}^T, \ldots, \mathbf{x}_{p_2 2}^T, \ldots, \mathbf{x}_{p_n n}^T)^T$ is a $(p_1 + \ldots + p_n) \times m$ matrix.

The methods of longitudinal data explained here follow in general the theory and ideas presented in the book, *Analysis of Longitudinal Data* by Diggle **et al.** [7], unless otherwise indicated.

### 1.1.1 Linear and linear-mixed models

Longitudinal linear and linear-mixed models rely upon the theory of the general linear model (GLM). The basic form of this model is given by

$$\mathbf{y} \quad = \quad \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{1.1}$$

where the indexing of $\boldsymbol{\epsilon}$ is analogous to that of $\mathbf{y}$. Assume that $\boldsymbol{\epsilon}$ is a realization of some random vector whose distribution is $N(\mathbf{0}, \mathbf{V})$, given $\mathbf{X}$. To obtain the usual classical GLM, one would set $\mathbf{V}$ to equal the identity matrix premultiplied by a constant: $\sigma^2\mathbf{I}$. For the special case of longitudinal data, the matrix $\mathbf{V}$ is block diagonal. We will denote these blocks by $\mathbf{V}_\alpha$, for $\alpha = 1, \ldots, n$.

Before elaborating upon the structure of $\mathbf{V}$, we define the general sources of variation that arise in longitudinal data. Measurement error is the variation arising from some measurement process. Variation due to random effects arises when units are randomly sampled from some population. Finally, serial correlation is variation arising from within-unit correlation, when units are monitored over time.

Consider the following additive decomposition

$$\epsilon_{i\alpha} \quad = \quad Z_{i\alpha} + \mathbf{d}_{i\alpha}^T\mathbf{U}_\alpha + W_\alpha(t_{i\alpha}) \tag{1.2}$$

where we define $\mathbf{d}_{i\alpha}$ to be an $r$-vector of unit-specific explanatory variables and $t_{i\alpha}$ to be the measurement time for the $i^{\text{th}}$ measurement on the $\alpha^{\text{th}}$ unit. We let the $Z_{i\alpha}$ form a set of mutually independent $N(0, \tau^2)$ random variables, and let the $\mathbf{U}_\alpha$ form a set of mutually independent length-$r$ $N(\mathbf{0}, \mathbf{G})$ random vectors. Finally, the $W_\alpha(t_{i\alpha})$ are sampled from $n$ independent copies of a zero-mean stationary Gaussian process, with variance $\sigma^2$ and correlation function $\rho(u)$ (Diggle **et al.**, [7]). In this decomposition, the $Z_{i\alpha}$, $\mathbf{U}_\alpha$, and $W_\alpha(t_{i\alpha})$ terms correspond to measurement error, random effects, and serial correlation, respectively.

In order to describe the $\mathbf{V}_\alpha$ we develop a matrix formulation of the above decomposition. Define $\boldsymbol{\epsilon}_\alpha = (\epsilon_{1\alpha}, \epsilon_{2\alpha}, \ldots, \epsilon_{p_\alpha\alpha})^T$. We let $\mathbf{D}_\alpha$ be the $p_\alpha \times r$

matrix with rows given by $\mathbf{d}_{i\alpha}^T$ (where $\alpha$ is fixed). Let $\mathbf{H}_\alpha$ be the $p_\alpha \times p_\alpha$ matrix given by $(h_{jk}) = \rho(|t_{j\alpha} - t_{k\alpha}|)$. Then,

$$\mathbf{V}_\alpha = VAR[\boldsymbol{\epsilon}_\alpha] = \tau^2\mathbf{I} + \mathbf{D}_\alpha\mathbf{G}\mathbf{D}_\alpha^T + \sigma^2\mathbf{H}_\alpha \tag{1.3}$$

Diggle **et al.** [7] give specific examples of the above covariance structure. Here are two of them:

**Example 1.1.2 (Serial correlation alone)** *Suppose that* $\epsilon_{i\alpha} = W_\alpha(t_{i\alpha})$. *Then,* $\mathbf{V}_\alpha$ *simplifies to* $\sigma^2\mathbf{H}_\alpha$. *Typically* $\rho(u)$ *is chosen to decrease as* $u$ *(time separation, or lag) increases. For instance, the exponential model uses* $\rho(u) = exp(-\phi u)$.

**Example 1.1.3 (Measurement error with random effects)** *If* $\epsilon_{i\alpha} = Z_{i\alpha} + \mathbf{d}_{i\alpha}^T\mathbf{U}_\alpha$, *then* $\mathbf{V}_\alpha = \tau^2\mathbf{I} + \mathbf{D}_\alpha\mathbf{G}\mathbf{D}_\alpha^T$. *The special case of* $r = 1$ *yields a well-known growth curve model.*

Now that we have identified the basic framework of the linear and linear-mixed models, we can identify how they are used. In particular we briefly consider three steps: model building, estimation, and inference.

The model building stage first involves exploratory data analysis, including time plots, scatterplot matrices, and empirical variograms. Time plots of the response are useful for ascertaining mean behaviour over time, where time is included as an explanatory variable. Once the model matrix $\mathbf{X}$ is identified, ordinary least-squares (OLS) residuals are calculated and used to identify a suitable covariance model. In particular, time plots, scatterplot matrices, and empirical variograms of the residuals are examined.

The next step is to estimate the parameters from the model identified at the model building stage. In particular, we have implicitly assumed that $\mathbf{y}$ is a realization of the random vector $\mathbf{Y}$ whose distribution is $N(\mathbf{X}\boldsymbol{\beta}, \mathbf{V})$. We may write $\mathbf{V}$ as $\mathbf{V}(\boldsymbol{\gamma})$ to explicitly note its dependence on various parameters (such as $\tau^2$, $\mathbf{G}$, $\sigma^2$). Taking

$$\hat{\boldsymbol{\beta}}(\boldsymbol{\gamma}) \;=\; (\mathbf{X}^T\mathbf{V}(\boldsymbol{\gamma})^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}(\boldsymbol{\gamma})^{-1}\mathbf{y} \tag{1.4}$$

and letting

$$RSS(\boldsymbol{\gamma}) \quad = \quad (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}(\boldsymbol{\gamma}))^T \mathbf{V}(\boldsymbol{\gamma})^{-1} (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}(\boldsymbol{\gamma})) \tag{1.5}$$

be the residual sum of squares (RSS), the restricted maximum likelihood estimator (REML) for $\boldsymbol{\gamma}$ maximizes:

$$L^*(\boldsymbol{\gamma}) \quad = \quad -\frac{1}{2}\{\log|\mathbf{V}(\boldsymbol{\gamma})| + \log|\mathbf{X}^T\mathbf{V}(\boldsymbol{\gamma})^{-1}\mathbf{X}| + RSS(\boldsymbol{\gamma})\} \tag{1.6}$$

Then, the REML estimate for $\boldsymbol{\beta}$ is $\hat{\boldsymbol{\beta}}(\tilde{\boldsymbol{\gamma}})$, where $\tilde{\boldsymbol{\gamma}}$ is the REML estimate for $\boldsymbol{\gamma}$.

Inference for this model is based on the fact that $\hat{\beta}(\boldsymbol{\gamma})$ is distributed as $N(\boldsymbol{\beta}, (\mathbf{X}^T\mathbf{V}(\boldsymbol{\gamma})^{-1}\mathbf{X})^{-1})$. This holds approximately if $\mathbf{V}(\boldsymbol{\gamma})$ is estimated using the REML estimates of $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$. General linear hypothesis testing proceeds as usual. Diggle **et al.** [7] also outline a log-likelihood ratio statistic for use when model selection is not obvious.

### 1.1.2 Marginal models

Marginal models are a natural extension of linear models for longitudinal data. There are two general components to a marginal model: a regression model of the response on explanatory variables, and a model of within-unit correlation. The theory of marginal models borrows extensively from that of generalized linear models.

As for the linear and linear-mixed models, the regression model for the response is one based on expectation. That is,

$$E[Y_{i\alpha}] \quad = \quad \mu_{i\alpha} \tag{1.7}$$

depends on $\mathbf{x}_{i\alpha}$ only through $h(\mu_{i\alpha}) = \mathbf{x}_{i\alpha}^T\boldsymbol{\beta}$. The function $h(\cdot)$ is called a *link* function. Marginal variance is assumed to be

$$VAR[Y_{i\alpha}] \quad = \quad \nu(\mu_{i\alpha})\phi \tag{1.8}$$

where $\nu(\cdot)$ is a known function, and $\phi$ may or may not be known. Similarly, the correlation between two within-unit observations is

$$CORR[Y_{i\alpha}, Y_{j\alpha}] \;=\; \rho(\mu_{i\alpha}, \mu_{j\alpha}, \boldsymbol{\gamma}) \tag{1.9}$$

where $\rho(\cdot)$ is also a known function, and $\boldsymbol{\gamma}$ is a vector of (potentially unknown) parameters.

If we use the identity link, $h(\mu) = \mu$, and further assume that the data is Gaussian, the class of models specified by the above equations is the same as the class of linear models developed in Section 1.1.1.

In the normal case, a likelihood approach is sufficient for estimation. It leads to well-known equations that are easy to solve. However, the general form of the marginal model does not require a distributional assumption. This precludes the use of likelihood methods for general estimation purposes. Instead, parameters are often estimated by solving the generalized estimating equations (GEE), given by

$$S(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \sum_{\alpha=1}^{n} \left( \frac{\partial \boldsymbol{\mu}_\alpha}{\partial \boldsymbol{\beta}} \right)^T VAR[\mathbf{Y}_\alpha](\boldsymbol{\beta}, \boldsymbol{\gamma})^{-1}(\mathbf{y}_\alpha - \boldsymbol{\mu}_\alpha) = 0 \tag{1.10}$$

where $\boldsymbol{\mu}_\alpha = (\mu_{1\alpha}, \mu_{2\alpha}, \ldots, \mu_{p_\alpha \alpha})^T$. Note that if $\boldsymbol{\gamma}$ is not known it must be estimated separately, prior to solving the GEE. Diggle **et al.** [7] suggest that $\boldsymbol{\gamma}$ should be replaced by an $m^{1/2}$-consistent estimate, $\hat{\boldsymbol{\gamma}}$.

According to Fitzmaurice **et al.** [8], a robust sandwich estimator for the variance matrix of $\hat{\boldsymbol{\beta}}$ should be used for making inferences regarding $\boldsymbol{\beta}$.

### 1.1.3 Repeated measures ANOVA

The analysis of longitudinal data using repeated measures analysis of variance (ANOVA) is an alternative to analyses based on linear, linear-mixed, and marginal models. Its structure parallels that of a split-plot ANOVA; units are plots, and unit-specific repeated measurements form subplots. A randomization argument is not justified however, as measurements are taken in sequential

order. The following description of repeated measures ANOVA follows that of Davis [6].

In the previous sections, we have considered only general longitudinal data. In particular, a single sample of units, each with corresponding repeated measurements on some variable. ANOVA methods can be used to compare two or more such samples. To do so however, we will require an extension of our existing notation:

**Definition 1.1.4** *Let $y_{i\alpha}^{(k)}$ be the $i^{th}$ measurement on the $\alpha^{th}$ unit in the $k^{th}$ group, for $i = 1, \ldots, p_\alpha^{(k)}$, $k = 1, \ldots, c$, and $\alpha = 1, \ldots, n_k$, with corresponding measurement time $t_{i\alpha}^{(k)}$. In vector notation, $\mathbf{y}_\alpha^{(k)} = (y_{1\alpha}^{(k)}, y_{2\alpha}^{(k)}, \ldots, y_{p_\alpha^{(k)}\alpha}^{(k)})^T$.*

This definition allows for flexibility in the number of groups as well as the number of individuals in each group. In what follows, we will fix $p_\alpha^{(k)} = p$ for $k = 1, \ldots, c$ and $\alpha = 1, \ldots, n_k$. In other words, we assume that we have the same number of repeated measurements on each unit.

The formulation of an ANOVA table first requires the specification of an underlying model. There are several possibilities, each of which leads to the same table. As in Davis [6], we choose the simplest one:

$$y_{i\alpha}^{(k)} = \mu + \gamma_k + \tau_i + (\gamma\tau)_{ki} + \pi_{\alpha(k)} + \epsilon_{i\alpha}^{(k)} \tag{1.11}$$

for $i = 1, \ldots, p$, $k = 1, \ldots, c$, and $\alpha = 1, \ldots, n_k$. The overall mean, common to all subjects, is represented by $\mu$. The $\gamma_k$ represent group effects, the $\tau_i$ represent time effects, and the $(\gamma\tau)_{ki}$ the interactions between the two. The effects are subject to the following constraints:

$$\sum_{k=1}^{c} \gamma_k = \sum_{i=1}^{p} \tau_i = \sum_{k=1}^{c} (\gamma\tau)_{ki} = \sum_{i=1}^{p} (\gamma\tau)_{ki} = 0. \tag{1.12}$$

We define the $\pi_{\alpha(k)}$ to be mutually independent random effects for units, and the $\epsilon_{i\alpha}^{(k)}$ to be mutually independent measurement errors.

Assume that the $\pi_{\alpha(k)}$ and $\epsilon_{i\alpha}^{(k)}$ are distributed as $N(0, \nu^2)$ and $N(0, \sigma^2)$, respectively. The ANOVA table is given by

Table 1.1: Sums of squares, degrees of freedom, and mean squares for repeated measures ANOVA.

| Source | SS | df | MS |
|---|---|---|---|
| Group | $SS_G$ | $c-1$ | $SS_G/(c-1)$ |
| Units(Group) | $SS_{U(G)}$ | $n-c$ | $SS_{U(G)}/(n-c)$ |
| Time | $SS_T$ | $p-1$ | $SS_T/(p-1)$ |
| Group $\times$ Time | $SS_{G\times T}$ | $(c-1)\times(p-1)$ | $SS_{G\times T}/[(c-1)\times(p-1)]$ |
| Residual | $SS_R$ | $(n-c)\times(p-1)$ | $SS_R/[(n-c)\times(p-1)]$ |

where

$$n = \sum_{k=1}^{c} n_k \tag{1.13}$$

is the total number of units under observation, and

$$SS_G = p\sum_{k=1}^{c} n_k(\bar{y}_{\cdot\cdot}^{(k)} - \bar{y}_{\cdot\cdot}^{(\cdot)})^2 \tag{1.14}$$

$$SS_{U(G)} = p\sum_{k=1}^{c}\sum_{\alpha=1}^{n_k}(\bar{y}_{\cdot\alpha}^{(k)} - \bar{y}_{\cdot\cdot}^{(k)})^2 \tag{1.15}$$

$$SS_T = n\sum_{i=1}^{p}(\bar{y}_{i\cdot}^{(\cdot)} - \bar{y}_{\cdot\cdot}^{(\cdot)})^2 \tag{1.16}$$

$$SS_{G\times T} = \sum_{k=1}^{c}\sum_{\alpha=1}^{n_k}\sum_{i=1}^{p}(\bar{y}_{i\cdot}^{(k)} - \bar{y}_{\cdot\cdot}^{(k)} - \bar{y}_{i\cdot}^{(\cdot)} + \bar{y}_{\cdot\cdot}^{(\cdot)})^2 \tag{1.17}$$

$$SS_R = \sum_{k=1}^{c}\sum_{\alpha=1}^{n_k}\sum_{i=1}^{p}(y_{i\alpha}^{(k)} - \bar{y}_{i\cdot}^{(k)} - \bar{y}_{\cdot\alpha}^{(k)} + \bar{y}_{\cdot\cdot}^{(k)})^2 \tag{1.18}$$

are the sums of squares. The bar-dot notation is taken to mean the average over the dotted indices. For example,

$$\bar{y}_{\cdot\cdot}^{(\cdot)} = \frac{1}{np}\sum_{k=1}^{c}\sum_{\alpha=1}^{n_k}\sum_{i=1}^{p} y_{i\alpha}^{(k)} \tag{1.19}$$

and

$$\bar{y}_{i\cdot}^{(k)} = \frac{1}{n_k}\sum_{\alpha=1}^{n_k} y_{i\alpha}^{(k)}. \tag{1.20}$$

There are several hypotheses of interest that are eligible for testing under the model specified in equation 1.11. In particular, to test

$$H_{01}: \text{ no difference between groups} \tag{1.21}$$

versus

$$H_{A1}: \text{ not } H_{01} \tag{1.22}$$

we would use

$$F_1 = \frac{MS_G}{MS_{U(G)}} = \frac{SS_G/(c-1)}{SS_{U(G)}/(n-c)} \tag{1.23}$$

which is distributed as $F$ with $c-1$ and $n-c$ degrees of freedom, under the null hypothesis. Similarly, to test

$$H_{02}: \text{ no difference over time} \tag{1.24}$$

versus

$$H_{A2}: \text{ not } H_{02} \tag{1.25}$$

we would use

$$F_2 = \frac{MS_T}{MS_R} = \frac{SS_T/(p-1)}{SS_R/[(n-c)\times(p-1)]} \tag{1.26}$$

which is distributed as $F$ with $p-1$ and $(n-c)\times(p-1)$ degrees of freedom, under the null hypothesis. Finally, we may also test whether there is an interaction between the effects of group and time, that is

$$H_{03}: \text{ no interaction between group and time} \tag{1.27}$$

versus

$$H_{A3}: \text{ not } H_{03}. \tag{1.28}$$

For this test, we would use

$$F_3 = \frac{MS_{G\times T}}{MS_R} = \frac{SS_{G\times T}/[(c-1)\times(p-1)]}{SS_R/[(n-c)\times(p-1)]} \tag{1.29}$$

which is distributed as $F$ with $(c-1) \times (p-1)$ and $(n-c) \times (p-1)$ degrees of freedom, under the null hypothesis.

Each ANOVA test requires that the within-unit variance matrices are the same, for each group. This is similar to the usual "equal variances" assumption in one-way ANOVA. In addition however, the tests of $H_{02}$ and $H_{03}$ require a so-called sphericity condition to hold. If we ignore groups, reverting to our previous notation, the sphericity condition can be expressed as:

$$VAR[y_{i\alpha} - y_{j\alpha}] \text{ is constant for all } i \text{ and } j \tag{1.30}$$

where $i, j = 1, \ldots, p$. This can be assessed using Mauchly's test for sphericity [6].

### 1.1.4 Other models and methods

Diggle **et al.** [7] outline two other general approaches that can be used to analyze longitudinal data. We describe both very briefly.

Random effects models are a natural extension of both generalized linear models, and the linear-mixed random effects error models. The latter are extended to incorporate link functions other than the identity link. These models are useful when it is the case that there is heterogeneity among units in some or all regression coefficients. In particular, they should be used when individuals are the target of inference, rather than the mean. As is the case with generalized linear models, inference here is likelihood-based.

Transition models assume that correlation arises between successive within-unit measurements because past values of the underlying random process influence current and future values. The distribution of $Y_{i\alpha}$ is specified as being conditional upon its past values $Y_{1\alpha}, \ldots, Y_{(i-1)\alpha}$, as well as the past and current values of covariates under consideration. Likelihood-based estimation and inference is generally used. A useful class of transition models is the class of Markov generalized linear models.

## 1.2 Sequential analysis

One of the characteristics of clinical trial data is that it is not usually available all at once. In other words, data accumulates gradually, over time. For ethical and sometimes economic reasons, researchers are motivated to examine the data as it arrives. The process of doing so may inflate the probability of Type I error. Sequential analysis is an area of statistics that aims to control this inflation, by formalizing the process of interim looks.

Historically, there have been two approaches: group and fully sequential analysis. Group sequential analysis restricts analyses to often pre-specified analysis times, called interim analyses. Fully, or continuous, sequential analysis involves monitoring the data continuously. That is, an analysis is performed after each new data point or observation is received.

Before describing how sequential analysis has been applied to longitudinal data, we must establish some notation, and define some sequential methods.

### 1.2.1 Group sequential analysis

A group sequential monitoring procedure involves examining the data at specific interim analysis times:

**Definition 1.2.1** *Let $T$ be the maximum number of interim analyses. Analyses $1, 2, \ldots, T$ take place at times $t_1, t_2, \ldots, t_T$. We will use the index $j$ to refer to an arbitrary analysis time, $t_j$.*

This should not be confused with $t_{i\alpha}^{(k)}$, which is the measurement time corresponding to the $i^{\text{th}}$ measurement on the subject indexed by $k$ and $\alpha$.

It is often the case that we are repeatedly testing some null hypothesis on a single parameter, with a symmetric alternative. To do so, we calculate some test statistic, say $S$, at each analysis time:

**Definition 1.2.2** *We say that $S_1, S_2, \ldots, S_T$ is a sequence of test statistics, and often place them in a vector $\mathbf{S} = (S_1, S_2, \ldots, S_T)^T$.*

To actually test the null hypothesis at time $t_j$, we compare $S_j$ to a critical value, or boundary:

**Definition 1.2.3** *We define $B_1, B_2, \ldots, B_T$ to be the boundaries used at interim analyses $1, 2, \ldots, T$ respectively.*

Most often a simple rejection rule is used, whereby if $|S_j| > B_j$, the trial stops. If on the other hand $|S_j| \leq B_j$, the trial continues at least until the $(j+1)^{\text{th}}$ interim analysis time.

Assuming an overall error rate of $\alpha$, we define two commonly used monitoring procedures. The first is due to Slud and Wei [39]. Our definition follows that of Spiessens **et al.** [40]:

**Definition 1.2.4 (Slud-Wei method)** *First, exit probabilities $\alpha_1, \alpha_2, \ldots, \alpha_T$ are chosen so that*

$$\sum_{j=1}^{T} \alpha_j \quad = \quad \alpha. \tag{1.31}$$

*Then, the boundaries can be calculated according to*

$$P\{|S_1| > B_1\} \quad = \quad \alpha_1 \tag{1.32}$$

*and*

$$P\{|S_1| \leq B_1, \ldots, |S_{j-1}| \leq B_{j-1}, |S_j| > B_j\} \quad = \quad \alpha_j \tag{1.33}$$

*for $j = 2, \ldots, T$, using numerical integration, under the assumption that the distribution of $\mathbf{S}$ is multivariate normal. We also define marginal significance levels $\alpha'_1, \alpha'_2, \ldots, \alpha'_T$ such that*

$$P\{|S_j| > B_j\} \quad = \quad \alpha'_j. \tag{1.34}$$

Unlike earlier group sequential methods, the Slud-Wei method does not require equally spaced intervals. It does however require $T$ to be pre-specified. The next method is due to Lan and DeMets [26], with the definition following the one given in Spiessens **et al.** [40]:

**Definition 1.2.5 (Lan-DeMets method)** *Let $\alpha^*(t)$ be a continuous, non-decreasing function satisfying $\alpha^*(0) = 0$ and $\alpha^*(1) = \alpha$. We call $\alpha^*(t)$ an $\alpha$-spending function, where $t$ is the fraction of total information available. Boundaries are determined according to*

$$P\{|S_1| > B_1\} = \alpha^*(t_1^*) \tag{1.35}$$

*and*

$$P\{|S_1| \leq B_1, \ldots, |S_{j-1}| \leq B_{j-1}, |S_j| > B_j\} = \alpha^*(t_j^*) - \alpha^*(t_{j-1}^*) \tag{1.36}$$

*for $j = 2, \ldots, T$, where $t_1^*, t_2^*, \ldots, t_T^*$ are the information fractions at each analysis time. In practice, boundary calculation requires numerical integration.*

The Lan-DeMets method is in fact quite flexible since it does not require pre-specification of the spacing or number of analyses. Its main disadvantage is that the information fractions must be estimated, which may be difficult if the total information is unknown.

A detailed development of group sequential methods can be found in the book by Jennison and Turnbull [21].

### 1.2.2 Fully sequential analysis

A fully sequential monitoring procedure involves examining the data after each new observation is received:

**Definition 1.2.6** *Analyses $1, 2, \ldots, j, \ldots$ take place at times $t_1, t_2, \ldots, t_j, \ldots$, corresponding in theory to the measurement times of the $1^{\text{st}}, 2^{\text{nd}}, \ldots, j^{\text{th}}, \ldots$ observations, respectively.*

A pre-specified maximal number of analyses is not required. Rather, the early philosophy of fully sequential methods involved sampling until the null hypothesis was accepted or rejected.

A description and historical account of early fully sequential methods can be found in the book by Ghosh and Sen [12].

### 1.2.3 Group versus fully sequential analysis

In this section we briefly examine the group and fully sequential dichotomy from the perspective of relative advantages and disadvantages. We follow the comparison given in Bogowicz **et al.** [2].

Group sequential methods are simple and flexible [38]. Trials employing these methods are less biased and shortened less often than fully sequential trials [45]. Moreover, group sequential techniques are preferred to fully sequential methods when the estimation of treatment effects is important [45]. On the other hand, fully sequential methods are lauded for having smaller expected sample sizes [31]. They are often of shorter duration, exposing fewer patients to inferior treatments [31].

Both have relative disadvantages however. Group sequential analyses may delay the potential for early stopping [42]. Unplanned interim analyses may cause interpretation problems. Indeed, treatment differences may be exaggerated [38]. In contrast, because of the frequency of data analysis, fully sequential analysis has been rarely applied in medicine [38]. Moreover, adjustments at the final analysis are more aggressive for fully than for group sequential trials [31].

# Chapter 2

# Sequential methods for longitudinal data

## 2.1 Group sequential methods

The longitudinal group sequential literature can be classified according to methodological approach. In particular, researchers have based their theories upon linear and linear-mixed models, marginal models and the GEE, and nonparametric methods. In addition, there are some articles intended for general applicability, encompassing two or more approaches. Finally, there are articles with no particular restrictions on underlying models or methods.

In what follows, we will use both the simple notation introduced with the linear and linear-mixed models, as well as the extended notation from the section on repeated measures ANOVA.

### 2.1.1 Linear and linear-mixed models

The application of group sequential methods to longitudinal data was first considered by Armitage **et al.** [1]. They develop a method for a simple linear-mixed model, with autoregressive errors. Using the linear-mixed model notation,

$$\mathbf{y} = \boldsymbol{\epsilon} \tag{2.1}$$

where they set

$$\epsilon_{i\alpha}^{(k)} \;=\; (1-\phi)u_{\alpha}^{(k)} + \phi\epsilon_{(i-1)\alpha}^{(k)} + \omega_{i\alpha}^{(k)}. \tag{2.2}$$

The distributions of the $u_{\alpha}^{(k)}$ and $\omega_{i\alpha}^{(k)}$ are $N(\mu, \sigma_0^2)$ and $N(0, \sigma^2(1-\phi^2))$, respectively. The $u_{\alpha}^{(k)}$ and $\omega_{i\alpha}^{(k)}$ are assumed to be mutually independent. The model matrix $\mathbf{X}$ is empty, since there are no covariates under consideration.

Armitage **et al.** [1] make a number of restrictive assumptions on the structure of the data: fixed and non-staggered entry, fixed follow-up times, analyses at equally spaced intervals. Their test statistics are taken to be differences in cumulative sums, between two treatment groups:

$$S_j \;=\; \sum_{\alpha=1}^{n_1}\left(\sum_{i=1}^{j} y_{i\alpha}^{(1)}\right) - \sum_{\alpha=1}^{n_2}\left(\sum_{i=1}^{j} y_{i\alpha}^{(2)}\right). \tag{2.3}$$

The authors show that the effect of sequentially testing correlated data at equally spaced intervals is the same as that of testing uncorrelated data at unequally spaced intervals. Hence theory for the latter is applied, with the same nominal significance level used at each analysis.

The approach of Armitage **et al.** [1] was extended by Geary [11]. The model given in equation (2.1) is extended by using errors given by

$$\epsilon_{i\alpha}^{(k)} = \begin{cases} u_{\alpha}^{(k)} + w_{i\alpha}^{(k)} & , i = 1 \\ u_{\alpha}^{(k)} + \phi\epsilon_{(i-1)\alpha}^{(k)} + \omega_{i\alpha}^{(k)} & , i > 1 \end{cases} \tag{2.4}$$

where the distributions of the $u_{\alpha}^{(k)}$, $\omega_{1\alpha}^{(k)}$, and $\omega_{i\alpha}^{(k)}$ (for $i > 1$) are $N(\mu, \sigma_0^2)$, $N(0, \sigma^2/(1-\kappa\phi^2))$, and $N(0, \sigma^2)$, respectively. The $u_{\alpha}^{(k)}$ and $\omega_{i\alpha}^{(k)}$ are again assumed to be mutually independent.

Geary [11] placed similar constraints to Armitage **et al.** [1] on data structure. Test statistics are also similar, again based upon differences of cumulative sums. Boundaries are calculated according to a multivariate normal numerical integration procedure. The procedure requires the *a priori* specification of the number of interim analyses, as well as the corresponding exit probabilities.

The ideas of both Armitage **et al.** [1] and Geary [11] were extended by Lee and DeMets [27], in a slightly more general linear-mixed model framework. The model is

$$\mathbf{y}_\alpha^{(k)} = \mathbf{X}_\alpha^{(k)}\boldsymbol{\beta}^{(k)} + \boldsymbol{\epsilon}_\alpha^{(k)} \tag{2.5}$$

where

$$\boldsymbol{\epsilon}_\alpha^{(k)} = \mathbf{D}_\alpha^{(k)}\mathbf{U}_\alpha^{(k)} + \boldsymbol{\omega}_\alpha^{(k)}. \tag{2.6}$$

Here $\mathbf{D}_\alpha^{(k)}$ is a $p_\alpha^{(k)} \times r$ design matrix, $\mathbf{U}_\alpha^{(k)}$ is distributed as $N(\mathbf{0}, \mathbf{G})$, and $\boldsymbol{\omega}_\alpha^{(k)}$ is distributed as $N(\mathbf{0}, \mathbf{R}_\alpha^{(k)})$. The matrix $\mathbf{R}_\alpha^{(k)}$ depends on $k$ and $\alpha$ only through its dimensionality, $p_\alpha^{(k)} \times p_\alpha^{(k)}$. The matrix $\mathbf{X}_\alpha^{(k)}$ and vector $\boldsymbol{\beta}^{(k)}$ will be specified below. This model is similar to, but different from the original linear-mixed model formulation from section 1.1.1, in that $\boldsymbol{\omega}_\alpha^{(k)}$ is accounting for the additional variation and possible correlation introduced by measurement error and serial correlation components, respectively.

The approach of Lee and DeMets [27] allows for staggered entry, unequally spaced measurement times, and some degree of missing data. Their procedure requires the matrix $\mathbf{X}_\alpha^{(k)}$ to be

$$\left[\begin{array}{cc} \mathbf{1} & \mathbf{t}_\alpha^{(k)} \end{array}\right] \tag{2.7}$$

where $\mathbf{t}_\alpha^{(k)}$ is the vector of measurement times corresponding to the measurements on the unit indexed by $k$ and $\alpha$. Corresponding to this form of $\mathbf{X}_\alpha^{(k)}$ is the vector $\boldsymbol{\beta}^{(k)} = (\beta_1^{(k)}, \beta_2^{(k)})^T$. The test statistic at analysis $j$, for comparing two treatment groups, is a normalized version of

$$\hat{\beta}_2^{(1)}(t_j) - \hat{\beta}_2^{(2)}(t_j). \tag{2.8}$$

The authors show that the joint distribution of their sequence of test statistics is multivariate normal. They indicate that both the Slud-Wei and Lan-DeMets methods can be used for boundary calculation.

Wu and Lan [47] define a linear model similar to that of Lee and DeMets [27]. They omit, however, the component corresponding to the fixed effects

$\boldsymbol{\beta}^{(k)}$. They develop instead a group sequential method based upon the expected response curve, essentially just the expected value, $E[\mathbf{Y}_\alpha^{(k)}]$. A Lan-DeMets $\alpha$-spending function should be used, along with multivariate normal numerical integration, to calculate boundaries. The procedure allows for staggered entry, unequally spaced measurement times, missing data, and informative censoring. Although the method is used to compare only two treatment groups, the authors suggest that three or more groups could be compared using isotonic regression.

The 1997 article by Jennison and Turnbull [20] provides a unified look at the distribution of sequences of estimators. In particular, they consider maximum likelihood estimators derived from normal linear models, generalized linear models, and the proportional hazards regression model. In each case, the distribution of sequences of estimators is multivariate normal, with a specific covariance structure. The theory is exact for normal linear models and asymptotic for generalized linear models.

Finally, the 2002 article by Cerutti **et al.** [4] extends the applicability of the linear-mixed model approach to situations in which a comparison between three or more treatment groups is the objective. The model is a special case of the one used by Lee and DeMets [27], defined here in equations (2.5) and (2.6). In particular, they take $\mathbf{X}_\alpha^{(k)}$ to be

$$\left[ \begin{array}{cc} \mathbf{1} & \mathbf{t}_\alpha^{(k)} \end{array} \right] \tag{2.9}$$

with $\boldsymbol{\beta}^{(k)} = (\beta_1^{(k)}, \beta_2^{(k)})^T$. The matrix $\mathbf{R}_\alpha^{(k)}$ is set to equal $\tau^2\mathbf{I}$. In other words, the $\boldsymbol{\omega}_\alpha^{(k)}$ represent measurement error.

The hypotheses of interest are

$$H_{0k}: \ \beta_2^{(k)} = \beta_2^{(1)} \tag{2.10}$$

versus

$$H_{Ak}: \ \beta_2^{(k)} \neq \beta_2^{(1)} \tag{2.11}$$

for $k = 2, \ldots, c$, so that the overall hypothesis to be tested is

$$H_0 : \bigcap_{k=2}^{c} H_{0k} \tag{2.12}$$

versus

$$H_A : \text{ not } H_0. \tag{2.13}$$

The test statistics corresponding to each $H_{0k}$ at time $t_j$ are standardized versions of

$$\hat{\beta}_2^{(k)}(t_j) - \hat{\beta}_2^{(1)}(t_j) \tag{2.14}$$

as in Lee and DeMets [27]. Testing proceeds according to one of the following:

1. Test each hypothesis according to boundaries depending on the set $\mathcal{H} = \{H_{0k} \mid k = 2, \ldots, c\}$. Drop inferior treatments and remove the corresponding rejected hypotheses from $\mathcal{H}$. Update the value of the error spending function. Continue testing until all $H_{0k}$ are rejected or time runs out.

2. Test according to (1), stopping as soon as the first $H_{0k}$ is rejected. Apply a treatment comparison method using some pairwise procedure.

The error spending function used in (1) and (2) is completely specified by Cerutti **et al.** [4].

### 2.1.2 Marginal models and the GEE

The usage of GEE-based estimators in longitudinal group sequential trials was first proposed by Wei **et al.** [46]. They assume a marginal model for $Y_{i\alpha}^{(k)}$ and further specify that its distribution is from the exponential family. The GEE are used to estimate the model parameters, $\boldsymbol{\beta}$. The procedure solves

$$\sum_{k=1}^{2} \sum_{\alpha=1}^{n_k} \left( \frac{\partial \boldsymbol{\mu}_\alpha^{(k)}}{\partial \boldsymbol{\beta}} \right)^T (\mathbf{y}_\alpha^{(k)} - \boldsymbol{\mu}_\alpha^{(k)}) \;\; = \;\; 0 \tag{2.15}$$

where the working covariance matrix is taken to be the identity, **I**. Although this is equivalent to the strong assumption that within-subject observations are uncorrelated, estimates of $\boldsymbol{\beta}$ are still consistent [46]. The procedure allows for staggered entry, unequally spaced measurement times, and of course, non-normal response distributions.

The test statistic is the standardized estimate of the $\beta$ coefficient corresponding to the indicator of treatment group. Boundaries are calculated using multivariate normal numerical integration, while the exit probabilities are determined using the Slud-Wei method. The authors hesitate to apply the Lan-DeMets method because of difficulties in estimating total information.

In 1996, Gange and DeMets [10] showed that the Lan-DeMets method can in fact be applied in a GEE-based longitudinal setting. They specify a marginal model, relaxing the assumption of uncorrelated within-subject observations of Wei **et al.** [46]. Moment estimators are used to estimate nuisance parameters, while the GEE are solved to find $\hat{\boldsymbol{\beta}}$, where $\boldsymbol{\beta} = (\boldsymbol{\beta}^*, \theta)^T$ and $\theta$ is the parameter of interest. Testing at time $t_j$ is based on a Wald-type statistic:

$$\frac{\hat{\theta}(t_j)}{\sqrt{VAR[\hat{\theta}(t_j)]}}. \tag{2.16}$$

The sequence of $T$ test statistics, each premultiplied by its corresponding information fraction, is shown to have an asymptotic multivariate normal distribution. Numerical integration is used along with the Lan-DeMets method to calculate boundaries. Gange and DeMets [10] note that essentially any monotonic process mapping to $[0, 1]$ is a valid surrogate for the information fraction. However, following the original ideas of the Lan-DeMets method, they suggest a data-based surrogate be used. It is not clear whether the procedure functions, for example, if entry is staggered, or data is missing.

Lee **et al.** [29] use the GEE in a similar fashion to Gange and DeMets [10]. They apply the Lan-DeMets method to sequences of score and Wald test statistics. Both types of statistics are based upon a single parameter of interest. Information at time $t_j$ is estimated as being the inverse of the variance of the

test statistic at $t_j$. This is divided by some test statistic-dependent estimate of maximum information, to obtain the information fraction. The authors note that the procedure can accommodate staggered entry, variable numbers of repeated measurements, and unequally spaced measurement times.

### 2.1.3 Nonparametric methods

Lee and DeMets [28] were first to propose the application of nonparametric group sequential methods in a longitudinal setting. Their procedure, which is rank-based, is designed for the comparison of two treatment groups, under a location-shift model. It allows for staggered entry, unequally spaced measurement times, and missing data. The authors accommodate unequal numbers of measurements by using a regression-like transformation. At the $j^{\text{th}}$ interim analysis, this transformation maps each individual's arbitrary number of measurements into a length-$j$ vector of statistics.

The authors define ranks $R_{j\alpha}$ to be the rank of the $\alpha^{\text{th}}$ subject's $j^{\text{th}}$ transformed measurement, among all other measurements sharing the same $j$ index. Their test statistic is based upon linear rank statistics

$$S_{n(t_j)} = \sum_{\alpha=1}^{n(t_j)} c_\alpha a_{n,j}(R_{j\alpha}) \tag{2.17}$$

for $j = 1, \ldots, T$. The constants $c_\alpha$ indicate treatment group (0 or 1), and the $a_{n,j}(\cdot)$ are score functions. Dependence on time is emphasized by writing $n(t_j)$, indicating that not all subjects are necessarily recruited by time $t_j$. The test statistic is

$$S_j = \frac{S_{n(t_j)}}{\sqrt{n(t_j)}} \tag{2.18}$$

for $j = 1, \ldots, T$. The authors show that the sequence of test statistics has an asymptotic multivariate normal distribution. They note that both the Slud-Wei and Lan-DeMets methods can be used for boundary calculation. They conclude by noting that the procedure could be used to compare three or more groups.

A different nonparametric longitudinal group sequential approach is due to Su and Lachin [41]. Their method, like that of Lee and DeMets [28], is based on a location-shift model

$$F^{(1)}(\mathbf{y}) \;=\; F^{(2)}(\mathbf{y}-\boldsymbol{\Delta}) \tag{2.19}$$

where $F^{(k)}(\cdot)$ is the distribution function for all measurements on a subject from group $k$, and where $\mathbf{y}$ is assumed to be of length $p$ for all subjects. The authors develop a multivariate Hodges-Lehmann estimator for the shift parameter, $\boldsymbol{\Delta}$. The entries of $\hat{\boldsymbol{\Delta}}$ are combined to form some scalar aggregate estimate, which, in standardized form, serves as the test statistic. The sequence of test statistics is shown to be asymptotically multivariate normal. Boundaries are calculated via the Slud-Wei method.

Lachin [24] proposed another rank-based longitudinal group sequential method, with some similarities to the method of Su and Lachin [41]. An estimate of

$$\theta_i \;=\; P\{Y_{i\alpha}^{(1)} \le Y_{i\alpha}^{(2)}\} - P\{Y_{i\alpha}^{(2)} \le Y_{i\alpha}^{(1)}\} \tag{2.20}$$

is obtained from a transformation of the rank statistic, for $i = 1, \ldots, p$, and arbitrary $\alpha$. The vector $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)^T$ is estimated at every interim analysis time. Various test statistics are proposed: a $p$-degrees of freedom chi-square test, a 1-degree of freedom test of association, and a 1-degree of freedom test of stochastic ordering. The actual group sequential procedure is developed using the test of association. The asymptotic distribution of the sequence of test statistics is multivariate normal, as usual. Lachin [24] suggests that the Lan-DeMets method may be preferred over the Slud-Wei method, for boundary calculation.

Lachin **et al.** [25] developed a nonparametric chi-square testing framework for the comparison of two treatment groups. The authors assume that measurements occur at fixed time points, though entry may be staggered. The test is based upon a multivariate Wilcoxon test, using a $p$-degrees of freedom

chi-square test statistic. Numerical integration and simulation are used to calculate boundaries, with exit probabilities specified by either of the Slud-Wei or Lan-DeMets methods.

### 2.1.4 General approaches and other methods

In this section, we briefly discuss additional longitudinal group sequential articles that are intended for general applicability.

Scharfstein **et al.** [37] developed an "information-based" monitoring procedure with a very broad range of application. Providing that the parameter of interest is unique, and can be efficiently estimated, the procedure applies to any type of model in any type of group sequential study. The "information-based" moniker results from basing design considerations upon maximum information. The authors employ normal numerical integration along with the Lan-DeMets method for boundary calculation. They note that their procedure could eventually be extended to enable testing of multiple parameters.

Another information-based monitoring procedure was developed by Scharfstein and Tsiatis [36]. Similarly to Scharfstein **et al.** [37], the procedure applies to any type of model in any type of group sequential study, given that a unique parameter of interest can be efficiently estimated. The authors describe a technique, using the bootstrap, to determine whether or not a midterm trial redesign is warranted. They emphasize that maintaining blinding is possible in any redesign.

Galbraith and Marschner [9] proposed a longitudinal group sequential procedure that forgoes the specification of mean and covariance structures. Instead, the mean and covariance matrix are estimated, via maximum likelihood, irrespective of covariate values. They are used to form Wald-type test statistics. The sequence of test statistics has an approximate joint multivariate normal distribution. This is exploited for boundary calculation using, for instance, the Lan-DeMets method. The authors suggest that the greatest efficiency gains from incorporating group sequential methods in a longitudinal

trial can be had by setting $T = 2$, with the first analysis taking place in the first half of the trial.

A longitudinal group sequential method based on summary statistics was established by Kittelson **et al.** [23]. The method is designed to compare the means of two treatment groups. Inference at time $t_j$ is based on a linear combination of

$$\boldsymbol{\mu}^{(1)}(t_j) - \boldsymbol{\mu}^{(2)}(t_j) \tag{2.21}$$

where the length of the resulting vector may be less than $p$. The $\boldsymbol{\mu}^{(k)}$ are estimated via maximum likelihood. It is not clear exactly what the test statistic is, or if a particular boundaries method should be used. The authors note that their method should be used for trials with fixed measurement times.

Troendle **et al.** [43] proposed a new type of group sequential analysis for longitudinal trials. The method allows for the testing of different null hypotheses at each analysis time. Moreover, different parameters may be tested at different times. The authors assume that treatment effects are non-transient, in the sense that if a treatment difference exists at one analysis time, a difference in favour of the same treatment will exist at subsequent analyses. They adopt a so-called "ordered multiple hypothesis testing framework." The Bonferroni correction is presented as an alternative to more formal error spending methods.

## 2.2 Fully sequential methods

We have thus far reviewed the theoretical papers concerning the application of sequential analysis to longitudinal data. It is our understanding that fully sequential methods have not yet been applied in this particular setting. Although there are some disadvantages to fully sequential procedures, as discussed in section 1.2.3, we believe that these contra are outweighed by the possibility of earlier stopping. Indeed, stopping a trial early minimizes overall

exposure to control or inferior treatments. With this in mind, we set out to ascertain the feasibility of a longitudinal fully sequential procedure.

We began by studying existing longitudinal group sequential methods, to find a suitable basis for a fully sequential procedure. Given the complexity of longitudinal data, we restricted our attention to nonparametric methods. After excluding methods that were deemed to be too complex, we were left with the approach of Lee and DeMets [28]. Their rank-based method, which was described in section 2.1.3, is for the comparison of two treatment groups. We thus had two problems: the task of extending the procedure for the comparison of multiple treatment groups; and the conversion of the procedure from group to fully sequential.

On the recommendations of the article by Lee and DeMets [28], we retrieved the text by Puri and Sen [34] to examine how the multiple treatment group comparison might be done. Although they consider the comparison of multiple groups using ranks, it was not obvious how we might use their theory to form a single test statistic. Moreover, a sequence of such test statistics would possess a complex covariance structure, on account of dependency among ranks. The resulting difficulties in the calculation of critical values effectively render this approach unfeasible in the fully sequential setting.

Abandoning the idea of extending the procedure of Lee and DeMets [28], we decided to focus on another type of sequential testing developed by Gombay [14]. In that article, she develops continuous versions of the Pocock and O'Brien-Fleming group sequential monitoring procedures, applying them to the problem of detecting change within a single sample, using sequential ranks. The two procedures have since been applied in a variety of parametric and nonparametric settings: in Gombay [15–17], and in Gombay and Serban [19].

The following chapters examine the application of the techniques of Gombay [14] to the longitudinal setting. We use sequential ranks along with test statistics in the form of partial sums. Sequential ranks are similar to ranks obtained by the usual method, and will be explained in greater detail in Chap-

ter 3. Their main advantage is that they are independent, while regular ranks are not. One might legitimately wonder why we use partial sum test statistics rather than an average. In fact, using averages would require the monitoring of pairwise differences, yielding the same sort of dependency problems as would have arisen from using regular ranks.

In addition, we develop an exact version of our procedures by applying theory on random walks from Csáki [5], as applied in Gombay [13].

# Chapter 3

# The fully sequential nonparametric procedure

In this section we develop a new nonparametric fully sequential procedure for the analysis of longitudinal clinical trial data. We begin with some basic notation and definitions, and then develop large- and small-sample versions of the procedure.

## 3.1 Preliminaries

Suppose that we have longitudinal data on individuals in $c$ treatment groups. Each group has a maximum sample size of $n_k$, for $k = 1, \ldots, c$. Each subject has a maximum of $p$ repeated, or longitudinal, measurements. We denote these measurements by $y_{i\alpha}^{(k)}$: the $i^{\text{th}}$ measurement on the $\alpha^{\text{th}}$ subject in the $k^{\text{th}}$ group, where $i = 1, \ldots, p$, $k = 1, \ldots, c$, and $\alpha = 1, \ldots, n_k$. We place a single subject's observations in a vector $\mathbf{y}_\alpha^{(k)} = (y_{1\alpha}^{(k)}, y_{2\alpha}^{(k)}, \ldots, y_{p\alpha}^{(k)})^T$.

Assume that $\mathbf{y}_\alpha^{(k)}$ is a randomly sampled realization of $\mathbf{Y}_\alpha^{(k)}$. We define $F^{(k)}$ to be the distribution function of $\mathbf{Y}_\alpha^{(k)}$, for $k = 1, \ldots, c$ and arbitrary $\alpha$. We are interested in testing

$$H_0 : \; F^{(1)} = F^{(2)} = \ldots = F^{(c)} = F \tag{3.1}$$

versus

$$H_A : \text{ there exist } k, l \in 1, \ldots, c \text{ such that } F^{(k)} \neq F^{(l)} \tag{3.2}$$

where $F$ is continuous, but unknown. Under the null hypothesis, for fixed $i$,

$$\{Y_{i\alpha}^{(k)} \mid k = 1, \ldots, c, \alpha = 1, \ldots, n_k\} \tag{3.3}$$

are independent and identically distributed (i.i.d.), with marginal distribution function $F_i^{(k)} = F_i$.

### 3.1.1 Sequential ranks and their properties

As each new observation arrives, its sequential rank may be computed among all $Y_{i\alpha}^{(k)}$ for $k = 1, \ldots, c$, $\alpha = 1, \ldots, n_k$, and fixed $i$, that have already been observed. We define, for fixed $i$,

$$\begin{aligned} R_{i\alpha}^{(k)} &= 1 + \#\{Y_{i\beta}^{(j)} \mid Y_{i\beta}^{(j)} \leq Y_{i\alpha}^{(k)}, Y_{i\beta}^{(j)} \text{ has already} \\ &\quad \text{been observed, not including } Y_{i\alpha}^{(k)} \text{ itself,} \\ &\quad \text{for } j = 1, \ldots, c \text{ and } \beta = 1, \ldots, n_j\} \end{aligned} \tag{3.4}$$

to be the sequential rank for $Y_{i\alpha}^{(k)}$, where $i = 1, \ldots, p$, $k = 1, \ldots, c$, and $\alpha = 1, \ldots, n_k$. Note that the sequential rank for $Y_{i\alpha}^{(k)}$ does not change as additional observations are received.

We next consider some of the basic properties of sequential ranks, first established by Parent [32]. Before doing so, we define the following quantity:

$$\begin{aligned} m_{i\alpha}^{(k)} &= 1 + \#\{Y_{i\beta}^{(j)} \mid Y_{i\beta}^{(j)} \text{ has been observed} \\ &\quad \text{before, but not including, } Y_{i\alpha}^{(k)}; \text{ for} \\ &\quad j = 1, \ldots, c \text{ and } \beta = 1, \ldots, n_j\}. \end{aligned} \tag{3.5}$$

Now suppose we are determining the ranked order of all $m_{i\alpha}^{(k)}$ observations that have been observed up to and including $Y_{i\alpha}^{(k)}$. Under $H_0$ and for fixed $i$, the $Y_{i\alpha}^{(k)}$ are i.i.d., so that any ordered permutation of those $m_{i\alpha}^{(k)}$ observations is equally likely. That is,

$$P\{R_{i\alpha}^{(k)} = j \mid j \leq m_{i\alpha}^{(k)}\} = \frac{1}{m_{i\alpha}^{(k)}}. \tag{3.6}$$

As $R_{i\alpha}^{(k)}$ takes values in $\{1, 2, \ldots, m_{i\alpha}^{(k)}\}$, the above relation defines the probability mass function of a discrete uniform random variable. The mean and variance of $R_{i\alpha}^{(k)}$ are

$$E[R_{i\alpha}^{(k)}] = \frac{m_{i\alpha}^{(k)} + 1}{2} \tag{3.7}$$

and

$$VAR[R_{i\alpha}^{(k)}] = \frac{(m_{i\alpha}^{(k)})^2 - 1}{12} \tag{3.8}$$

respectively. It follows that

$$Z_{i\alpha}^{(k)} = \frac{R_{i\alpha}^{(k)} - \frac{m_{i\alpha}^{(k)}+1}{2}}{\sqrt{\frac{(m_{i\alpha}^{(k)})^2-1}{12}}} \tag{3.9}$$

has a mean of zero and a variance of one.

Next, we will examine why it is that sequential ranks are independent. Consider the joint distribution of any two sequential ranks from those $m_{i\alpha}^{(k)}$ observations that have been observed before $Y_{i\alpha}^{(k)}$. Suppose our two sequential ranks are $R_{i\alpha}^{(k)}$ and $R_{i\beta}^{(l)}$, and without loss of generality assume that $m_{i\beta}^{(l)} < m_{i\alpha}^{(k)}$. Now, because the ranking processes for $R_{i\alpha}^{(k)}$ and $R_{i\beta}^{(l)}$ are done separately, it follows that they take values in $\{1, 2, \ldots, m_{i\alpha}^{(k)}\}$ and $\{1, 2, \ldots, m_{i\beta}^{(l)}\}$, respectively. Thus there are a total of $m_{i\alpha}^{(k)} \times m_{i\beta}^{(l)}$ possibilities. Once again, under $H_0$ all permutations are equally likely, so that

$$P\{R_{i\alpha}^{(k)} = j, R_{i\beta}^{(l)} = h \mid j \leq m_{i\alpha}^{(k)}, h \leq m_{i\beta}^{(l)}\} = \frac{1}{m_{i\alpha}^{(k)} m_{i\beta}^{(l)}}. \tag{3.10}$$

On the other hand, we have that

$$P\{R_{i\alpha}^{(k)} = j \mid j \leq m_{i\alpha}^{(k)}\} P\{R_{i\beta}^{(l)} = h \mid h \leq m_{i\beta}^{(l)}\} = \frac{1}{m_{i\alpha}^{(k)}} \times \frac{1}{m_{i\beta}^{(l)}} \tag{3.11}$$

verifying the independence condition for any two sequential ranks. This can be extended to verify the independence conditions for any three ranks, four

ranks, and so on. Indeed, mathematical induction can be used to prove that the sequential ranks are independent.

Examples of sequential and regular ranking, as well as illustrations of their respective independence and dependence, are given in Appendix A.

## 3.2 Large-sample versions

### 3.2.1 The test statistic

Assume that at time $t$, there are $n(t)$ subjects in the trial on whom all $p$ repeated measures have been observed. This $n(t)$ can be broken down into group totals, $n_k(t)$, for $k = 1, \ldots, c$, where $n(t) = n_1(t) + n_2(t) + \ldots + n_c(t)$. Each subject has an associated vector of repeated measurements, with length equal to $p$.

**Definition 3.2.1** *A subject enrolled in the trial at time $t$ is called* ***complete*** *if all $p$ repeated measurements on that subject have been observed by $t$. Subjects are otherwise said to be* ***incomplete****.*

Note that the actual number of subjects enrolled in the trial at time $t$ may be larger than $n(t)$, as there are potentially subjects on whom we have less than $p$ repeated measures. Data on these incomplete subjects is still used, however, for the calculation of sequential ranks and for the estimation of variance components, used in the standardization of statistics on complete subjects.

The within-subject correlation, inherent to longitudinal data, implies that

$$S_\alpha^{(k)} \;=\; \sum_{i=1}^{p} Z_{i\alpha}^{(k)} \tag{3.12}$$

will have dependent components. On the other hand, the quantities $S_\alpha^{(k)}$ are independent for $k = 1, \ldots, c$ and $\alpha = 1, \ldots, n_k(t)$, by the independence of sequential ranks. It seems very natural, then, to form test statistics based on these quantities. Before doing so, however, we need to standardize them.

Based on the available data at time $t$, we need to estimate

$$\begin{aligned} VAR[S_\alpha^{(k)}] &= VAR[\sum_{i=1}^{p} Z_{i\alpha}^{(k)}] \\ &= \sum_{i=1}^{p}\sum_{j=1}^{p} COV[Z_{i\alpha}^{(k)}, Z_{j\alpha}^{(k)}] \end{aligned} \tag{3.13}$$

to standardize the $S_\alpha^{(k)}$. Simple empirical estimates of $COV[Z_{i\alpha}^{(k)}, Z_{j\alpha}^{(k)}]$, for $i, j = 1, \ldots, p$ are given by

$$q_{i,j} = \frac{1}{n-1}\sum_{k=1}^{c}\sum_{\alpha=1}^{n_k}(Z_{i\alpha}^{(k)} - \bar{Z}_i)(Z_{j\alpha}^{(k)} - \bar{Z}_j) \tag{3.14}$$

for $n = n_1 + n_2 + \ldots + n_c$, and where

$$\bar{Z}_i = \frac{1}{n}\sum_{k=1}^{c}\sum_{\alpha=1}^{n_k} Z_{i\alpha}^{(k)}. \tag{3.15}$$

Replacing $COV[Z_{i\alpha}^{(k)}, Z_{j\alpha}^{(k)}]$ on the right hand side of equation (3.13) with $q_{i,j}$, and denoting the estimator of $VAR[S_\alpha^{(k)}]$ by $\hat{\sigma}(t)^2$, we have that

$$\hat{\sigma}(t)^2 = \sum_{i=1}^{p}\sum_{j=1}^{p} q_{i,j}. \tag{3.16}$$

Note that $n$ from equations (3.14) and (3.15) depends upon $t$ in the sense that $q_{i,j}$ may be calculated at a time when patients are still being recruited. However, writing $n(t)$ would imply that only observations from complete subjects are used, while better estimates could surely be obtained by pooling all available data. This dependence on $t$ is made explicit only on the left hand side of equation (3.16).

The estimator defined in equation (3.16) has a number of interesting properties. First, as

$$E[q_{i,j}] = COV[Z_{i\alpha}^{(k)}, Z_{j\alpha}^{(k)}] \tag{3.17}$$

and as expectation is a linear operator, it follows that

$$\begin{aligned} E[\hat{\sigma}(t)^2] &= E[\sum_{i=1}^{p}\sum_{j=1}^{p} q_{i,j}] \\ &= \sum_{i=1}^{p}\sum_{j=1}^{p} E[q_{i,j}] \\ &= \sum_{i=1}^{p}\sum_{j=1}^{p} COV[Z_{i\alpha}^{(k)}, Z_{j\alpha}^{(k)}] \\ &= \sigma^2. \end{aligned} \tag{3.18}$$

In other words, $\hat{\sigma}(t)^2$ is an unbiased estimator of $\sigma^2 = VAR[S_\alpha^{(k)}]$. It is also the case that $\hat{\sigma}(t)^2$ is consistent. To see this, note that the independence of sequential ranks, for fixed $i$, ensures that $q_{i,j}$ is a consistent estimator of $COV[Z_{i\alpha}^{(k)}, Z_{j\alpha}^{(k)}]$ [22]. The consistency of $\hat{\sigma}(t)^2$ follows since the sum of finitely many consistent estimators is a consistent estimator of the corresponding sum of parameters [3].

Finally, the quantities

$$S^{(k)}(t) = \sum_{\alpha=1}^{n_k(t)} \frac{S_\alpha^{(k)}}{\hat{\sigma}(t)} \tag{3.19}$$

each have $n_k(t)$ standardized and independent components. The number of $S_\alpha^{(k)}$, for $k = 1, \ldots, c$ and $\alpha = 1, \ldots, n_k(t)$, and the value of $\hat{\sigma}(t)$ are to be updated each time the data is analyzed.

### 3.2.2 Monitoring

We consider two different monitoring procedures, continuous versions of the group sequential monitoring procedures due to Pocock [33] and O'Brien and Fleming [30], that were developed by Gombay [14]. The procedures, which we call Test 1 and Test 2, will be explained below. Each test tests the null and alternative hypotheses given by equations (3.1) and (3.2), respectively.

First we consider the special case of $c = 2$, where there are two groups to be compared:

**Test 1 - Pocock**. Stop sampling and reject the null hypothesis if

$$\left| \frac{S^{(1)}(t)}{\sqrt{n_1(t)}} \right| > C(\alpha, n_1) \tag{3.20}$$

where $n_1(t)$ is the number of complete subjects enrolled at time $t$, that received treatment 1. If equation (3.20) does not hold, sampling and testing should continue. The bound $C(\alpha, n_1)$ is a critical value depending upon the overall significance level $\alpha$, and the number of subjects receiving treatment 1, $n_1$. This test is two-sided.

**Test 2 - O'Brien-Fleming**. Stop sampling and reject the null hypothesis if

$$\left| \frac{S^{(1)}(t)}{\sqrt{n_1}} \right| > C(\alpha) \tag{3.21}$$

holds. The number $n_1$ is the fixed total number of subjects receiving treatment 1. The bound $C(\alpha)$ depends upon the overall significance level $\alpha$. Sampling continues if equation (3.21) does not hold. This test is also two-sided.

Notice that we only monitor one of the groups; without loss of generality, group 1. This may be justified as follows. Suppose that the test statistic for group 1 is not significant. This implies that the $Z_{i\alpha}^{(1)}$ have a mean of zero, and further, for fixed $i$, that the $Y_{i\alpha}^{(k)}$ are i.i.d. Hence,

$$P\{Y_{i\alpha}^{(1)} \geq Y_{i\beta}^{(2)}\} = P\{Y_{i\alpha}^{(1)} \leq Y_{i\beta}^{(2)}\} = \frac{1}{2}. \tag{3.22}$$

On the other hand, the above relation implies that the $Z_{i\alpha}^{(2)}$ have a mean of zero, and hence that the test statistic for group 2 is a mean zero process. Here we assume that the distribution of $\mathbf{Y}_{\alpha}^{(k)}$ is symmetric.

Now consider the comparison of three treatment groups, in other words setting $c = 3$:

**Test 1 - Pocock**. Stop sampling and reject the null hypothesis if either

$$\left| \frac{S^{(1)}(t)}{\sqrt{n_1(t)}} \right| > C(\alpha^*, n_1) \tag{3.23}$$

or

$$\left| \frac{S^{(2)}(t)}{\sqrt{n_2(t)}} \right| > C(\alpha^*, n_2) \tag{3.24}$$

where $n_1(t)$ and $n_2(t)$ are the numbers of complete subjects enrolled at time $t$, that received treatments 1 and 2, respectively. If neither equation (3.23) nor equation (3.24) holds, sampling and testing should continue. The bounds $C(\alpha^*, n_1)$ and $C(\alpha^*, n_2)$ are critical values depending upon the overall significance level $\alpha$ through $\alpha^*$, and the number of subjects receiving treatments 1 and 2, $n_1$ and $n_2$, respectively. This test is two-sided.

**Test 2 - O'Brien-Fleming**. Stop sampling and reject the null hypothesis if either

$$\left| \frac{S^{(1)}(t)}{\sqrt{n_1}} \right| > C(\alpha^*) \tag{3.25}$$

or

$$\left| \frac{S^{(2)}(t)}{\sqrt{n_2}} \right| > C(\alpha^*) \tag{3.26}$$

holds. The numbers $n_1$ and $n_2$ are the fixed total numbers of subjects receiving treatments 1 and 2, respectively. The bound $C(\alpha^*)$ depends upon the overall significance level $\alpha$ through $\alpha^*$. Sampling continues if neither equation (3.25) nor equation (3.26) holds. This test is also two-sided.

Note that here we only monitor two out of the three groups. Without loss of generality, we monitor the first and second. We claim that if these two groups yield insignificant test statistics, a Type II error will not be made by not rejecting $H_0$. We justify this as follows. Suppose that the test statistic for group 1 is not significant, and is hence a zero mean process, under $H_0$. Then, assuming that $n_1 = n_2 = n_3$,

$$P\{Y_{i\alpha}^{(1)} \geq Y_{i\beta}^{(l)} \mid m_{i\beta}^{(l)} < m_{i\alpha}^{(1)}\} \;=\; \frac{1}{2}P\{Y_{i\alpha}^{(1)} \geq Y_{i\beta}^{(2)}\} + \frac{1}{2}P\{Y_{i\alpha}^{(1)} \geq Y_{i\beta}^{(3)}\} \tag{3.27}$$

and

$$P\{Y_{i\alpha}^{(1)} \leq Y_{i\beta}^{(l)} \mid m_{i\beta}^{(l)} < m_{i\alpha}^{(1)}\} \;=\; \frac{1}{2}P\{Y_{i\alpha}^{(1)} \leq Y_{i\beta}^{(2)}\} + \frac{1}{2}P\{Y_{i\alpha}^{(1)} \leq Y_{i\beta}^{(3)}\} \tag{3.28}$$

since if $Y_{i\alpha}^{(1)}$ is larger than some other observation $Y_{i\beta}^{(l)}$, it is equally likely that that observation comes from either group 2 or group 3. Then, noting that under $H_0$ the relation from equation (3.22) holds for any two of groups 1, 2, and 3, we have that:

$$\frac{1}{2}P\{Y_{i\alpha}^{(1)} \geq Y_{i\beta}^{(2)}\} + \frac{1}{2}P\{Y_{i\alpha}^{(1)} \geq Y_{i\beta}^{(3)}\} = \frac{1}{2} \tag{3.29}$$

$$\frac{1}{2}P\{Y_{i\alpha}^{(1)} \leq Y_{i\beta}^{(2)}\} + \frac{1}{2}P\{Y_{i\alpha}^{(1)} \leq Y_{i\beta}^{(3)}\} = \frac{1}{2}. \tag{3.30}$$

Similarly, if the test statistic for group 2 is insignificant, then under $H_0$, it is also a mean zero process implying that:

$$\frac{1}{2}P\{Y_{i\alpha}^{(2)} \geq Y_{i\beta}^{(1)}\} + \frac{1}{2}P\{Y_{i\alpha}^{(2)} \geq Y_{i\beta}^{(3)}\} = \frac{1}{2} \tag{3.31}$$

$$\frac{1}{2}P\{Y_{i\alpha}^{(2)} \leq Y_{i\beta}^{(1)}\} + \frac{1}{2}P\{Y_{i\alpha}^{(2)} \leq Y_{i\beta}^{(3)}\} = \frac{1}{2}. \tag{3.32}$$

Adding and rearranging equations (3.30) and (3.32), and equations (3.29) and (3.31), yields

$$\begin{aligned}\frac{1}{2}P\{Y_{i\alpha}^{(1)} \leq Y_{i\beta}^{(3)}\} + \frac{1}{2}P\{Y_{i\alpha}^{(2)} \leq Y_{i\beta}^{(3)}\} \\ = 1 - \frac{1}{2}P\{Y_{i\alpha}^{(1)} \leq Y_{i\beta}^{(2)}\} - \frac{1}{2}P\{Y_{i\alpha}^{(2)} \leq Y_{i\beta}^{(1)}\}\end{aligned} \tag{3.33}$$

and

$$\begin{aligned}\frac{1}{2}P\{Y_{i\alpha}^{(2)} \geq Y_{i\beta}^{(3)}\} + \frac{1}{2}P\{Y_{i\alpha}^{(1)} \geq Y_{i\beta}^{(3)}\} \\ = 1 - \frac{1}{2}P\{Y_{i\alpha}^{(2)} \geq Y_{i\beta}^{(1)}\} - \frac{1}{2}P\{Y_{i\alpha}^{(1)} \geq Y_{i\beta}^{(2)}\}\end{aligned} \tag{3.34}$$

respectively. Noticing that the right-hand-sides of equations (3.33) and (3.34) are equivalent, we have that,

$$\begin{aligned}\frac{1}{2}P\{Y_{i\alpha}^{(1)} \leq Y_{i\beta}^{(3)}\} + \frac{1}{2}P\{Y_{i\alpha}^{(2)} \leq Y_{i\beta}^{(3)}\} \\ = \frac{1}{2}P\{Y_{i\alpha}^{(2)} \geq Y_{i\beta}^{(3)}\} + \frac{1}{2}P\{Y_{i\alpha}^{(1)} \geq Y_{i\beta}^{(3)}\}\end{aligned} \tag{3.35}$$

which implies that the test statistic for group 3 is also a mean zero process.

Note that this proof could be modified for unequal group sizes, relaxing the assumption that $n_1 = n_2 = n_3$. It can also be shown that if $c$ populations are to be compared, then it is sufficient to monitor $c-1$ of them, via mathematical induction. In light of this, and the fact that the test statistics for each group are independent partial sum processes, we have that

$$(1-\alpha) \;=\; (1-\alpha^*)^{c-1} \tag{3.36}$$

where $\alpha$ is the overall error rate, and $\alpha^*$ is the nominal significance level. This $\alpha^*$ can be calculated for any given $\alpha$.

The basic idea behind the monitoring procedures of Test 1 and Test 2 is that if the null hypothesis is not true, in that locations differ among treatment groups, we should expect to see

$$E[R_{i\alpha}^{(k)}] > \frac{m_{i\alpha}^{(k)}+1}{2} \quad \text{or} \quad E[R_{i\alpha}^{(k)}] < \frac{m_{i\alpha}^{(k)}+1}{2} \tag{3.37}$$

for at least one $k \in \{1, 2, \ldots, c\}$. In turn, that would lead to large values of $Z_{i\alpha}^{(k)}$, and hence the $S_\alpha^{(k)}$ and $S^{(k)}(t)$.

Although both Test 1 and Test 2 are fully sequential, the data is not necessarily analyzed after each new observation is received. Rather, for the observation received at time $t$, the number of complete subjects $n(t)$ is calculated. The data are then analyzed if this total number of complete subjects is larger (by one) than it was at the last analysis time. Otherwise, sampling continues.

### 3.2.3 Boundary calculation

The bounds used for both Test 1 and Test 2 are based on large sample approximations. The theory presented here is from Gombay [14].

To be consistent with notation from previous chapters, in particular Section 1.2.1, we write $S_j^{(k)}$ to be the value of the test statistic for group $k$ at time $t_j$,

i.e. $S^{(k)}(t_j)$. To ease the notation, we omit the superscript corresponding to treatment group, so that the quantity of interest is $S_j$. Then,

$$\lim_{N\to\infty} P\{a(N) \max_{1\leq j\leq N} j^{-1/2}|S_j| - b(N) \leq y\} \quad = \quad exp(-2e^{-y}) \qquad (3.38)$$

where $N$ is the number of subjects in any given group (i.e. $n_k$), $a(N) = (2\log\log N)^{1/2}$, and $b(N) = 2\log\log N + \frac{1}{2}\log\log\log N - \frac{1}{2}\log\pi$. The critical values for Test 1 can be found according to equation (3.38). However, Gombay [18] notes that better approximations are available, citing the work of Vostrikova [44]. Select Vostrikova critical values for Test 1 can be found in Tables 3.1 and 3.2.

Table 3.1: Critical values $C(\alpha, n_1)$ for Test 1 with $c = 2$.

| $n_1$ | $\alpha = 0.01$ | $\alpha = 0.05$ | $\alpha = 0.10$ |
|---|---|---|---|
| 10 | 3.41 | 2.84 | 2.55 |
| 15 | 3.45 | 2.90 | 2.61 |
| 20 | 3.48 | 2.93 | 2.65 |
| 50 | 3.56 | 3.02 | 2.74 |
| 100 | 3.60 | 3.07 | 2.80 |
| 150 | 3.63 | 3.10 | 2.83 |
| 200 | 3.64 | 3.12 | 2.85 |

Table 3.2: Critical values $C(\alpha^*, n_1)$ for Test 1 with $c = 3$ and overall error rate $\alpha$.

| $n_1$ | $\alpha = 0.01$ | $\alpha = 0.05$ | $\alpha = 0.10$ |
|---|---|---|---|
| 10 | 3.62 | 3.10 | 2.83 |
| 15 | 3.66 | 3.15 | 2.88 |
| 20 | 3.69 | 3.18 | 2.92 |
| 50 | 3.76 | 3.26 | 3.01 |
| 100 | 3.80 | 3.31 | 3.06 |
| 150 | 3.82 | 3.33 | 3.09 |
| 200 | 3.84 | 3.35 | 3.11 |

Now, suppose $W(t)$ is a standard Brownian motion. That is, $W(0) = 0$, $\{W(t), t \geq 0\}$ has stationary and independent increments, and for all $t > 0$, $W(t)$ is $N(0, t)$ [35]. Then, for Test 2, the critical values $C(\alpha)$ and $C(\alpha^*)$ are found according to the distribution of

$$\sup_{0 \leq u \leq 1} |W(u)|. \tag{3.39}$$

Select critical values $C(\alpha)$ and $C(\alpha^*)$ are presented in Table 3.3.

Table 3.3: Critical values $C(\alpha)$ for Test 2 with $c = 2$, and $C(\alpha^*)$ with $c = 3$ and overall error rate $\alpha$.

| | $\alpha = 0.01$ | $\alpha = 0.05$ | $\alpha = 0.10$ |
|---|---|---|---|
| $C(\alpha)$ | 2.80 | 2.24 | 1.96 |
| $C(\alpha^*)$ | 3.00 | 2.49 | 2.23 |

Formulas for calculating the Test 1 and Test 2 boundaries are given in Gombay [17]. According to Gombay [16], Test 2 has higher power when there are small differences between the population distributions. On the other hand, Test 1 has shorter stopping times if the differences are large.

## 3.3 Small-sample version

### 3.3.1 The test statistic

Recall that at time $t$, we assume that a total of $n(t)$ complete subjects are enrolled in the trial, where $n(t) = n_1(t) + n_2(t) + \ldots + n_c(t)$. Each of these patients has all $p$ repeated measurements observed.

We have previously considered sums

$$S_\alpha^{(k)} \;=\; \sum_{i=1}^{p} Z_{i\alpha}^{(k)} \tag{3.40}$$

for $k = 1, \ldots, c$ and $\alpha = 1, \ldots, n_k(t)$, where $Z_{i\alpha}^{(k)}$ was the standardized sequential rank for the measurement indexed by $i$, $k$, and $\alpha$. Now, we define the following quantity,

$$\text{sgn}(S_\alpha^{(k)}) = \begin{cases} 1 & , \text{ if } S_\alpha^{(k)} > 0 \\ 0 & , \text{ if } S_\alpha^{(k)} = 0 \\ -1 & , \text{ if } S_\alpha^{(k)} < 0 \end{cases} \tag{3.41}$$

using the so-called sign function. Under the null hypothesis, the $\text{sgn}(S_\alpha^{(k)})$ are i.i.d for $\alpha = 1, \ldots, n_k(t)$ and for fixed $k = 1, \ldots, c$, with

$$P\{\text{sgn}(S_\alpha^{(k)}) = 1\} = P\{\text{sgn}(S_\alpha^{(k)}) = -1\} = \frac{1}{2}. \tag{3.42}$$

Thus the sequences

$$S^{(k)}(t) \;=\; \sum_{\alpha=1}^{n_k(t)} \text{sgn}(S_\alpha^{(k)}) \tag{3.43}$$

for $k = 1, \ldots, c$, over time, are simple symmetric random walks.

### 3.3.2 Monitoring

We consider a single fully sequential monitoring procedure, which we call Test 3. Again, for this procedure the data is analyzed only after a newly received observation makes a subject complete.

First we consider the special case of $c = 2$. As with the large-sample version of Section 3.2.2, we only need to monitor one of the two groups. Without loss of generality, monitor group 1.

**Test 3**. Stop sampling and reject the null hypothesis if

$$\left|S^{(1)}(t)\right| \geq C(\alpha, n_1) \tag{3.44}$$

holds. The bound $C(\alpha, n_1)$ depends upon the overall significance level $\alpha$ and the fixed total number of subjects receiving treatment 1, $n_1$. Sampling continues if equation (3.44) does not hold. This test is two-sided.

Next we consider the comparison of three groups, that is where $c = 3$. We only need to monitor two of the groups, without loss of generality, groups 1 and 2. We again employ the relation

$$(1 - \alpha) \;\; = \;\; (1 - \alpha^*)^{c-1}. \tag{3.45}$$

with $c = 3$.

**Test 3**. Stop sampling and reject the null hypothesis if either

$$\left|S^{(1)}(t)\right| \geq C(\alpha^*, n_1) \tag{3.46}$$

or

$$\left|S^{(2)}(t)\right| \geq C(\alpha^*, n_2) \tag{3.47}$$

holds. The bounds $C(\alpha^*, n_1)$ and $C(\alpha^*, n_2)$ depend upon the overall significance level $\alpha$ through $\alpha^*$, and the fixed total numbers of subjects receiving treatments 1 and 2, $n_1$ and $n_2$. Sampling continues if neither equation (3.46) nor equation (3.47) holds. This test is two-sided.

### 3.3.3 Boundary calculation

The bounds for Test 3 are calculated using theory on simple random walks found in Csáki [5], as it was applied in Gombay [13]. In contrast to the critical values for Test 1 and Test 2, the bounds here are exact, not asymptotic approximations.

Similarly to the boundary calculation for the large-sample version, we write $S_j$ to be the value of the test statistic at $t_j$. The superscript corresponding to treatment group is once again omitted. First we define the following two quantities

$$M_N = \max_{0 \le j \le N} S_j \tag{3.48}$$

$$m_N = \min_{0 \le j \le N} S_j \tag{3.49}$$

where $N$ is the number of patients in any given group (i.e. $n_k$). Then, for $x_j = \frac{j\pi}{2a}$,

$$P\{-a < m_N < M_N < a\} = \frac{1}{a} \sum_{j=0}^{2a-1} (\cos x_j)^N \sin a x_j \, \frac{1 + \cos x_j}{\sin x_j} \left( \frac{1 - (-1)^j}{2} \right) \tag{3.50}$$

where $a > 1$ is an integer. Select critical values for Test 3 can be found in Tables 3.4 and 3.5. Note that as $a$ is restricted to the set of integers, the bounds might not correspond exactly to the desired $\alpha$. For this reason, we include the exact $\alpha$ in brackets after each bound, within both tables.

Table 3.4: Critical values $a = C(\alpha, n_1)$ for Test 3 with $c = 2$. The exact value for $\alpha$ is given in brackets after each critical value.

| $n_1$ | $\alpha = 0.01$ | $\alpha = 0.05$ | $\alpha = 0.10$ |
|---|---|---|---|
| 10 | 8 (0.0234) | 7 (0.0430) | 6 (0.1309) |
| 15 | 10 (0.0148) | 9 (0.0425) | 8 (0.0703) |
| 20 | 12 (0.0144) | 10 (0.0532) | 9 (0.0828) |
| 30 | 15 (0.0105) | 12 (0.0589) | 11 (0.0856) |
| 50 | 20 (0.0092) | 16 (0.0482) | 14 (0.0978) |
| 100 | 28 (0.0102) | 22 (0.0562) | 20 (0.0921) |
| 150 | 34 (0.0110) | 27 (0.0543) | 24 (0.1009) |
| 200 | 39 (0.0114) | 32 (0.0475) | 28 (0.0960) |

Table 3.5: Critical values $a = C(\alpha^*, n_1)$ for Test 3 with $c = 3$ and overall error rate $\alpha$. The exact value for $\alpha$ is given in brackets after each critical value.

| $n_1$ | $\alpha = 0.01$ | $\alpha = 0.05$ | $\alpha = 0.10$ |
|---|---|---|---|
| 10 | 9 (0.0078) | 8 (0.0463) | 7 (0.0841) |
| 15 | 12 (0.0039) | 10 (0.0293) | 9 (0.0833) |
| 20 | 13 (0.0103) | 11 (0.0467) | 10 (0.1036) |
| 30 | 16 (0.0133) | 14 (0.0422) | 12 (0.1143) |
| 50 | 21 (0.0104) | 18 (0.0434) | 16 (0.0941) |
| 100 | 30 (0.0106) | 25 (0.0476) | 22 (0.1092) |
| 150 | 37 (0.0096) | 31 (0.0440) | 27 (0.1057) |
| 200 | 43 (0.0091) | 35 (0.0518) | 32 (0.0928) |

# Chapter 4

# Simulation

## 4.1 Preliminaries

This simulation study examines the performance of Test 1, Test 2, and Test 3 in the context of comparing either two or three treatment groups. Each subject from each group is assumed to have a maximum of three repeated measurements. In other words, we set $c = 2$ or 3 and $p = 3$. We assume that there are equal numbers of subjects from each group: $n_1 = n_2$, or $n_1 = n_2 = n_3$.

Data is simulated in Fortran using the IMSL Numerical Libraries. To induce correlation among repeated observations, we use the following moving average (MA) time series model, for its simplicity:

$$Y_{i\alpha}^{(k)} \quad = \quad w_{i\alpha}^{(k)} + \theta_1 w_{(i-1)\alpha}^{(k)} + \theta_2 w_{(i-2)\alpha}^{(k)}. \tag{4.1}$$

The $w_{j\alpha}^{(k)}$ are simulated independent random variates, for $j = -1, 0, \ldots, 3$, $k = 1, 2$ or $k = 1, \ldots, 3$, and $\alpha = 1, \ldots, n_k$, that are either normally distributed, with mean $\mu_k$ and variance 1, or exponentially distributed with mean $1/\lambda_k$. In the normal case, we set $\mu_2 = 0$ or $\mu_2 = \mu_3 = 0$, while $\mu_1$ is varied over the set $\{0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0\}$. For the exponential case, we set $\lambda_2 = 1$ or $\lambda_2 = \lambda_3 = 1$, while $1/\lambda_1$ is varied over the set $\{1.00, 1.25, 1.50, 1.75, 2.00, 2.25, 2.50, 2.75, 3.00\}$.

Within-subject correlation arises when we restrict $\theta_1$ and $\theta_2$ to be nonzero. In our case, we choose $\theta_1 = \theta_2 = \frac{1}{2}$. Then, omitting the $k$ and $\alpha$ indices for

brevity,

$$\begin{aligned} COV[y_{i+u}, y_i] &= COV[w_{i+u} + \frac{1}{2}w_{i+u-1} + \frac{1}{2}w_{i+u-2}, w_i + \frac{1}{2}w_{i-1} + \frac{1}{2}w_{i-2}] \\ &= COV[w_{i+u}w_i] + \frac{1}{2}COV[w_{i+u}w_{i-1}] + \frac{1}{2}COV[w_{i+u}w_{i-2}] \\ &\quad + \frac{1}{2}COV[w_{i+u-1}w_i] + \frac{1}{4}COV[w_{i+u-1}w_{i-1}] \\ &\quad + \frac{1}{4}COV[w_{i+u-1}w_{i-2}] + \frac{1}{2}COV[w_{i+u-2}w_i] \\ &\quad + \frac{1}{4}COV[w_{i+u-2}w_{i-1}] + \frac{1}{4}COV[w_{i+u-2}w_{i-2}]. \end{aligned} \tag{4.2}$$

Recalling that for $i \neq j$, $COV[w_i, w_j] = 0$, and that $COV[w_i, w_i] = VAR[w_i]$, we have:

$$\begin{aligned} u = 0 &\Rightarrow COV[y_i \quad, y_i] = VAR[w_i] + \frac{1}{4}VAR[w_{i-1}] + \frac{1}{4}VAR[w_{i-2}] \\ u = 1 &\Rightarrow COV[y_{i+1}, y_i] = \frac{1}{2}VAR[w_i] + \frac{1}{4}VAR[w_{i-1}] \\ u = 2 &\Rightarrow COV[y_{i+2}, y_i] = \frac{1}{2}VAR[w_i]. \end{aligned}$$

As the variance of $w_i$ does not change, for various $i$, we may rewrite the above as:

$$\begin{aligned} COV[y_i \quad, y_i] &= \frac{3}{2}VAR[w_i] \\ COV[y_{i+1}, y_i] &= \frac{3}{4}VAR[w_i] \\ COV[y_{i+2}, y_i] &= \frac{1}{2}VAR[w_i]. \end{aligned}$$

Now, using

$$CORR[y_{i+u}, y_i] = \frac{COV[y_{i+u}, y_i]}{\sqrt{VAR[y_{i+u}]VAR[y_i]}} \tag{4.3}$$

and writing $\rho(u) = CORR[y_{i+u}, y_i]$, we have

$$\rho(u) = \begin{cases} 1 & , u = 0 \\ \frac{1}{2} & , |u| = 1 \\ \frac{1}{3} & , |u| = 2 \end{cases} \tag{4.4}$$

where $u$ is the lag value. Values of $\rho(u)$ for $|u| \geq 3$ are not relevant for this study, as setting $p = 3$ implies that repeated measurements can be a maximum of two time points apart. In any case, $\rho(u) = 0$ for $|u| \geq 3$.

For computational purposes, it is also necessary to simulate the order in which the data are received. We simulate the orders all at once, in a process that is independent from the data generation. This is done for two reasons: first, it guarantees an equal distribution of subjects across groups; and second, the corresponding code was thought to be less computationally expensive than alternative algorithms.

Given that we have three measurements on $n$ subjects, we must generate a random permutation of the numbers $\{1, 2, \ldots, 3n\}$, where the numbers $1, 2, \ldots, n$ represent the first measurements on the $n$ subjects, $n + 1, n + 2, \ldots, 2n$ represent the second measurements, and so on. We must ensure however that, for example, 1 is not observed *after* $n + 1$ or $2n + 1$, as this would be nonsensical. Denote $\mathbf{o}$ to be the $3n \times 1$ ordering vector. Initially, this vector is empty. The simulation algorithm proceeds as follows, where, to begin with $\mathcal{H} = \{1, 2, \ldots, n\}$.

1. Randomly choose an integer $h$ from $\mathcal{H}$.
2. Place $h$, $n + h$, and $2n + h$ in that order, in the next three available entries of $\mathbf{o}$ (the first three entries if $\mathbf{o}$ is empty).
3. Remove $h$ from $\mathcal{H}$.
4. Repeat steps 1-3 until $\mathcal{H}$ is empty.

Then, the first entry of $\mathbf{o}$ is taken to be the observation number corresponding to the first observation received, the second entry is taken to be the number corresponding to the second observation, and so on. The group index $k$ is accounted for in the structure of our data storage. In particular, the first $n_1$ subjects out of $1, \ldots, n$ belong to group 1, the next $n_2$ subjects belong to group 2, and so on.

We define empirical power (P) to be the relative frequency of null hypothesis rejection, for a single set of replicates, and given a treatment difference. When there is no treatment difference, empirical power is equivalent to the estimate of empirical Type I error, $\hat{\alpha}$. The average stopping time (AVST) is taken to be the average of either the number of complete observations that are observed at the time of rejection (for replicates that reject the null hypothesis), or the maximal total sample size, within a single set of replicates.

Finally, the results of each simulation are based upon $10^4$ replications.

## 4.2 Comparing $c$=2 groups

### 4.2.1 Algorithms

**Test 1**

1. Generate data according to equation (4.1) for specified $n_1 = n_2$, and $\mu_1$ or $\lambda_1$, depending on the response distribution.

2. Generate the hypothetical orders of observation.

3. Calculate $n$ test statistics using the complete subjects from group 1; treat the data as if it arrives according to the ordering generated in step (2).

4. Compare the test statistics from step (3) to a critical value, according to equation (3.20) and the values of $n_1$ and $\alpha$.

5. Record the first observation for which equation (3.20) is satisfied, nothing otherwise.

6. Repeat steps $(1) - (5)$ 9999 additional times.

7. Calculate and record $\hat{\alpha}$ (for $\mu_1 = 0$, $\lambda_1 = 1$) or empirical power (for $\mu_1 \neq 0$, $\lambda_1 \neq 1$), along with the average stopping time.

**Test 2**

1. Generate data according to equation (4.1) for specified $n_1 = n_2$, and $\mu_1$ or $\lambda_1$, depending on the response distribution.

2. Generate the hypothetical orders of observation.

3. Calculate $n$ test statistics using the complete subjects from group 1; treat the data as if it arrives according to the ordering generated in step (2).

4. Compare the test statistics from step (3) to a critical value, according to equation (3.21) and the value of $\alpha$.

5. Record the first observation for which equation (3.21) is satisfied, nothing otherwise.

6. Repeat steps $(1) - (5)$ 9999 additional times.

7. Calculate and record $\hat{\alpha}$ (for $\mu_1 = 0$, $\lambda_1 = 1$) or empirical power (for $\mu_1 \neq 0$, $\lambda_1 \neq 1$), along with the average stopping time.

**Test 3**

1. Generate data according to equation (4.1) for specified $n_1 = n_2$, and $\mu_1$ or $\lambda_1$, depending on the response distribution.

2. Generate the hypothetical orders of observation.

3. Calculate $n$ test statistics using the complete subjects from group 1; treat the data as if it arrives according to the ordering generated in step (2).

4. Compare the test statistics from step (3) to a critical value, according to equation (3.44) and the values of $n_1$ and $\alpha$.

5. Record the first observation for which equation (3.44) is satisfied, nothing otherwise.

6. Repeat steps $(1) - (5)$ 9999 additional times.

7. Calculate and record $\hat{\alpha}$ (for $\mu_1 = 0$, $\lambda_1 = 1$) or empirical power (for $\mu_1 \neq 0$, $\lambda_1 \neq 1$), along with the average stopping time.

### 4.2.2 Results

Here we present the results of our simulation study for $c = 2$. The results corresponding to Test 1, Test 2, and Test 3 are shown in Tables 4.1 to 4.8.

Table 4.1: Simulated power (P) and average stopping time (AVST) for Test 1 (Pocock) with $c = 2$ and normally distributed responses. We set $\mu_2 = 0$ and $\sigma^2 = 1$, while $\mu_1$ is allowed to vary. Critical values are obtained from Table 3.1. For this simulation, $n_1 = n_2 = 20$ or 50.

| | | $\alpha = 0.01$ | | $\alpha = 0.05$ | | $\alpha = 0.10$ | |
|---|---|---|---|---|---|---|---|
| $n$ | $\mu_1$ | P | AVST | P | AVST | P | AVST |
| 40 | | | | | | | |
| | 0.0 | 0.0173 | 39.5133 | 0.0554 | 38.4350 | 0.1096 | 36.8780 |
| | 0.1 | 0.0194 | 39.4797 | 0.0603 | 38.3560 | 0.1104 | 37.0020 |
| | 0.2 | 0.0260 | 39.3964 | 0.0805 | 38.0311 | 0.1498 | 36.3116 |
| | 0.3 | 0.0337 | 39.2752 | 0.1115 | 37.5975 | 0.1994 | 35.6633 |
| | 0.4 | 0.0591 | 38.9247 | 0.1695 | 36.7517 | 0.2708 | 34.5776 |
| | 0.5 | 0.0891 | 38.6271 | 0.2433 | 35.7444 | 0.3612 | 33.2796 |
| | 0.6 | 0.1378 | 38.0450 | 0.3257 | 34.6365 | 0.4602 | 31.7124 |
| | 0.7 | 0.1978 | 37.4216 | 0.4244 | 33.3640 | 0.5636 | 29.8730 |
| | 0.8 | 0.2547 | 36.8803 | 0.5157 | 31.9556 | 0.6553 | 28.2752 |
| | 0.9 | 0.3246 | 36.0526 | 0.6004 | 30.4940 | 0.7275 | 26.7073 |
| | 1.0 | 0.3934 | 35.2575 | 0.6642 | 29.3492 | 0.7819 | 25.4803 |
| 100 | | | | | | | |
| | 0.0 | 0.0151 | 98.7768 | 0.0607 | 95.3353 | 0.1044 | 91.6534 |
| | 0.1 | 0.0213 | 98.4530 | 0.0758 | 94.5691 | 0.1279 | 90.8122 |
| | 0.2 | 0.0453 | 97.6264 | 0.1322 | 92.6771 | 0.2198 | 87.0641 |
| | 0.3 | 0.1084 | 95.6903 | 0.2733 | 87.0563 | 0.3874 | 79.7445 |
| | 0.4 | 0.2333 | 91.8371 | 0.4694 | 79.2748 | 0.5941 | 70.5520 |
| | 0.5 | 0.4228 | 85.0126 | 0.6648 | 70.6474 | 0.7839 | 59.7643 |
| | 0.6 | 0.6283 | 77.2578 | 0.8332 | 60.2236 | 0.8985 | 50.9883 |
| | 0.7 | 0.7948 | 68.0099 | 0.9130 | 52.7937 | 0.9603 | 43.4149 |
| | 0.8 | 0.8872 | 61.1412 | 0.9674 | 45.3806 | 0.9843 | 37.4322 |
| | 0.9 | 0.9424 | 55.2382 | 0.9838 | 40.6740 | 0.9949 | 33.2430 |
| | 1.0 | 0.9714 | 50.2889 | 0.9943 | 36.7755 | 0.9972 | 30.9528 |

Table 4.2: Simulated power (P) and average stopping time (AVST) for Test 2 (O'Brien-Fleming) with $c = 2$ and normally distributed responses. We set $\mu_2 = 0$ and $\sigma^2 = 1$, while $\mu_1$ is allowed to vary. Critical values are obtained from Table 3.3. For this simulation, $n_1 = n_2 = 20$ or 50.

| | | $\alpha = 0.01$ | | $\alpha = 0.05$ | | $\alpha = 0.10$ | |
|---|---|---|---|---|---|---|---|
| $n$ | $\mu_1$ | P | AVST | P | AVST | P | AVST |
| 40 | | | | | | | |
| | 0.0 | 0.0113 | 39.9116 | 0.0485 | 39.4809 | 0.0992 | 38.8507 |
| | 0.1 | 0.0156 | 39.8808 | 0.0681 | 39.3601 | 0.1172 | 38.7396 |
| | 0.2 | 0.0343 | 39.7747 | 0.1100 | 39.0462 | 0.1849 | 38.0607 |
| | 0.3 | 0.0689 | 39.5443 | 0.1931 | 38.3686 | 0.2928 | 36.9608 |
| | 0.4 | 0.1254 | 39.2127 | 0.3014 | 37.3615 | 0.4250 | 35.5745 |
| | 0.5 | 0.2150 | 38.6921 | 0.4255 | 36.1377 | 0.5653 | 33.7774 |
| | 0.6 | 0.3020 | 38.0834 | 0.5481 | 34.7093 | 0.6857 | 32.0186 |
| | 0.7 | 0.4154 | 37.1331 | 0.6734 | 33.2374 | 0.7801 | 30.3460 |
| | 0.8 | 0.5206 | 36.2689 | 0.7610 | 31.7551 | 0.8438 | 28.8237 |
| | 0.9 | 0.6134 | 35.3126 | 0.8272 | 30.4689 | 0.8960 | 27.4258 |
| | 1.0 | 0.6904 | 34.4186 | 0.8749 | 29.3694 | 0.9271 | 26.3040 |
| 100 | | | | | | | |
| | 0.0 | 0.0102 | 99.8554 | 0.0481 | 98.8953 | 0.0945 | 97.4671 |
| | 0.1 | 0.0268 | 99.5778 | 0.0902 | 98.0556 | 0.1560 | 96.1170 |
| | 0.2 | 0.0926 | 98.4398 | 0.2445 | 94.4426 | 0.3399 | 91.0472 |
| | 0.3 | 0.2440 | 95.6965 | 0.4689 | 88.2038 | 0.5859 | 82.6154 |
| | 0.4 | 0.4752 | 90.4047 | 0.7066 | 79.6203 | 0.8075 | 72.4948 |
| | 0.5 | 0.6920 | 83.6821 | 0.8683 | 70.9685 | 0.9230 | 63.6770 |
| | 0.6 | 0.8504 | 76.6676 | 0.9512 | 63.3259 | 0.9752 | 56.2283 |
| | 0.7 | 0.9388 | 69.8403 | 0.9851 | 57.2303 | 0.9928 | 50.6957 |
| | 0.8 | 0.9757 | 64.9024 | 0.9957 | 52.5814 | 0.9981 | 46.4053 |
| | 0.9 | 0.9914 | 60.7696 | 0.9987 | 49.3626 | 0.9994 | 43.5393 |
| | 1.0 | 0.9957 | 57.5219 | 0.9996 | 46.7485 | 0.9997 | 41.2584 |

Table 4.3: Simulated power (P) and average stopping time (AVST) for Test 3 with $c = 2$ and normally distributed responses. We set $\mu_2 = 0$ and $\sigma^2 = 1$, while $\mu_1$ is allowed to vary. Critical values are obtained from Table 3.4. For this simulation, $n_1 = n_2 = 10$ or 15.

| | | $\alpha = 0.01$ | | $\alpha = 0.05$ | | $\alpha = 0.10$ | |
|---|---|---|---|---|---|---|---|
| $n$ | $\mu_1$ | P | AVST | P | AVST | P | AVST |
| 20 | | | | | | | |
| | 0.0 | 0.0135 | 19.9691 | 0.0342 | 19.8944 | 0.0958 | 19.6232 |
| | 0.1 | 0.0138 | 19.9693 | 0.0446 | 19.8668 | 0.1034 | 19.5845 |
| | 0.2 | 0.0244 | 19.9443 | 0.0589 | 19.8175 | 0.1423 | 19.4238 |
| | 0.3 | 0.0362 | 19.9271 | 0.0835 | 19.7386 | 0.1749 | 19.3135 |
| | 0.4 | 0.0625 | 19.8740 | 0.1300 | 19.5972 | 0.2353 | 19.0525 |
| | 0.5 | 0.0916 | 19.7970 | 0.1810 | 19.4404 | 0.3073 | 18.7081 |
| | 0.6 | 0.1354 | 19.7030 | 0.2470 | 19.2351 | 0.3889 | 18.3226 |
| | 0.7 | 0.1907 | 19.5669 | 0.3155 | 18.9659 | 0.4675 | 17.9176 |
| | 0.8 | 0.2423 | 19.4401 | 0.3837 | 18.7125 | 0.5375 | 17.5350 |
| | 0.9 | 0.3022 | 19.3008 | 0.4542 | 18.4306 | 0.6019 | 17.1138 |
| | 1.0 | 0.3612 | 19.1005 | 0.5237 | 18.1477 | 0.6599 | 16.7249 |
| 30 | | | | | | | |
| | 0.0 | 0.0125 | 29.9475 | 0.0328 | 29.8495 | 0.0617 | 29.6430 |
| | 0.1 | 0.0172 | 29.9346 | 0.0379 | 29.8136 | 0.0743 | 29.5589 |
| | 0.2 | 0.0291 | 29.8836 | 0.0652 | 29.6889 | 0.1114 | 29.3495 |
| | 0.3 | 0.0533 | 29.7879 | 0.1017 | 29.5078 | 0.1601 | 29.0229 |
| | 0.4 | 0.0987 | 29.5874 | 0.1570 | 29.2143 | 0.2508 | 28.4765 |
| | 0.5 | 0.1416 | 29.3891 | 0.2328 | 28.8069 | 0.3312 | 27.9046 |
| | 0.6 | 0.2199 | 29.0468 | 0.3278 | 28.2333 | 0.4348 | 27.1160 |
| | 0.7 | 0.3006 | 28.6415 | 0.4208 | 27.6018 | 0.5467 | 26.1137 |
| | 0.8 | 0.3890 | 28.1130 | 0.5159 | 26.8844 | 0.6391 | 25.2488 |
| | 0.9 | 0.4795 | 27.5732 | 0.5980 | 26.2403 | 0.7094 | 24.4556 |
| | 1.0 | 0.5754 | 26.9621 | 0.6778 | 25.4138 | 0.7824 | 23.5164 |

Table 4.4: Simulated power (P) and average stopping time (AVST) for Test 3 with $c = 2$ and normally distributed responses. We set $\mu_2 = 0$ and $\sigma^2 = 1$, while $\mu_1$ is allowed to vary. Critical values are obtained from Table 3.4. For this simulation, $n_1 = n_2 = 20$ or 50.

| | | $\alpha = 0.01$ | | $\alpha = 0.05$ | | $\alpha = 0.10$ | |
|---|---|---|---|---|---|---|---|
| $n$ | $\mu_1$ | P | AVST | P | AVST | P | AVST |
| 40 | | | | | | | |
| | 0.0 | 0.0098 | 39.9491 | 0.0422 | 39.6785 | 0.0781 | 39.3054 |
| | 0.1 | 0.0157 | 39.9190 | 0.0486 | 39.6390 | 0.0944 | 39.1801 |
| | 0.2 | 0.0299 | 39.8271 | 0.0946 | 39.3086 | 0.1391 | 38.7755 |
| | 0.3 | 0.0610 | 39.6658 | 0.1559 | 38.8287 | 0.2314 | 37.9492 |
| | 0.4 | 0.1170 | 39.3273 | 0.2497 | 37.9685 | 0.3416 | 36.8532 |
| | 0.5 | 0.1835 | 38.9078 | 0.3611 | 36.9027 | 0.4613 | 35.4598 |
| | 0.6 | 0.2956 | 38.1492 | 0.4782 | 35.7502 | 0.5896 | 33.8157 |
| | 0.7 | 0.3958 | 37.3613 | 0.5846 | 34.4587 | 0.6912 | 32.1288 |
| | 0.8 | 0.4989 | 36.4311 | 0.6958 | 32.8527 | 0.7864 | 30.4702 |
| | 0.9 | 0.6033 | 35.2340 | 0.7761 | 31.3424 | 0.8490 | 28.8872 |
| | 1.0 | 0.6854 | 34.1926 | 0.8353 | 30.0287 | 0.8919 | 27.7650 |
| 100 | | | | | | | |
| | 0.0 | 0.0063 | 99.9020 | 0.0409 | 99.2356 | 0.0876 | 97.9282 |
| | 0.1 | 0.0204 | 99.6872 | 0.0728 | 98.4573 | 0.1335 | 96.6854 |
| | 0.2 | 0.0631 | 99.0528 | 0.1802 | 96.1684 | 0.2769 | 92.9619 |
| | 0.3 | 0.1573 | 97.4256 | 0.3578 | 91.6262 | 0.4757 | 86.7851 |
| | 0.4 | 0.3179 | 94.1400 | 0.5591 | 85.4084 | 0.6799 | 78.8087 |
| | 0.5 | 0.5133 | 89.2990 | 0.7534 | 77.6799 | 0.8345 | 70.2216 |
| | 0.6 | 0.7088 | 83.2849 | 0.8751 | 70.0850 | 0.9234 | 62.5338 |
| | 0.7 | 0.8417 | 76.5417 | 0.9484 | 62.7229 | 0.9741 | 55.5933 |
| | 0.8 | 0.9278 | 70.1710 | 0.9794 | 57.3000 | 0.9904 | 50.7586 |
| | 0.9 | 0.9697 | 65.1492 | 0.9911 | 52.9423 | 0.9973 | 46.6311 |
| | 1.0 | 0.9841 | 61.1805 | 0.9970 | 49.5758 | 0.9993 | 43.7730 |

Table 4.5: Simulated power (P) and average stopping time (AVST) for Test 1 (Pocock) with $c = 2$ and exponentially distributed responses. We set $\lambda_2 = 1$ while $\lambda_1$ is allowed to vary. Critical values are obtained from Table 3.1. For this simulation, $n_1 = n_2 = 20$ or 50.

| | | $\alpha = 0.01$ | | $\alpha = 0.05$ | | $\alpha = 0.10$ | |
|---|---|---|---|---|---|---|---|
| $n$ | $1/\lambda_1$ | P | AVST | P | AVST | P | AVST |
| 40 | | | | | | | |
| | 1.00 | 0.0187 | 39.4615 | 0.0558 | 38.4326 | 0.1053 | 37.0108 |
| | 1.25 | 0.0244 | 39.4253 | 0.0853 | 37.9654 | 0.1441 | 36.4624 |
| | 1.50 | 0.0529 | 39.0391 | 0.1592 | 36.9231 | 0.2660 | 34.5801 |
| | 1.75 | 0.1015 | 38.5135 | 0.2789 | 35.3512 | 0.4068 | 32.4878 |
| | 2.00 | 0.1613 | 37.8548 | 0.3956 | 33.6558 | 0.5352 | 30.4308 |
| | 2.25 | 0.2430 | 37.0023 | 0.4944 | 32.3860 | 0.6461 | 28.5310 |
| | 2.50 | 0.3104 | 36.2751 | 0.5961 | 30.6735 | 0.7155 | 27.0975 |
| | 2.75 | 0.3603 | 35.7146 | 0.6482 | 29.7022 | 0.7629 | 26.0088 |
| | 3.00 | 0.4179 | 34.9914 | 0.7004 | 28.7951 | 0.8132 | 24.9169 |
| 100 | | | | | | | |
| | 1.00 | 0.0164 | 98.6860 | 0.0545 | 95.6698 | 0.1049 | 91.5986 |
| | 1.25 | 0.0510 | 97.4387 | 0.1534 | 91.5293 | 0.2471 | 85.7574 |
| | 1.50 | 0.2164 | 92.3186 | 0.4451 | 80.5251 | 0.5820 | 71.2014 |
| | 1.75 | 0.5036 | 82.0065 | 0.7420 | 65.9060 | 0.8412 | 55.9079 |
| | 2.00 | 0.7429 | 71.3736 | 0.9017 | 54.2947 | 0.9421 | 45.6626 |
| | 2.25 | 0.8659 | 63.5733 | 0.9617 | 46.8587 | 0.9832 | 38.3031 |
| | 2.50 | 0.9371 | 56.4107 | 0.9822 | 41.5005 | 0.9928 | 34.1177 |
| | 2.75 | 0.9661 | 51.8288 | 0.9911 | 38.2637 | 0.9982 | 31.3079 |
| | 3.00 | 0.9798 | 48.4616 | 0.9964 | 35.4533 | 0.9985 | 29.3085 |

Table 4.6: Simulated power (P) and average stopping time (AVST) for Test 2 (O'Brien-Fleming) with $c = 2$ and exponentially distributed responses. We set $\lambda_2 = 1$ while $\lambda_1$ is allowed to vary. Critical values are obtained from Table 3.3. For this simulation, $n_1 = n_2 = 20$ or 50.

| | | $\alpha = 0.01$ | | $\alpha = 0.05$ | | $\alpha = 0.10$ | |
|---|---|---|---|---|---|---|---|
| $n$ | $1/\lambda_1$ | P | AVST | P | AVST | P | AVST |
| 40 | | | | | | | |
| | 1.00 | 0.0118 | 39.9012 | 0.0531 | 39.4518 | 0.0951 | 38.8878 |
| | 1.25 | 0.0361 | 39.7707 | 0.1215 | 38.9322 | 0.1979 | 37.9518 |
| | 1.50 | 0.1151 | 39.2990 | 0.2901 | 37.5493 | 0.4131 | 35.6887 |
| | 1.75 | 0.2425 | 38.4504 | 0.4862 | 35.6353 | 0.6111 | 33.2663 |
| | 2.00 | 0.3729 | 37.5814 | 0.6415 | 33.6804 | 0.7563 | 30.9513 |
| | 2.25 | 0.4905 | 36.5747 | 0.7531 | 32.0924 | 0.8369 | 29.1562 |
| | 2.50 | 0.5954 | 35.5478 | 0.8202 | 30.8010 | 0.8977 | 27.5710 |
| | 2.75 | 0.6639 | 34.8116 | 0.8578 | 29.7516 | 0.9222 | 26.6656 |
| | 3.00 | 0.7240 | 34.0519 | 0.8990 | 28.9439 | 0.9380 | 25.8254 |
| 100 | | | | | | | |
| | 1.00 | 0.0089 | 99.8459 | 0.0458 | 98.9693 | 0.0935 | 97.5444 |
| | 1.25 | 0.1111 | 98.2016 | 0.2747 | 93.7279 | 0.3782 | 89.7691 |
| | 1.50 | 0.4498 | 91.1298 | 0.6975 | 80.2560 | 0.7842 | 73.7791 |
| | 1.75 | 0.7622 | 80.9086 | 0.9082 | 67.7756 | 0.9517 | 60.3091 |
| | 2.00 | 0.9227 | 72.2142 | 0.9762 | 59.2226 | 0.9884 | 52.4821 |
| | 2.25 | 0.9709 | 66.0885 | 0.9945 | 53.6038 | 0.9976 | 47.2075 |
| | 2.50 | 0.9887 | 61.7309 | 0.9974 | 49.9684 | 0.9991 | 43.9182 |
| | 2.75 | 0.9955 | 58.4604 | 0.9993 | 47.4265 | 0.9995 | 42.0699 |
| | 3.00 | 0.9980 | 56.2937 | 0.9996 | 45.7383 | 0.9999 | 40.4613 |

Table 4.7: Simulated power (P) and average stopping time (AVST) for Test 3 with $c = 2$ and exponentially distributed responses. We set $\lambda_2 = 1$ while $\lambda_1$ is allowed to vary. Critical values are obtained from Table 3.4. For this simulation, $n_1 = n_2 = 10$ or 15.

| | | $\alpha = 0.01$ | | $\alpha = 0.05$ | | $\alpha = 0.10$ | |
|---|---|---|---|---|---|---|---|
| $n$ | $1/\lambda_1$ | P | AVST | P | AVST | P | AVST |
| 20 | | | | | | | |
| | 1.00 | 0.0125 | 19.9721 | 0.0388 | 19.8820 | 0.1025 | 19.5772 |
| | 1.25 | 0.0247 | 19.9483 | 0.0607 | 19.8154 | 0.1351 | 19.4750 |
| | 1.50 | 0.0585 | 19.8833 | 0.1215 | 19.6287 | 0.2232 | 19.1301 |
| | 1.75 | 0.1032 | 19.7793 | 0.1894 | 19.4256 | 0.3128 | 18.7106 |
| | 2.00 | 0.1569 | 19.6611 | 0.2775 | 19.1165 | 0.4208 | 18.1860 |
| | 2.25 | 0.2164 | 19.5009 | 0.3439 | 18.8905 | 0.5112 | 17.7164 |
| | 2.50 | 0.2625 | 19.3987 | 0.4261 | 18.5693 | 0.5769 | 17.3481 |
| | 2.75 | 0.3035 | 19.2948 | 0.4871 | 18.3403 | 0.6283 | 16.9803 |
| | 3.00 | 0.3589 | 19.1403 | 0.5353 | 18.1301 | 0.6796 | 16.6863 |
| 30 | | | | | | | |
| | 1.00 | 0.0117 | 29.9575 | 0.0305 | 29.8549 | 0.0647 | 29.6278 |
| | 1.25 | 0.0312 | 29.8775 | 0.0627 | 29.6882 | 0.1188 | 29.3127 |
| | 1.50 | 0.0782 | 29.6856 | 0.1461 | 29.2803 | 0.2209 | 28.6703 |
| | 1.75 | 0.1526 | 29.3781 | 0.2539 | 28.6867 | 0.3591 | 27.7185 |
| | 2.00 | 0.2527 | 28.9031 | 0.3626 | 28.0436 | 0.4984 | 26.6228 |
| | 2.25 | 0.3419 | 28.4504 | 0.4673 | 27.3269 | 0.5979 | 25.7248 |
| | 2.50 | 0.4386 | 27.8970 | 0.5638 | 26.5706 | 0.6857 | 24.8404 |
| | 2.75 | 0.5073 | 27.4298 | 0.6225 | 25.9931 | 0.7481 | 24.1080 |
| | 3.00 | 0.5722 | 26.9911 | 0.6857 | 25.4013 | 0.8004 | 23.4199 |

Table 4.8: Simulated power (P) and average stopping time (AVST) for Test 3 with $c = 2$ and exponentially distributed responses. We set $\lambda_2 = 1$ while $\lambda_1$ is allowed to vary. Critical values are obtained from Table 3.4. For this simulation, $n_1 = n_2 = 20$ or 50.

| | | $\alpha = 0.01$ | | $\alpha = 0.05$ | | $\alpha = 0.10$ | |
|---|---|---|---|---|---|---|---|
| $n$ | $1/\lambda_1$ | P | AVST | P | AVST | P | AVST |
| 40 | | | | | | | |
| | 1.00 | 0.0097 | 39.9455 | 0.0423 | 39.6858 | 0.0773 | 39.3524 |
| | 1.25 | 0.0294 | 39.8466 | 0.0989 | 39.2916 | 0.1507 | 38.7172 |
| | 1.50 | 0.0994 | 39.4503 | 0.2292 | 38.2280 | 0.3128 | 37.1374 |
| | 1.75 | 0.2084 | 38.7867 | 0.3861 | 36.8053 | 0.4938 | 35.1434 |
| | 2.00 | 0.3286 | 37.9632 | 0.5413 | 35.0343 | 0.6498 | 32.9573 |
| | 2.25 | 0.4492 | 36.9257 | 0.6579 | 33.4954 | 0.7565 | 31.2254 |
| | 2.50 | 0.5517 | 35.9283 | 0.7396 | 32.2339 | 0.8271 | 29.7844 |
| | 2.75 | 0.6425 | 34.9190 | 0.8120 | 30.8671 | 0.8746 | 28.4851 |
| | 3.00 | 0.7129 | 34.0637 | 0.8568 | 29.8053 | 0.9055 | 27.5556 |
| 100 | | | | | | | |
| | 1.00 | 0.0069 | 99.8942 | 0.0466 | 99.0283 | 0.0867 | 97.9341 |
| | 1.25 | 0.0607 | 99.0467 | 0.1948 | 95.8104 | 0.2926 | 92.5216 |
| | 1.50 | 0.2856 | 95.0681 | 0.5201 | 86.9291 | 0.6484 | 80.4260 |
| | 1.75 | 0.5739 | 87.7758 | 0.7948 | 75.6497 | 0.8727 | 68.1847 |
| | 2.00 | 0.7836 | 80.0483 | 0.9224 | 66.3601 | 0.9575 | 58.9282 |
| | 2.25 | 0.9043 | 73.0946 | 0.9725 | 59.5967 | 0.9848 | 52.6946 |
| | 2.50 | 0.9601 | 67.3859 | 0.9898 | 55.0059 | 0.9949 | 48.7150 |
| | 2.75 | 0.9816 | 63.4632 | 0.9945 | 51.6112 | 0.9975 | 45.3563 |
| | 3.00 | 0.9901 | 60.2169 | 0.9986 | 48.7659 | 0.9995 | 43.2986 |

## 4.3 Comparing $c$=3 groups

### 4.3.1 Algorithms

**Test 1**

1. Generate data according to equation (4.1) for specified $n_1 = n_2 = n_3$, and $\mu_1$ or $\lambda_1$, depending on the response distribution.

2. Generate the hypothetical orders of observation.

3. Calculate $n$ test statistics corresponding to group 1, and $n$ corresponding to group 2; treat the data as if it arrives according to the ordering generated in step (2).

4. Compare the test statistics from step (3) to a critical value, according to equations (3.23) and (3.24), and the values of $n_1 = n_2$, and $\alpha^*$.

5. Record the first observation for which one of equations (3.23) or (3.24) is satisfied, nothing otherwise.

6. Repeat steps $(1) - (5)$ 9999 additional times.

7. Calculate and record $\hat{\alpha}$ (for $\mu_1 = 0$, $\lambda_1 = 1$) or empirical power (for $\mu_1 \neq 0$, $\lambda_1 \neq 1$), along with the average stopping time.

**Test 2**

1. Generate data according to equation (4.1) for specified $n_1 = n_2 = n_3$, and $\mu_1$ or $\lambda_1$, depending on the response distribution.

2. Generate the hypothetical orders of observation.

3. Calculate $n$ test statistics corresponding to group 1, and $n$ corresponding to group 2; treat the data as if it arrives according to the ordering generated in step (2).

4. Compare the test statistics from step (3) to a critical value, according to equations (3.25) and (3.26), and the value of $\alpha^*$.

5. Record the first observation for which one of equations (3.25) or (3.26) is satisfied, nothing otherwise.

6. Repeat steps $(1) - (5)$ 9999 additional times.

7. Calculate and record $\hat{\alpha}$ (for $\mu_1 = 0$, $\lambda_1 = 1$) or empirical power (for $\mu_1 \neq 0$, $\lambda_1 \neq 1$), along with the average stopping time.

**Test 3**

1. Generate data according to equation (4.1) for specified $n_1 = n_2 = n_3$, and $\mu_1$ or $\lambda_1$, depending on the response distribution.

2. Generate the hypothetical orders of observation.

3. Calculate $n$ test statistics corresponding to group 1, and $n$ corresponding to group 2; treat the data as if it arrives according to the ordering generated in step (2).

4. Compare the test statistics from step (3) to a critical value, according to equations (3.46) and (3.47), and the values of $n_1 = n_2$ and $\alpha^*$.

5. Record the first observation for which one of equations (3.46) or (3.47) is satisfied, nothing otherwise.

6. Repeat steps $(1) - (5)$ 9999 additional times.

7. Calculate and record $\hat{\alpha}$ (for $\mu_1 = 0$, $\lambda_1 = 1$) or empirical power (for $\mu_1 \neq 0$, $\lambda_1 \neq 1$), along with the average stopping time.

### 4.3.2 Results

Here we present the results of our simulation study for $c = 3$. The raw results may be found in Tables 4.9 to 4.16, while plots of P and AVST are given in Figures 4.1 and 4.2, for all three tests, with $\alpha = 0.05$ and normal data.

Table 4.9: Simulated power (P) and average stopping time (AVST) for Test 1 (Pocock) with $c = 3$ and normally distributed responses. We set $\mu_2 = \mu_3 = 0$ and $\sigma^2 = 1$, while $\mu_1$ is allowed to vary. Critical values are obtained from Table 3.2. For this simulation, $n_1 = n_2 = n_3 = 20$ or 50.

| | | $\alpha = 0.01$ | | $\alpha = 0.05$ | | $\alpha = 0.10$ | |
|---|---|---|---|---|---|---|---|
| $n$ | $\mu_1$ | P | AVST | P | AVST | P | AVST |
| 60 | | | | | | | |
| | 0.0 | 0.0112 | 59.4918 | 0.0421 | 58.1473 | 0.0781 | 56.6818 |
| | 0.1 | 0.0114 | 59.5664 | 0.0522 | 57.8831 | 0.0970 | 56.1211 |
| | 0.2 | 0.0191 | 59.3742 | 0.0821 | 57.2639 | 0.1434 | 55.1444 |
| | 0.3 | 0.0361 | 59.0950 | 0.1420 | 56.1916 | 0.2422 | 52.9045 |
| | 0.4 | 0.0905 | 58.2109 | 0.2505 | 54.1084 | 0.3750 | 50.1662 |
| | 0.5 | 0.1648 | 57.2173 | 0.4063 | 51.0016 | 0.5608 | 45.7705 |
| | 0.6 | 0.2941 | 55.1023 | 0.5829 | 46.9640 | 0.7288 | 40.9650 |
| | 0.7 | 0.4588 | 52.2672 | 0.7503 | 42.6465 | 0.8645 | 36.2469 |
| | 0.8 | 0.6252 | 48.9990 | 0.8648 | 38.6983 | 0.9372 | 32.2200 |
| | 0.9 | 0.7537 | 45.7909 | 0.9340 | 34.7796 | 0.9718 | 29.4000 |
| | 1.0 | 0.8582 | 42.6957 | 0.9734 | 31.8365 | 0.9891 | 26.4426 |
| 150 | | | | | | | |
| | 0.0 | 0.0086 | 148.9492 | 0.0433 | 144.8926 | 0.0824 | 140.5927 |
| | 0.1 | 0.0159 | 148.4534 | 0.0677 | 143.5271 | 0.1184 | 138.4873 |
| | 0.2 | 0.0571 | 146.6917 | 0.1808 | 137.5318 | 0.2812 | 128.2342 |
| | 0.3 | 0.2126 | 139.4573 | 0.4372 | 123.0427 | 0.5726 | 109.8453 |
| | 0.4 | 0.5087 | 124.6682 | 0.7495 | 100.2828 | 0.8467 | 85.6244 |
| | 0.5 | 0.8084 | 103.1930 | 0.9328 | 78.6921 | 0.9679 | 65.0683 |
| | 0.6 | 0.9537 | 83.2035 | 0.9910 | 61.4096 | 0.9964 | 50.5286 |
| | 0.7 | 0.9940 | 68.3289 | 0.9994 | 50.3649 | 0.9999 | 41.7366 |
| | 0.8 | 0.9995 | 58.6947 | 0.9999 | 42.7846 | 1.0000 | 35.8507 |
| | 0.9 | 1.0000 | 51.8400 | 1.0000 | 38.0242 | 1.0000 | 31.6645 |
| | 1.0 | 1.0000 | 46.2481 | 1.0000 | 34.4057 | 1.0000 | 28.8427 |

Table 4.10: Simulated power (P) and average stopping time (AVST) for Test 2 (O'Brien-Fleming) with $c = 3$ and normally distributed responses. We set $\mu_2 = \mu_3 = 0$ and $\sigma^2 = 1$, while $\mu_1$ is allowed to vary. Critical values are obtained from Table 3.3. For this simulation, $n_1 = n_2 = n_3 = 20$ or 50.

| | | $\alpha = 0.01$ | | $\alpha = 0.05$ | | $\alpha = 0.10$ | |
|---|---|---|---|---|---|---|---|
| $n$ | $\mu_1$ | P | AVST | P | AVST | P | AVST |
| 60 | | | | | | | |
| | 0.0 | 0.0078 | 59.9238 | 0.0409 | 59.5271 | 0.0826 | 58.8365 |
| | 0.1 | 0.0148 | 59.8609 | 0.0652 | 59.2237 | 0.1220 | 58.3655 |
| | 0.2 | 0.0450 | 59.6100 | 0.1398 | 58.3835 | 0.2364 | 56.9005 |
| | 0.3 | 0.1150 | 58.9987 | 0.2913 | 56.6210 | 0.4159 | 54.3650 |
| | 0.4 | 0.2360 | 57.8644 | 0.4906 | 53.8017 | 0.6107 | 50.8125 |
| | 0.5 | 0.4144 | 56.0512 | 0.6870 | 50.4219 | 0.8080 | 46.3183 |
| | 0.6 | 0.6161 | 53.2047 | 0.8418 | 46.5305 | 0.9199 | 42.2828 |
| | 0.7 | 0.7776 | 50.3510 | 0.9335 | 42.9881 | 0.9704 | 38.6886 |
| | 0.8 | 0.8925 | 47.4663 | 0.9770 | 40.0687 | 0.9918 | 35.8449 |
| | 0.9 | 0.9516 | 44.8595 | 0.9931 | 37.5970 | 0.9982 | 33.7374 |
| | 1.0 | 0.9798 | 42.7174 | 0.9977 | 35.8021 | 0.9995 | 32.0828 |
| 150 | | | | | | | |
| | 0.0 | 0.0085 | 149.8272 | 0.0469 | 148.6833 | 0.0905 | 146.8945 |
| | 0.1 | 0.0340 | 149.2424 | 0.1147 | 146.8049 | 0.1904 | 143.6319 |
| | 0.2 | 0.1692 | 146.1499 | 0.3522 | 138.8879 | 0.5048 | 131.2216 |
| | 0.3 | 0.4593 | 136.9078 | 0.7096 | 122.3051 | 0.8104 | 112.5175 |
| | 0.4 | 0.7865 | 121.8153 | 0.9237 | 103.6689 | 0.9658 | 92.9724 |
| | 0.5 | 0.9582 | 104.9302 | 0.9911 | 87.9098 | 0.9974 | 78.7036 |
| | 0.6 | 0.9950 | 91.9151 | 0.9993 | 76.7315 | 0.9998 | 68.9431 |
| | 0.7 | 0.9995 | 82.8947 | 1.0000 | 69.0984 | 1.0000 | 61.8037 |
| | 0.8 | 1.0000 | 76.3594 | 1.0000 | 63.5991 | 1.0000 | 56.9427 |
| | 0.9 | 1.0000 | 71.2220 | 1.0000 | 59.4892 | 1.0000 | 53.3318 |
| | 1.0 | 1.0000 | 67.4635 | 1.0000 | 56.3992 | 1.0000 | 50.6889 |

Table 4.11: Simulated power (P) and average stopping time (AVST) for Test 3 with $c = 3$ and normally distributed responses. We set $\mu_2 = \mu_3 = 0$ and $\sigma^2 = 1$, while $\mu_1$ is allowed to vary. Critical values are obtained from Table 3.5. For this simulation, $n_1 = n_2 = n_3 = 10$ or 15.

| | | $\alpha = 0.01$ | | $\alpha = 0.05$ | | $\alpha = 0.10$ | |
|---|---|---|---|---|---|---|---|
| $n$ | $\mu_1$ | P | AVST | P | AVST | P | AVST |
| 30 | | | | | | | |
| | 0.0 | 0.0077 | 29.9740 | 0.0344 | 29.8735 | 0.0807 | 29.5674 |
| | 0.1 | 0.0088 | 29.9719 | 0.0363 | 29.8657 | 0.0908 | 29.5071 |
| | 0.2 | 0.0162 | 29.9472 | 0.0576 | 29.7830 | 0.1322 | 29.2815 |
| | 0.3 | 0.0314 | 29.8941 | 0.0947 | 29.6549 | 0.1865 | 28.9907 |
| | 0.4 | 0.0523 | 29.8268 | 0.1593 | 29.4314 | 0.2732 | 28.4790 |
| | 0.5 | 0.0834 | 29.7287 | 0.2344 | 29.0978 | 0.3755 | 27.8859 |
| | 0.6 | 0.1409 | 29.5350 | 0.3281 | 28.6765 | 0.4845 | 27.0953 |
| | 0.7 | 0.2106 | 29.3107 | 0.4374 | 28.1689 | 0.5994 | 26.2964 |
| | 0.8 | 0.2839 | 29.0399 | 0.5452 | 27.5870 | 0.7061 | 25.4152 |
| | 0.9 | 0.3709 | 28.7496 | 0.6392 | 27.0679 | 0.7884 | 24.5649 |
| | 1.0 | 0.4639 | 28.3809 | 0.7228 | 26.4768 | 0.8531 | 23.8692 |
| 45 | | | | | | | |
| | 0.0 | 0.0038 | 44.9789 | 0.0313 | 44.7739 | 0.0627 | 44.5309 |
| | 0.1 | 0.0057 | 44.9672 | 0.0381 | 44.7418 | 0.0847 | 44.3385 |
| | 0.2 | 0.0104 | 44.9479 | 0.0643 | 44.5523 | 0.1355 | 43.9335 |
| | 0.3 | 0.0266 | 44.8612 | 0.1179 | 44.1838 | 0.2264 | 43.1788 |
| | 0.4 | 0.0580 | 44.7254 | 0.2139 | 43.4716 | 0.3580 | 41.9787 |
| | 0.5 | 0.1146 | 44.4174 | 0.3327 | 42.4979 | 0.4959 | 40.5046 |
| | 0.6 | 0.1981 | 43.9886 | 0.4672 | 41.2406 | 0.6433 | 38.6421 |
| | 0.7 | 0.3015 | 43.3596 | 0.6132 | 39.6171 | 0.7710 | 36.7532 |
| | 0.8 | 0.4292 | 42.5024 | 0.7301 | 38.1086 | 0.8601 | 34.8593 |
| | 0.9 | 0.5501 | 41.6571 | 0.8336 | 36.5983 | 0.9236 | 33.2521 |
| | 1.0 | 0.6663 | 40.7826 | 0.9041 | 35.1223 | 0.9599 | 31.8088 |

Table 4.12: Simulated power (P) and average stopping time (AVST) for Test 3 with $c = 3$ and normally distributed responses. We set $\mu_2 = \mu_3 = 0$ and $\sigma^2 = 1$, while $\mu_1$ is allowed to vary. Critical values are obtained from Table 3.5. For this simulation, $n_1 = n_2 = n_3 = 20$ or 50.

| | | $\alpha = 0.01$ | | $\alpha = 0.05$ | | $\alpha = 0.10$ | |
|---|---|---|---|---|---|---|---|
| $n$ | $\mu_1$ | P | AVST | P | AVST | P | AVST |
| 60 | | | | | | | |
| | 0.0 | 0.0088 | 59.9320 | 0.0442 | 59.5478 | 0.0861 | 59.0302 |
| | 0.1 | 0.0158 | 59.8751 | 0.0640 | 59.3175 | 0.1210 | 58.6446 |
| | 0.2 | 0.0366 | 59.7149 | 0.1152 | 58.7282 | 0.2041 | 57.6249 |
| | 0.3 | 0.0831 | 59.3402 | 0.2219 | 57.6036 | 0.3331 | 55.7869 |
| | 0.4 | 0.1727 | 58.4767 | 0.3662 | 55.7209 | 0.5069 | 53.1087 |
| | 0.5 | 0.3023 | 57.2007 | 0.5492 | 52.9723 | 0.6704 | 49.8642 |
| | 0.6 | 0.4698 | 55.3353 | 0.7078 | 50.0092 | 0.8212 | 46.2371 |
| | 0.7 | 0.6244 | 53.1322 | 0.8331 | 47.0388 | 0.9088 | 42.9753 |
| | 0.8 | 0.7674 | 50.8110 | 0.9199 | 44.0723 | 0.9596 | 40.0907 |
| | 0.9 | 0.8620 | 48.6000 | 0.9637 | 41.5481 | 0.9841 | 37.7370 |
| | 1.0 | 0.9335 | 46.3139 | 0.9865 | 39.5229 | 0.9956 | 35.9282 |
| 150 | | | | | | | |
| | 0.0 | 0.0102 | 149.7985 | 0.0410 | 148.8816 | 0.0880 | 147.3017 |
| | 0.1 | 0.0277 | 149.4389 | 0.0849 | 147.7077 | 0.1549 | 145.1954 |
| | 0.2 | 0.1030 | 147.7213 | 0.2438 | 142.8844 | 0.3604 | 137.4381 |
| | 0.3 | 0.2935 | 142.4214 | 0.5087 | 133.0608 | 0.6578 | 123.0977 |
| | 0.4 | 0.5896 | 132.0983 | 0.7739 | 118.3120 | 0.8767 | 106.8096 |
| | 0.5 | 0.8209 | 118.5655 | 0.9416 | 102.8100 | 0.9711 | 91.8074 |
| | 0.6 | 0.9480 | 105.7963 | 0.9872 | 90.4770 | 0.9952 | 80.4723 |
| | 0.7 | 0.9920 | 94.8924 | 0.9985 | 81.2443 | 1.0000 | 72.0168 |
| | 0.8 | 0.9989 | 86.7986 | 0.9999 | 74.2384 | 1.0000 | 66.4159 |
| | 0.9 | 1.0000 | 80.8194 | 0.9999 | 69.4605 | 1.0000 | 61.9039 |
| | 1.0 | 1.0000 | 76.4510 | 1.0000 | 65.6134 | 1.0000 | 58.3219 |

Table 4.13: Simulated power (P) and average stopping time (AVST) for Test 1 (Pocock) with $c = 3$ and exponentially distributed responses. We set $\lambda_2 = \lambda_3 = 1$ while $\lambda_1$ is allowed to vary. Critical values are obtained from Table 3.2. For this simulation, $n_1 = n_2 = n_3 = 20$ or 50.

| | | $\alpha = 0.01$ | | $\alpha = 0.05$ | | $\alpha = 0.10$ | |
|---|---|---|---|---|---|---|---|
| $n$ | $1/\lambda_1$ | P | AVST | P | AVST | P | AVST |
| 60 | | | | | | | |
| | 1.00 | 0.0111 | 59.5043 | 0.0445 | 58.0785 | 0.0766 | 56.6822 |
| | 1.25 | 0.0231 | 59.2486 | 0.0832 | 57.4152 | 0.1571 | 54.7946 |
| | 1.50 | 0.0821 | 58.3215 | 0.2326 | 54.5203 | 0.3711 | 50.3864 |
| | 1.75 | 0.2176 | 56.4521 | 0.4813 | 49.4895 | 0.6318 | 43.8494 |
| | 2.00 | 0.4123 | 53.1151 | 0.7061 | 44.0520 | 0.8271 | 37.7385 |
| | 2.25 | 0.5939 | 49.8125 | 0.8468 | 39.0657 | 0.9212 | 33.3328 |
| | 2.50 | 0.7341 | 46.6340 | 0.9216 | 35.8426 | 0.9662 | 29.6429 |
| | 2.75 | 0.8266 | 43.7871 | 0.9630 | 32.8091 | 0.9859 | 27.2422 |
| | 3.00 | 0.8869 | 41.5038 | 0.9812 | 30.6726 | 0.9945 | 25.6025 |
| 150 | | | | | | | |
| | 1.00 | 0.0100 | 148.7463 | 0.0463 | 144.6410 | 0.0852 | 140.0826 |
| | 1.25 | 0.0702 | 145.9093 | 0.2115 | 135.5367 | 0.3200 | 126.9630 |
| | 1.50 | 0.4811 | 125.4595 | 0.7279 | 102.0638 | 0.8276 | 87.2194 |
| | 1.75 | 0.8801 | 95.2217 | 0.9689 | 70.7442 | 0.9852 | 58.4110 |
| | 2.00 | 0.9866 | 73.1737 | 0.9979 | 53.5627 | 0.9996 | 44.2208 |
| | 2.25 | 0.9989 | 60.7293 | 0.9999 | 44.5231 | 1.0000 | 36.9201 |
| | 2.50 | 1.0000 | 52.7399 | 1.0000 | 38.8416 | 1.0000 | 32.4302 |
| | 2.75 | 1.0000 | 48.1233 | 1.0000 | 35.5908 | 1.0000 | 29.7305 |
| | 3.00 | 1.0000 | 44.7669 | 1.0000 | 32.9585 | 1.0000 | 27.8222 |

Table 4.14: Simulated power (P) and average stopping time (AVST) for Test 2 (O'Brien-Fleming) with $c = 3$ and exponentially distributed responses. We set $\lambda_2 = \lambda_3 = 1$ while $\lambda_1$ is allowed to vary. Critical values are obtained from Table 3.3. For this simulation, $n_1 = n_2 = n_3 = 20$ or 50.

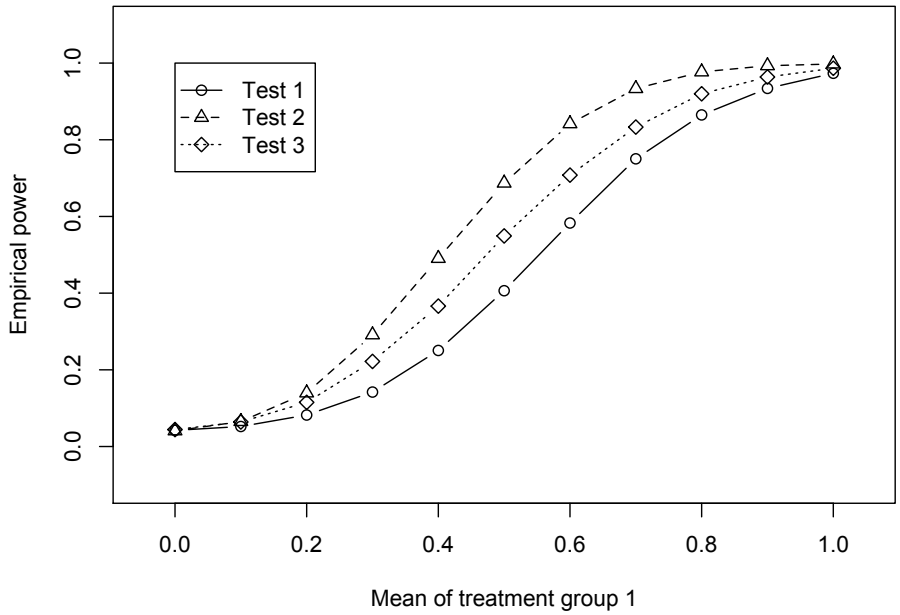
| | | $\alpha = 0.01$ | | $\alpha = 0.05$ | | $\alpha = 0.10$ | |
|---|---|---|---|---|---|---|---|
| $n$ | $1/\lambda_1$ | P | AVST | P | AVST | P | AVST |
| 60 | | | | | | | |
| | 1.00 | 0.0073 | 59.9313 | 0.0428 | 59.4659 | 0.0842 | 58.8269 |
| | 1.25 | 0.0466 | 59.6065 | 0.1565 | 58.1682 | 0.2561 | 56.5803 |
| | 1.50 | 0.2300 | 58.0235 | 0.4756 | 54.1139 | 0.5953 | 51.0572 |
| | 1.75 | 0.4969 | 54.8683 | 0.7484 | 49.0646 | 0.8483 | 44.7725 |
| | 2.00 | 0.7222 | 51.3871 | 0.9082 | 44.1643 | 0.9542 | 40.0286 |
| | 2.25 | 0.8662 | 48.1642 | 0.9691 | 40.5743 | 0.9877 | 36.5112 |
| | 2.50 | 0.9397 | 45.5588 | 0.9909 | 38.2797 | 0.9968 | 34.3952 |
| | 2.75 | 0.9708 | 43.5415 | 0.9971 | 36.3442 | 0.9986 | 32.6027 |
| | 3.00 | 0.9867 | 42.0094 | 0.9993 | 35.1142 | 0.9996 | 31.6309 |
| 150 | | | | | | | |
| | 1.00 | 0.0093 | 149.7966 | 0.0409 | 148.8153 | 0.0918 | 146.8402 |
| | 1.25 | 0.1910 | 145.3541 | 0.4153 | 136.3195 | 0.5453 | 128.8503 |
| | 1.50 | 0.7628 | 122.9867 | 0.9142 | 105.0798 | 0.9570 | 94.6540 |
| | 1.75 | 0.9776 | 100.4006 | 0.9950 | 83.2943 | 0.9987 | 74.9540 |
| | 2.00 | 0.9989 | 86.0597 | 0.9999 | 71.5144 | 1.0000 | 64.2597 |
| | 2.25 | 1.0000 | 77.4560 | 1.0000 | 64.7975 | 1.0000 | 58.1422 |
| | 2.50 | 1.0000 | 72.4695 | 1.0000 | 60.5805 | 1.0000 | 54.4879 |
| | 2.75 | 1.0000 | 68.6527 | 1.0000 | 57.4353 | 1.0000 | 51.5617 |
| | 3.00 | 1.0000 | 66.0040 | 1.0000 | 55.2850 | 1.0000 | 49.6979 |

Table 4.15: Simulated power (P) and average stopping time (AVST) for Test 3 with $c = 3$ and exponentially distributed responses. We set $\lambda_2 = \lambda_3 = 1$ while $\lambda_1$ is allowed to vary. Critical values are obtained from Table 3.5. For this simulation, $n_1 = n_2 = n_3 = 10$ or 15.

| | | $\alpha = 0.01$ | | $\alpha = 0.05$ | | $\alpha = 0.10$ | |
|---|---|---|---|---|---|---|---|
| $n$ | $1/\lambda_1$ | P | AVST | P | AVST | P | AVST |
| 30 | | | | | | | |
| | 1.00 | 0.0064 | 29.9782 | 0.0352 | 29.8744 | 0.0790 | 29.5650 |
| | 1.25 | 0.0191 | 29.9324 | 0.0585 | 29.7901 | 0.1260 | 29.3175 |
| | 1.50 | 0.0495 | 29.8402 | 0.1366 | 29.4777 | 0.2575 | 28.5697 |
| | 1.75 | 0.0885 | 29.7097 | 0.2545 | 29.0206 | 0.3942 | 27.7066 |
| | 2.00 | 0.1560 | 29.4615 | 0.3794 | 28.4946 | 0.5357 | 26.7563 |
| | 2.25 | 0.2315 | 29.2277 | 0.4860 | 27.9472 | 0.6535 | 25.8331 |
| | 2.50 | 0.3078 | 28.9495 | 0.5763 | 27.4712 | 0.7370 | 25.1051 |
| | 2.75 | 0.3847 | 28.6756 | 0.6640 | 26.9171 | 0.8058 | 24.3598 |
| | 3.00 | 0.4452 | 28.4433 | 0.7178 | 26.5123 | 0.8479 | 23.8833 |
| 45 | | | | | | | |
| | 1.00 | 0.0028 | 44.9868 | 0.0271 | 44.8124 | 0.0649 | 44.5025 |
| | 1.25 | 0.0130 | 44.9324 | 0.0683 | 44.5180 | 0.1500 | 43.8320 |
| | 1.50 | 0.0504 | 44.7520 | 0.1885 | 43.6366 | 0.3227 | 42.3010 |
| | 1.75 | 0.1295 | 44.3222 | 0.3577 | 42.2533 | 0.5276 | 40.1540 |
| | 2.00 | 0.2393 | 43.7269 | 0.5380 | 40.5536 | 0.7007 | 37.9380 |
| | 2.25 | 0.3480 | 43.0296 | 0.6729 | 38.9604 | 0.8142 | 35.9111 |
| | 2.50 | 0.4719 | 42.2252 | 0.7815 | 37.4621 | 0.8888 | 34.2313 |
| | 2.75 | 0.5686 | 41.5116 | 0.8486 | 36.2909 | 0.9363 | 32.9321 |
| | 3.00 | 0.6587 | 40.7824 | 0.9038 | 35.2312 | 0.9592 | 31.8807 |

Table 4.16: Simulated power (P) and average stopping time (AVST) for Test 3 with $c = 3$ and exponentially distributed responses. We set $\lambda_2 = \lambda_3 = 1$ while $\lambda_1$ is allowed to vary. Critical values are obtained from Table 3.5. For this simulation, $n_1 = n_2 = n_3 = 20$ or 50.

| | | $\alpha = 0.01$ | | $\alpha = 0.05$ | | $\alpha = 0.10$ | |
|---|---|---|---|---|---|---|---|
| $n$ | $1/\lambda_1$ | P | AVST | P | AVST | P | AVST |
| 60 | | | | | | | |
| | 1.00 | 0.0088 | 59.9384 | 0.0455 | 59.5051 | 0.0915 | 58.9156 |
| | 1.25 | 0.0377 | 59.7028 | 0.1306 | 58.6151 | 0.2062 | 57.5873 |
| | 1.50 | 0.1470 | 58.8247 | 0.3327 | 56.1877 | 0.4772 | 53.7623 |
| | 1.75 | 0.3312 | 56.8978 | 0.5867 | 52.3068 | 0.7084 | 49.2820 |
| | 2.00 | 0.5288 | 54.6244 | 0.7572 | 48.8314 | 0.8573 | 45.0540 |
| | 2.25 | 0.7022 | 52.0999 | 0.8803 | 45.4285 | 0.9387 | 41.6053 |
| | 2.50 | 0.8067 | 50.0038 | 0.9439 | 43.1553 | 0.9733 | 39.1647 |
| | 2.75 | 0.8796 | 48.1477 | 0.9724 | 41.0395 | 0.9891 | 37.2456 |
| | 3.00 | 0.9343 | 46.5193 | 0.9882 | 39.4830 | 0.9946 | 36.0308 |
| 150 | | | | | | | |
| | 1.00 | 0.0094 | 149.7916 | 0.0368 | 148.9459 | 0.0886 | 147.3095 |
| | 1.25 | 0.1262 | 147.1535 | 0.2666 | 142.1740 | 0.3933 | 136.0882 |
| | 1.50 | 0.5257 | 134.5031 | 0.7394 | 121.6317 | 0.8489 | 109.9476 |
| | 1.75 | 0.8637 | 116.2698 | 0.9537 | 100.4810 | 0.9829 | 89.3358 |
| | 2.00 | 0.9780 | 101.2626 | 0.9952 | 86.8219 | 0.9986 | 76.9177 |
| | 2.25 | 0.9973 | 91.1417 | 0.9997 | 77.8712 | 1.0000 | 69.4758 |
| | 2.50 | 0.9998 | 80.2413 | 1.0000 | 72.7223 | 1.0000 | 64.3905 |
| | 2.75 | 1.0000 | 80.2273 | 1.0000 | 68.6684 | 1.0000 | 61.1518 |
| | 3.00 | 1.0000 | 77.0099 | 1.0000 | 65.8899 | 1.0000 | 58.5831 |



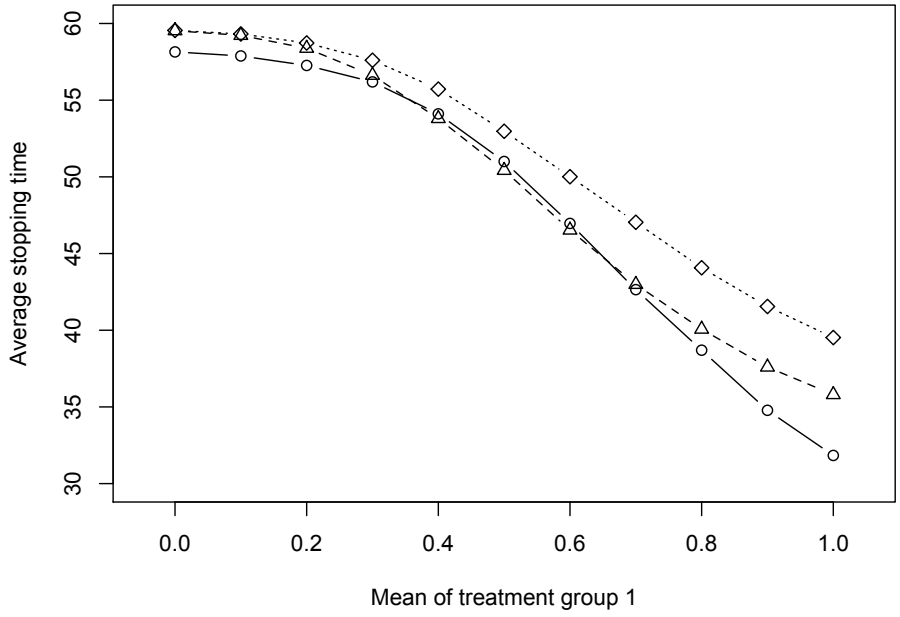
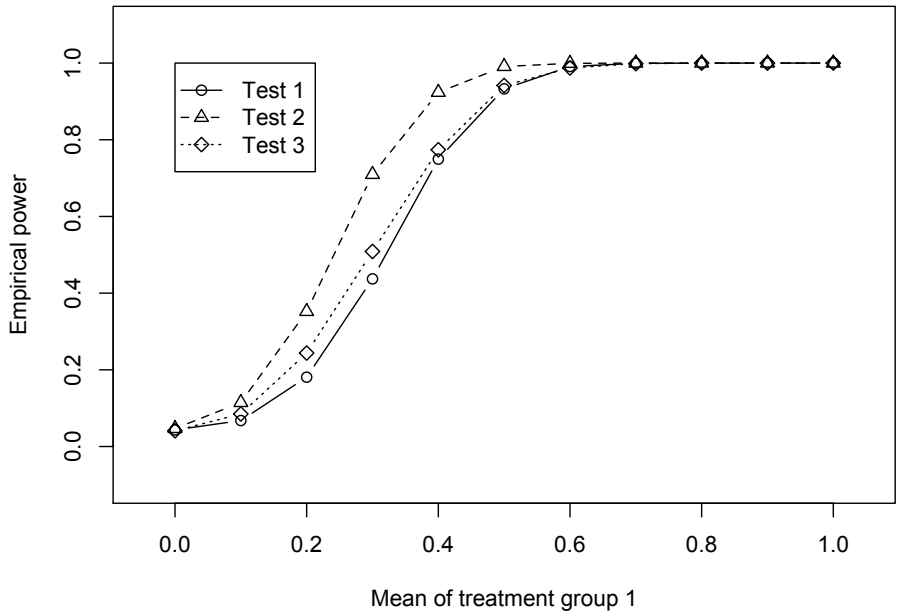




Figure 4.1: Plots of empirical power (P) and average stopping time (AVST) for Tests 1, 2, and 3, where $c = 3$, $n_1 = n_2 = n_3 = 20$, $\alpha = 0.05$, and where the response distribution is normal.



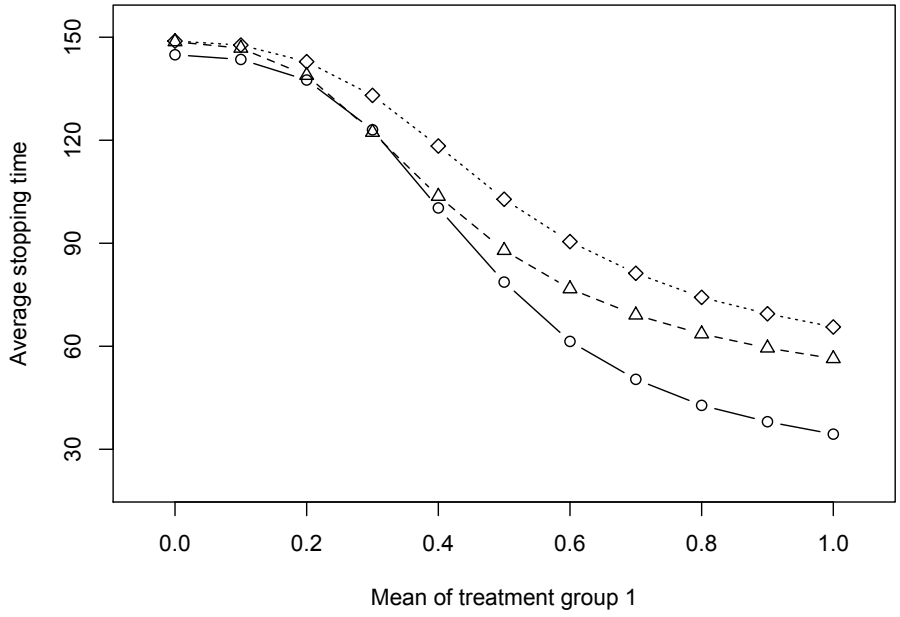




Figure 4.2: Plots of empirical power (P) and average stopping time (AVST) for Tests 1, 2, and 3, where $c = 3$, $n_1 = n_2 = n_3 = 50$, $\alpha = 0.05$, and where the response distribution is normal.

## 4.4 Discussion

Overall, we observed that all three tests perform similarly for both normal and exponential responses. Power is higher when $c = 3$ compared to $c = 2$, for all tests and both response types. This may be attributed to the fact that the total sample size, $n$, is larger for simulations involving three groups. Rather than matching the total sample size, we chose to maintain the same group sizes between all simulations.

We also made some test-specific observations:

**Test 1**

1. Test 1 is reasonably powered for $n_1 = 20$ and well-powered for $n_1 = 50$.
2. Test 1 has earlier average stopping times than both Test 2 and Test 3, for large treatment differences.

**Test 2**

1. Test 2 appears to have higher power than Test 1 and Test 3, for small treatment differences.
2. We observed average stopping times for Test 2 that were larger than those of Test 1, similar to those of Test 3 for $c = 2$, and smaller than those of Test 3 for $c = 3$.

**Test 3**

1. The performance of Test 3 is acceptable for $n_1 = 10, 15$, with decent power and somewhat lengthy average stopping times.
2. Overall, Test 3 has the longest average stopping times; this is to be expected however, as the test is nonparametric, and suffers from some degree of information loss.

# Chapter 5

# Application

In this section, we apply Tests 1, 2, and 3 to data from an orthodontic clinical trial. The objective of the trial was to determine whether the skeletal and dental effects of maxillary expansion are the same between subjects, and for different treatment groups. A total of $n = 62$ patients were recruited from the Graduate Orthodontic Clinic patient pool. They were randomly assigned to one of three treatment groups: Hyrax (1), bone-anchored expander (2), or control (3). Two patients were excluded from this analysis due to missing data, leaving group sample sizes of $n_1 = 21$, $n_2 = 20$, and $n_3 = 19$.

Three-dimensional volumetric scans were obtained for each patient via cone-beam computerized tomography (CBCT), a maximum of three times over the course of the trial. From these scans, three-dimensional landmarks, descriptors of shape, were derived. The response variables are taken to be the Euclidean distance between pairs of landmarks. We will restrict our analysis here to the response derived from comparing landmarks 1 and 6. Finally, we refer to the three measures of our response as T1, T2, and T3.

For all three tests, we use $\alpha = 0.05$ with corresponding $\alpha^* = 0.0253$. For Test 1 and both groups 1 and 2 we use boundaries corresponding to $n_k = 20$. These may be found in Table 3.2 of Chapter 3. The boundaries for Test 2 on the other hand, which are not dependent on sample size, are found in Table 3.3. As Test 3 is exact, we calculate the appropriate boundaries for our specific values of $n_1$ and $n_2$. We arrive at the same boundary value for both groups.

The critical values for all three tests may be found in Table 5.1.

Table 5.1: Critical values for Test 1, Test 2, and Test 3, for the skeletal and dental effects study. For Test 3, the exact significance levels are provided in brackets.

| | Group 1 | Group 2 |
|---|---|---|
| Test 1 | 3.18 | 3.18 |
| Test 2 | 2.49 | 2.49 |
| Test 3 | 11 (0.0467) | 11 (0.0665) |

The test statistics for Tests 1, 2, and 3, are presented in Tables 5.2 and 5.3 for analyses 1-30 and 31-60, respectively. Plots of the sample paths of each test statistic along with the corresponding critical values are given in Figure 5.1. For Test 1, $H_0$ is rejected at analysis 31, for Test 2 we reject at analysis 36, and for Test 3 we reject at analysis 50. Again, all tests use an overall significance level of $\alpha = 0.05$.

Table 5.2: Test statistics at analyses 1-30, to be compared to a critical value, for landmarks 1 and 6, and Tests 1, 2, and 3.

| | Test 1 | | Test 2 | | Test 3 | |
|---|---|---|---|---|---|---|
| Analysis | Group 1 | Group 2 | Group 1 | Group 2 | Group 1 | Group 2 |
| 1 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0 | 0 |
| 2 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0 | 0 |
| 3 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0 | 0 |
| 4 | 0.9440 | 0.0000 | 0.2060 | 0.0000 | 1 | 0 |
| 5 | 0.9316 | 0.0000 | 0.2033 | 0.0000 | 1 | 0 |
| 6 | 0.9556 | 0.9172 | 0.2085 | 0.2051 | 1 | 1 |
| 7 | 1.0171 | 0.9763 | 0.2219 | 0.2183 | 1 | 1 |
| 8 | 1.0786 | 0.6448 | 0.2354 | 0.2039 | 1 | 0 |
| 9 | 0.9939 | 0.5942 | 0.2169 | 0.1879 | 1 | 0 |
| 10 | 1.0327 | 0.6173 | 0.2253 | 0.1952 | 1 | 0 |
| 11 | 1.0488 | 0.6270 | 0.2289 | 0.1983 | 1 | 0 |
| 12 | 1.0121 | 1.4784 | 0.2209 | 0.5726 | 1 | 1 |
| 13 | 1.2835 | 1.5096 | 0.3961 | 0.5847 | 2 | 1 |
| 14 | 1.3148 | 1.5436 | 0.4058 | 0.6903 | 2 | 2 |
| 15 | 1.2948 | 2.0731 | 0.3996 | 1.0365 | 2 | 3 |
| 16 | 2.0288 | 2.0160 | 0.7668 | 1.0080 | 3 | 3 |
| 17 | 2.4371 | 2.0422 | 1.0636 | 1.0211 | 4 | 3 |
| 18 | 2.4878 | 2.0847 | 1.0858 | 1.0423 | 4 | 3 |
| 19 | 2.5229 | 2.3356 | 1.1011 | 1.2793 | 4 | 4 |
| 20 | 2.4971 | 2.3710 | 1.2185 | 1.2987 | 5 | 4 |
| 21 | 2.4760 | 1.9474 | 1.2082 | 1.1521 | 5 | 3 |
| 22 | 2.6646 | 1.9819 | 1.4243 | 1.1725 | 6 | 3 |
| 23 | 2.6940 | 2.2357 | 1.4400 | 1.4140 | 6 | 4 |
| 24 | 2.7244 | 2.2078 | 1.4562 | 1.4811 | 6 | 5 |
| 25 | 2.7639 | 2.2399 | 1.4774 | 1.5026 | 6 | 5 |
| 26 | 2.7464 | 2.6540 | 1.4680 | 1.8767 | 6 | 6 |
| 27 | 2.9639 | 2.5809 | 1.7112 | 1.8250 | 7 | 6 |
| 28 | 3.0562 | 2.6040 | 1.8863 | 1.8413 | 8 | 6 |
| 29 | 3.0729 | 2.8646 | 1.8966 | 2.1244 | 8 | 7 |
| 30 | 2.3583 | 2.7857 | 1.5439 | 2.0659 | 7 | 7 |

Table 5.3: Test statistics at analyses 31-60, to be compared to a critical value, for landmarks 1 and 6, and Tests 1, 2, and 3.

| | Test 1 | | Test 2 | | Test 3 | |
|---|---|---|---|---|---|---|
| Analysis | Group 1 | Group 2 | Group 1 | Group 2 | Group 1 | Group 2 |
| 31 | 2.3434 | 3.1845 | 1.5341 | 2.4667 | 7 | 8 |
| 32 | 2.3252 | 3.1598 | 1.5222 | 2.4475 | 7 | 8 |
| 33 | 2.3036 | 3.1304 | 1.5081 | 2.4248 | 7 | 8 |
| 34 | 2.3503 | 3.1938 | 1.5386 | 2.4739 | 7 | 8 |
| 35 | 1.8957 | 3.1729 | 1.3082 | 2.4577 | 6 | 8 |
| 36 | 1.8910 | 3.3935 | 1.3049 | 2.7359 | 6 | 9 |
| 37 | 1.9176 | 3.4412 | 1.3233 | 2.7744 | 6 | 9 |
| 38 | 1.9287 | 3.5856 | 1.3309 | 2.9999 | 6 | 10 |
| 39 | 2.1948 | 3.5934 | 1.5885 | 3.0065 | 7 | 10 |
| 40 | 2.0415 | 3.6133 | 1.5432 | 3.0231 | 6 | 10 |
| 41 | 1.9792 | 3.6423 | 1.5572 | 3.0474 | 7 | 10 |
| 42 | 1.9953 | 3.6719 | 1.5699 | 3.0721 | 7 | 10 |
| 43 | 2.4412 | 3.6534 | 1.9933 | 3.0566 | 8 | 10 |
| 44 | 2.4571 | 3.5504 | 2.0063 | 3.0747 | 8 | 9 |
| 45 | 2.8425 | 3.5399 | 2.4024 | 3.0657 | 9 | 9 |
| 46 | 2.7669 | 3.4457 | 2.3385 | 2.9841 | 9 | 9 |
| 47 | 2.7815 | 3.5632 | 2.3508 | 3.1870 | 9 | 10 |
| 48 | 2.7488 | 3.0925 | 2.3232 | 2.8511 | 9 | 9 |
| 49 | 3.0831 | 3.0798 | 2.6912 | 2.8394 | 10 | 9 |
| 50 | 3.3172 | 3.0867 | 2.9846 | 2.8458 | 11 | 9 |
| 51 | 3.3405 | 3.1084 | 3.0056 | 2.8658 | 11 | 9 |
| 52 | 3.3401 | 3.1080 | 3.0052 | 2.8654 | 11 | 9 |
| 53 | 3.3156 | 3.0852 | 2.9832 | 2.8444 | 11 | 9 |
| 54 | 3.3290 | 3.2085 | 2.9952 | 3.0438 | 11 | 10 |
| 55 | 3.5382 | 3.2095 | 3.2758 | 3.0448 | 12 | 10 |
| 56 | 3.5564 | 3.2554 | 3.2925 | 3.1730 | 12 | 11 |
| 57 | 3.5695 | 3.3836 | 3.3047 | 3.3836 | 12 | 12 |
| 58 | 3.6868 | 3.3961 | 3.5069 | 3.3961 | 13 | 12 |
| 59 | 3.8380 | 3.4105 | 3.7455 | 3.4105 | 14 | 12 |
| 60 | 4.0777 | 3.4083 | 4.0777 | 3.4083 | 15 | 12 |

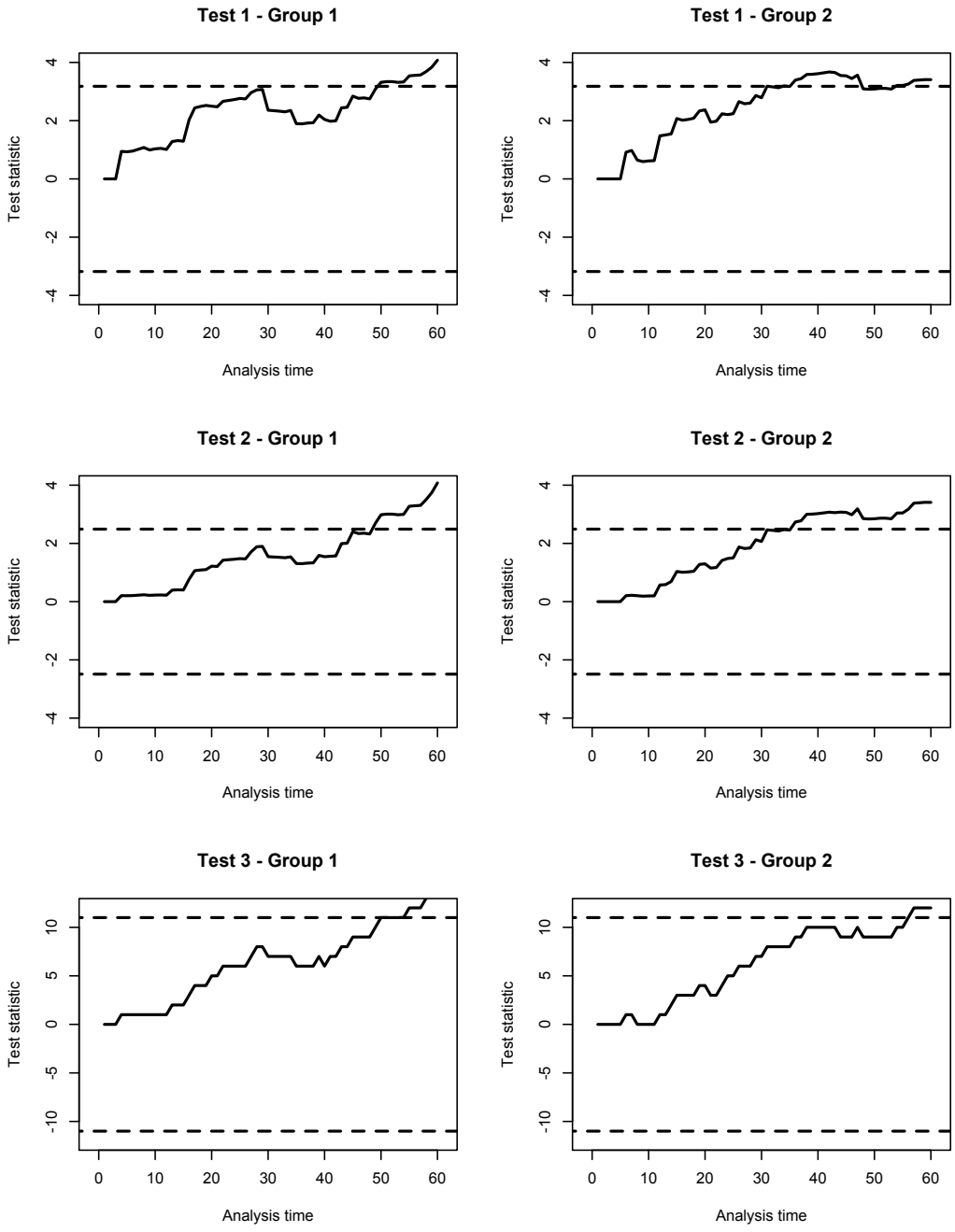



Figure 5.1: Plots of the sample paths of the test statistics for Tests 1, 2, and 3, along with the corresponding critical values, for landmarks 1 and 6.

# Chapter 6

# Summary and conclusions

This thesis has examined the application of sequential methods to longitudinal data. We developed several new, nonparametric, fully sequential monitoring procedures to be used for the comparison of two or more groups. Performance of the new procedures was assessed via a Monte Carlo simulation study. Finally, the procedures were applied to data from an orthodontic clinical trial, for illustrative purposes.

The following are our recommendations for application of the new procedures:

- Test 1 should be used if early stopping is of the utmost importance.
- Test 2 should be used if high power is paramount.
- Test 3 should be used when group sample sizes are small, where $n_k < 20$.

Given the results of this thesis, we believe that there is potential for future research in the following directions:

- Performance of the new procedures could be examined for scenarios in which there are more than three repeated measures, and more than three treatment groups, as well as for various other response distributions.
- Theory could be developed for the pairwise comparison of treatment groups, to be applied when the null hypothesis is rejected.

- The procedure could be revised to accommodate complicated forms of missing data.
- The procedure could be compared to some of the longitudinal group sequential methods.
- Finally, the procedure could be adapted to accommodate multiple outcome response variables.

# Bibliography


[1] P. Armitage, I. M. Stratton, and H. V. Worthington. Repeated significance tests for clinical trials with a fixed number of patients and variable follow-up. *Biometrics*, 41(2):353–359, 1985.

[2] P. Bogowicz, C. Flores-Mir, P. W. Major, and G. Heo. Sequential analysis applied to clinical trials in dentistry: a systematic review. *Evidence-Based Dentistry*, 9(2):55–60, 2008.

[3] H. D. Brunk. *An Introduction to Mathematical Statistics*. Blaisdell: New York, 2nd edition, 1965.

[4] B. Cerutti, T. Zoubeidi, and D. M. Reboussin. Adaptive group sequential procedures for multiple comparison of treatments. *Communications in Statistics - Theory and Methods*, 31(5):781–793, 2002.

[5] E. Csáki. Some applications of the classical formula on ruin probabilities. *Journal of Statistical Planning and Inference*, 14:35–42, 1986.

[6] C. S. Davis. *Statistical Methods for the Analysis of Repeated Measurements*. Springer-Verlag: New York, 1st edition, 2002.

[7] P. J. Diggle, P. Heagerty, K. Liang, and S. L. Zeger. *Analysis of Longitudinal Data*. Oxford University Press: New York, 2nd edition, 2002.

[8] G. M. Fitzmaurice, N. M. Laird, and J. H. Ware. *Applied Longitudinal Analysis*. Wiley-Interscience: Hoboken, 1st edition, 2004.

[9] S. Galbraith and I. C. Marschner. Interim analysis of continuous long-term endpoints in clinical trials with longitudinal outcomes. *Statistics in Medicine*, 22:1787–1805, 2003.

[10] S. J. Gange and D. L. DeMets. Sequential monitoring of clinical trials with correlated responses. *Biometrika*, 83(1):157–167, 1996.

[11] D. N. Geary. Sequential testing in clinical trials with repeated measurements. *Biometrika*, 75(2):311–318, 1988.

[12] B. K. Ghosh and P. K. Sen. *Handbook of Sequential Analysis*. Marcel Dekker: New York, 1st edition, 1991.

[13] E. Gombay. Testing for change-points with rank and sign statistics. *Statistics and probability letters*, 20:49–55, 1994.

[14] E. Gombay. Nonparametric truncated sequential change-point detection. *Statistics & Decisions*, 13:71–82, 1995.

[15] E. Gombay. The weighted sequential likelihood ratio. *Canadian Journal of Statistics*, 24(2):229–239, 1996.

[16] E. Gombay. Parametric sequential tests in the presence of nuisance parameters. *Theory of Stochastic Processes*, 8(24):106–119, 2002.

[17] E. Gombay. Sequential testing of composite hypotheses. *Limit Theorems in Probability and Statistics: Fourth Hungarian Colloquium on Limit Theorems in Probability and Statistics, Balatonlelle, Hungary, June 28-July 2, 1999, Volume* II, 2:107–125, 2002.

[18] E. Gombay. Weighted logrank statistics in sequential tests. *Sequential Analysis*, 27(1):97–104, 2008.

[19] E. Gombay and D. Serban. Sequential comparison of $d$ populations and related tables. *Communications in Statistics - Simulation and Computation*, 36:55–72, 2007.

[20] C. Jennison and B. W. Turnbull. Group-sequential analysis incorporating covariate information. *Journal of the American Statistical Association*, 92(440):1330–1341, 1997.

[21] C. Jennison and B. W. Turnbull. *Group Sequential Methods with Applications to Clinical Trials*. Chapman and Hall-CRC: Boca Raton, 1st edition, 2000.

[22] R. A. Johnson and D. W. Wichern. *Applied Multivariate Statistical Analysis*. Prentice-Hall: New Jersey, 2nd edition, 1988.

[23] J. M. Kittelson, K. Sharples, and S. S. Emerson. Group sequential clinical trials for longitudinal data with analyses using summary statistics. *Statistics in Medicine*, 24:2457–2475, 2005.

[24] J. M. Lachin. Group sequential monitoring of distribution-free analyses of repeated measures. *Statistics in Medicine*, 16:653–668, 1997.

[25] J. M. Lachin, S. W. Greenhouse, and O. M. Bautista. Group sequential large sample $T^2$-like $\chi^2$ tests for multivariate observations. *Statistics in Medicine*, 22:3357–3368, 2003.

[26] K. K. G. Lan and D. L. DeMets. Discrete sequential boundaries for clinical trials. *Biometrika*, 70(3):659–663, 1983.

[27] J. W. Lee and D. L. DeMets. Sequential comparison of changes with repeated measurements data. *Journal of the American Statistical Association*, 86(415):757–762, 1991.

[28] J. W. Lee and D. L. DeMets. Sequential rank tests with repeated measurements in clinical trials. *Journal of the American Statistical Association*, 87(417):136–142, 1992.

[29] S. J. Lee, K. Kim, and A. A. Tsiatis. Repeated significance testing in longitudinal clinical trials. *Biometrika*, 83(4):779–789, 1996.

[30] P. C. O'Brien and T. R. Fleming. A multiple testing procedure for clinical trials. *Biometrics*, 35:549–556, 1979.

[31] C. R. Palmer and W. F. Rosenberger. Ethics and practice: Alternative designs for phase III randomized clinical trials. *Controlled Clinical Trials*, 20:172–186, 1999.

[32] E. A. Parent, Jr. *Sequential ranking procedures*. PhD thesis, Statistics, Stanford University, 1965.

[33] S. J. Pocock. Group sequential methods in the design and analysis of clinical trials. *Biometrika*, 64:191–199, 1977.

[34] M. L. Puri and P. K. Sen. *Nonparametric Methods in Multivariate Analysis*. Wiley: New York, 1st edition, 1971.

[35] S. M. Ross. *Introduction to Probability Models*. Academic Press: San Diego, 8th edition, 2003.

[36] D. O. Scharfstein and A. A. Tsiatis. The use of simulation and bootstrap in information-based group sequential studies. *Statistics in Medicine*, 17:75–87, 1998.

[37] D. O. Scharfstein, A. A. Tsiatis, and J. M. Robins. Semiparametric efficiency and its implication on the design and analysis of group-sequential studies. *Journal of the American Statistical Association*, 92(440):1342–1350, 1997.

[38] V. Sébille and E. Bellissant. Sequential methods and group sequential designs for comparative clinical trials. *Fundamental and Clinical Pharmacology*, 17:505–516, 2003.

[39] E. Slud and L. J. Wei. Two-sample repeated significance tests based on the modified wilcoxon statistic. *Journal of the American Statistical Association*, 77(380):862–868, 1982.

[40] B. Spiessens, E. Lesaffre, and G. Verbeke. An overview of group sequential methods in longitudinal clinical trials. *Statistical Methods in Medical Research*, 9:497–515, 2000.

[41] J. Q. Su and J. M. Lachin. Group sequential distribution-free methods for the analysis of multivariate observations. *Biometrics*, 48(4):1033–1042, 1992.

[42] S. Todd, A. Whitehead, N. Stallard, and J. Whitehead. Interim analyses and sequential designs in phase III studies. *British Journal of Clinical Pharmacology*, 51:394–399, 2001.

[43] J. F. Troendle, A. Liu, C. Wu, and K. F. Yu. Sequential testing for efficacy in clinical trials with non-transient effects. *Statistics in Medicine*, 24:3239–3250, 2005.

[44] L. J. Vostrikova. Detection of a "disorder" in a wiener process. *Theory of Probability and Its Applications*, 26:356–362, 1981.

[45] K. Wegscheider. Advantages and disadvantages of sequential designs. *Herzschrittmachertherapie und Elektrophysiologie*, 9:248–254, 1998.

[46] L. J. Wei, J. Q. Su, and J. M. Lachin. Interim analyses with repeated measurements in a sequential clinical trial. *Biometrika*, 77(2):359–364, 1990.

[47] M. C. Wu and K. K. G. Lan. Sequential monitoring for comparison of changes in a response variable in clinical studies. *Biometrics*, 48(3):765–779, 1992.

# Appendix A

# Examples

**Example A.0.1 (Sequential ranking)** *Suppose that we have received five observations on some random process:*

| Time | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| Observation | 0.1 | 0.5 | 0.2 | 0.7 | 0.8 |

*where hypothetical times are given to indicate the order in which the data were received. The process of sequentially ranking the data would proceed as follows:*

| Time | | | | | |
|---|---|---|---|---|---|
| 1 | 0.1 | | | | |
| | $\downarrow$ | | | | |
| | **1** | | | | |
| 2 | 0.1 | 0.5 | | | |
| | $\vdots$ | $\downarrow$ | | | |
| | **1** | **2** | | | |
| 3 | 0.1 | 0.5 | 0.2 | | |
| | $\vdots$ | $\vdots$ | $\downarrow$ | | |
| | **1** | **2** | **2** | | |
| 4 | 0.1 | 0.5 | 0.2 | 0.7 | |
| | $\vdots$ | $\vdots$ | $\vdots$ | $\downarrow$ | |
| | **1** | **2** | **2** | **4** | |
| 5 | 0.1 | 0.5 | 0.2 | 0.7 | 0.8 |
| | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\downarrow$ |
| | **1** | **2** | **2** | **4** | **5** |

**Example A.0.2 (Sequential ranking)** *Suppose that we have received five observations on another random process:*

| Time | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| Observation | 0.5 | 0.3 | 1.2 | 0.1 | 0.9 |

*where once again hypothetical times are given to indicate the order in which the data were received. The process of sequentially ranking the data would proceed as follows:*

| Time | | | | | |
|---|---|---|---|---|---|
| 1 | 0.5 | | | | |
| | $\downarrow$ | | | | |
| | **1** | | | | |
| 2 | 0.5 | 0.3 | | | |
| | $\vdots$ | $\downarrow$ | | | |
| | **1** | **1** | | | |
| 3 | 0.5 | 0.3 | 1.2 | | |
| | $\vdots$ | $\vdots$ | $\downarrow$ | | |
| | **1** | **1** | **3** | | |
| 4 | 0.5 | 0.3 | 1.2 | 0.1 | |
| | $\vdots$ | $\vdots$ | $\vdots$ | $\downarrow$ | |
| | **1** | **1** | **3** | **1** | |
| 5 | 0.5 | 0.3 | 1.2 | 0.1 | 0.9 |
| | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\downarrow$ |
| | **1** | **1** | **3** | **1** | **4** |

**Example A.0.3 (Regular ranking)** *Suppose that we have received the same five observations as in Example A.0.2:*

| Time | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| Observation | 0.5 | 0.3 | 1.2 | 0.1 | 0.9 |

*The process of ranking the data by the usual method would proceed as follows:*

| Time | | | | | |
|---|---|---|---|---|---|
| 1 | 0.5 | | | | |
| | ↓ | | | | |
| | **1** | | | | |
| 2 | 0.5 | 0.3 | | | |
| | ↓ | ↓ | | | |
| | **2** | **1** | | | |
| 3 | 0.5 | 0.3 | 1.2 | | |
| | ↓ | ↓ | ↓ | | |
| | **2** | **1** | **3** | | |
| 4 | 0.5 | 0.3 | 1.2 | 0.1 | |
| | ↓ | ↓ | ↓ | ↓ | |
| | **3** | **2** | **4** | **1** | |
| 5 | 0.5 | 0.3 | 1.2 | 0.1 | 0.9 |
| | ↓ | ↓ | ↓ | ↓ | ↓ |
| | **3** | **2** | **5** | **1** | **4** |

*Notice how the ranks found by the regular method differ from sequential ranks in that they do not generally stay the same after each new observation is received.*

**Example A.0.4 (Independence of sequential ranks)** *Consider the ranking of three i.i.d. observations, similar to what would be done in our testing framework, under the null hypothesis of no distributional difference. Denote these observations by $y_1$, $y_2$, and $y_3$, and assume that they have arrived in some particular order, without loss of generality, $y_1$ first, $y_2$ second, and $y_3$ third. In keeping with the notation of Chapter 3, we denote the sequential ranks of $y_1$, $y_2$, and $y_3$, by $R_1$, $R_2$, and $R_3$, respectively. The corresponding parameters are $m_1 = 1$, $m_2 = 2$, and $m_3 = 3$. Under the assumption that the data are i.i.d., we have six possibilities for their sequential ranking:*

| $R_1$ | $R_2$ | $R_3$ |
|---|---|---|
| 1 | 1 | 1 |
| 1 | 2 | 1 |
| 1 | 1 | 2 |
| 1 | 2 | 2 |
| 1 | 1 | 3 |
| 1 | 2 | 3 |

*The i.i.d. assumption implies that each of these rankings is equally likely to occur, in other words that*

$$\begin{aligned}
P\{R_1 = 1, R_2 = 1, R_3 = 1\} &= \frac{1}{6} \\
P\{R_1 = 1, R_2 = 2, R_3 = 1\} &= \frac{1}{6} \\
P\{R_1 = 1, R_2 = 1, R_3 = 2\} &= \frac{1}{6} \\
P\{R_1 = 1, R_2 = 2, R_3 = 2\} &= \frac{1}{6} \\
P\{R_1 = 1, R_2 = 1, R_3 = 3\} &= \frac{1}{6} \\
P\{R_1 = 1, R_2 = 2, R_3 = 3\} &= \frac{1}{6}.
\end{aligned} \tag{A.1}$$

*Similarly, it can easily be verified that*

$$\begin{aligned}
&P\{R_1 = 1, R_2 = 1\} = P\{R_1 = 1, R_2 = 2\} = \frac{1}{2} \\
&P\{R_1 = 1, R_3 = 1\} = P\{R_1 = 1, R_3 = 2\} = P\{R_1 = 1, R_3 = 3\} = \frac{1}{3} \\
&P\{R_2 = 1, R_3 = 1\} = P\{R_2 = 1, R_3 = 2\} = P\{R_2 = 1, R_3 = 3\} = \frac{1}{6} \\
&P\{R_2 = 2, R_3 = 1\} = P\{R_2 = 2, R_3 = 2\} = P\{R_2 = 2, R_3 = 3\} = \frac{1}{6}
\end{aligned} \tag{A.2}$$

*and that*

$$\begin{aligned}
&P\{R_1 = 1\} = 1 \\
&P\{R_2 = 1\} = P\{R_2 = 2\} = \frac{1}{2} \\
&P\{R_3 = 1\} = P\{R_3 = 2\} = P\{R_3 = 3\} = \frac{1}{3}.
\end{aligned} \tag{A.3}$$

*One can then show that for any integers $\alpha$, $\beta$, and $\gamma$,*

$$P\{R_1 = \alpha, R_2 = \beta, R_3 = \gamma\} \;=\; P\{R_1 = \alpha\} \times P\{R_2 = \beta\} \times P\{R_3 = \gamma\} \tag{A.4}$$

*and that for $i, j \in \{1, 2, 3\}$ with $i \neq j$, and integers $\alpha$ and $\beta$,*

$$P\{R_i = \alpha, R_j = \beta\} \quad = \quad P\{R_i = \alpha\} \times P\{R_j = \beta\} \tag{A.5}$$

*implying that $R_1$, $R_2$, and $R_3$ are independent.*

**Example A.0.5 (Dependence of regular ranks)** *Now consider the ranking of three i.i.d. observations by the usual ranking method. Again, observations $y_1$, $y_2$, and $y_3$ have corresponding ranks $R_1$, $R_2$, and $R_3$. Under the i.i.d. assumption, the following six rankings are equally likely:*

| $R_1$ | $R_2$ | $R_3$ |
|---|---|---|
| 1 | 2 | 3 |
| 1 | 3 | 2 |
| 2 | 1 | 3 |
| 2 | 3 | 1 |
| 3 | 1 | 2 |
| 3 | 2 | 1 |

*In particular, it can be verified that*

$$P\{R_1 = 1, R_2 = 2, R_3 = 3\} = \frac{1}{6} \tag{A.6}$$

*and that*

$$P\{R_1 = 1\} = P\{R_2 = 2\} = P\{R_3 = 3\} = \frac{1}{3}. \tag{A.7}$$

*Hence,*

$$P\{R_1 = 1\} \times P\{R_2 = 2\} \times P\{R_3 = 3\} = \frac{1}{9} \tag{A.8}$$

*which is not equal to the probability on the right hand side of equation (A.6). Thus, $R_1$, $R_2$, and $R_3$ cannot be independent, and hence must be dependent.*

# Appendix B

# Code

The following Fortran code was written for the simulation study of Chapter 4. The program requests as inputs the sample size for group 1, the number of groups, the method of analysis, the response distribution, and the overall error rate, alpha.

```
program simseq

use IMSL_LIBRARIES

implicit none

real, dimension(:,:), allocatable :: X, output
real, dimension(10000,2) :: reps
real, dimension(:), allocatable :: means
real, dimension(:), allocatable :: progres
real :: alpha, avst
integer, dimension(:,:), allocatable :: O
integer :: method, d, n1, k, i, j, n

! Read in parameter values:
write(*,*) 'n1 = '; read(*,*) n1
```

```
write(*,*) 'k (# of groups) = '; read(*,*) k
write(*,*) 'method (1=P, 2=OBF, 3=SS) = '; read(*,*) method
write(*,*) 'd (1=N, 2=E) = '; read(*,*) d
write(*,*) 'alpha = '; read(*,*) alpha

! Set means, the vector for mu1 or lambda1 depending on d,
! also allocating eventual output dimensions:
if (d==1) then
    allocate(means(11))
    do i=1, 11
        means(i) = (real(i)-1.0)/10.0
    end do
    allocate(output(11,2))
else if(d==2) then
    allocate(means(9))
    do i=1, 9
         means(i) = (real(i)+3.0)/4.0
    end do
    allocate(output(9,2))
end if

! Total number of subjects:
n = n1*k

allocate(X(3,n))
allocate(O(3,n))
allocate(progres(2))

do i=1, size(means)
    do j=1, 10000
```

```
          X = datagen(n1, k, d, n, means(i))
          O = ordergen(n)
          progres = seqrank(X, O, alpha, method, n, k, n1)
          reps(j,1) = progres(1)
          reps(j,2) = progres(2)
     end do
     output(i,1) = sum(reps(1:10000,1))/10000.0
     output(i,2) = sum(reps(1:10000,2))/10000.0
end do

! Output values
do i=1, size(means)
     write (*,"(f9.4,4x,f9.4)") (output(i,j), j=1, 2)
end do



contains
function datagen(n1, k, d, n, m1)
     ! This function generates a 3xn matrix of raw data.
     real, dimension(:,:), allocatable :: datagen
     real, dimension(5) :: rand
     real :: m1, temp1, temp2, temp3
     integer :: n1, k, d, n, dgi, dgj

     allocate(datagen(3,n))

     do dgi=1, n
          if (d==1) then
               call RNNOR(rand)
               if ((m1.gt.0.0).and.(dgi.le.n1)) call SADD(5,m1,rand,1)
```

```
          else if (d==2) then
               call RNEXP(rand)
               if ((m1.gt.1.0).and.(dgi.le.n1)) call SSCAL(5,m1,rand,1)
          end if

          do dgj=1, 3
               temp1 = rand(dgj+2)
               temp2 = (1.0/2.0)*rand(dgj+1)
               temp3 = (1.0/2.0)*rand(dgj)
               datagen(dgj,dgi) = temp1 + temp2 + temp3
          end do
     end do

     return
end function datagen



function ordergen(n)
     ! This function generates a 3xn random matrix of orders.
     integer, dimension(:,:), allocatable :: ordergen
     integer, dimension(:), allocatable :: temp
     integer :: n, ogi, ogj

     allocate(ordergen(3,n))
     allocate(temp(n))

     call RNPER(temp)

     do ogi=1, n
```

```
          do ogj=1, 3
               ordergen(ogj,ogi) = 3*temp(ogi) + ogj - 3
          end do
     end do

     deallocate(temp)

     return
end function ordergen



function seqrank(X, O, alpha, method, n, k, n1)
     ! This function returns the results of one simulation.
     real, dimension(:,:,:), allocatable :: C
     real, dimension(:,:), allocatable :: X, R, M, Z
     real, dimension(:,:), allocatable :: Ztemp
     real, dimension(2) :: seqrank
     real, dimension(:), allocatable :: tempr1
     real, dimension(:), allocatable :: snt, snt2
     real, dimension(:), allocatable :: nsubs1, nsubs2
     real, dimension(:), allocatable :: tsum1, tsum2
     real, dimension(:), allocatable :: tstat1, tstat2
     real :: alpha, sd, b
     integer, dimension(:,:), allocatable :: O
     integer, dimension(:), allocatable :: rowobs, colobs, Orow1
     integer, dimension(:), allocatable :: tempi1, tempi2
     integer, dimension(3) :: tempi3
     integer :: method, n, k, n1
     integer :: nt, sri, srj, srk, srl, count, rind
```

```
nt = 3*n

allocate(R(3,n),M(3,n),Z(3,n))
allocate(rowobs(nt),colobs(nt))
allocate(Orow1(n),tempi1(n),tempi2(n))
allocate(snt(n**2), snt2(n**2))

! Pull out the permutation that rearranges the first
! row of the matrix O, and place it in tempi2:
Orow1(1:n) = O(1,1:n)

do sri=1, n
     tempi1(sri) = sri
end do

call SVIGP(Orow1,tempi2, tempi1)

do sri=1, 3
     tempi3(sri) = sri
end do

do sri=1, n
     rowobs( (3*(sri-1)+1) : (3*(sri-1)+3) ) = tempi3(1:3)
     colobs( (3*(sri-1)+1) : (3*(sri-1)+3) ) = tempi1(sri)
end do

deallocate(tempi1, tempi2)
```

```
! Calculate the sequential ranks:
do sri=1, nt

     ! Figure out how many observations occurred before the one
     ! indexed by sri.
     count = 0
     do srj=1, n
          if (O(rowobs(sri),srj).le.sri) count = count + 1
     end do

     allocate(tempr1(count))

     ! Extract the X's for the observations that occurred
     ! before the one indexed by sri.
     srk=1
     do srj=1, n
          if (O(rowobs(sri),srj).lt.sri) then
                tempr1(srk) = X(rowobs(sri),srj)
                srk = srk + 1
          else if (O(rowobs(sri),srj)==sri) then
                tempr1(srk) = X(rowobs(sri),srj)
                rind = srk
                srk = srk + 1
          end if
     end do

     ! Find and store the sequential rank for sri, as well as its
     ! distributional parameter.
     if (count.ge.2) then
          call RANKS(tempr1, tempr1, ITIE=3)
```

```
          R(rowobs(sri),colobs(sri)) = tempr1(rind)
          M(rowobs(sri),colobs(sri)) = real(count)
     else
          R(rowobs(sri),colobs(sri)) = 1.0
          M(rowobs(sri),colobs(sri)) = 1.0
     end if

     deallocate(tempr1)
end do



! Standardize the sequential ranks:
do sri=1, n
     do srj=1, 3
          sd = sqrt((M(srj,sri)**2.0-1.0)/12.0)
          if (sd.gt.0.0) then
               Z(srj,sri) = (R(srj,sri)-(M(srj,sri)+1.0)/2.0)/sd
          else
               Z(srj,sri) = 0.0
          end if
     end do
end do



! Estimate covariances, calculate within-subject sums, and
! determine how many patients are in the trial at any
! given analysis time:
if (method.ne.3) then
     allocate(C(n,3,3))
     allocate(nsubs1(n),nsubs2(n))
```

```
end if

do sri=1, n
     srk = 3*sri

     allocate(Ztemp(3,sri))
     count = 1

     do srj=1, n
          if (O(3,srj).le.srk) then
               snt(n*(sri-1)+srj) = Z(1,srj)+Z(2,srj)+Z(3,srj)
               Ztemp(1:3,count) = Z(1:3,srj)
               count = count + 1
          else
               snt(n*(sri-1)+srj) = 0
          end if
     end do

     if(method.ne.3) then
          if (sri.ge.2) then
               call CORVC(3,.t.Ztemp,C(sri,1:3,1:3))
          else
               C(sri,1:3,1:3) = 0.0
          end if

          count = 0
          do srj=1, n1
               if (O(3,srj).le.srk) then
                    count = count + 1
               end if
```

```
            end do
            nsubs1(sri) = real(count)

            count = 0
            do srj=n1+1, 2*n1
                 if (O(3,srj).le.srk) then
                      count = count + 1
                 end if
            end do
            nsubs2(sri) = real(count)
       end if

       deallocate(Ztemp)
  end do


  ! Calculate standardized sums, for Test 1/Test 2, and
  ! signs of the sums for Test 3:
  do sri=1, n
       if (method.ne.3) then
             sd = sqrt(sum(C(sri,1:3,1:3)))
             if((sri.ge.2).and.(sd.gt.0.0)) then
                  do srj=1, n
                       snt2(n*(sri-1)+srj) = snt(n*(sri-1)+srj)/sd
                  end do
             else
                  snt2(n*(sri-1)+1:n*(sri-1)+n) = 0.0
             end if
       else
             do srj=1, n
```

```
                if(snt(n*(sri-1)+srj).gt.0.0) then
                     snt2(n*(sri-1)+srj) = 1.0
                else if (snt(n*(sri-1)+srj).lt.0.0) then
                     snt2(n*(sri-1)+srj) = -1.0
                else
                     snt2(n*(sri-1)+srj) = 0.0
                end if
           end do
     end if
end do


! Calculate test sums:
allocate(tsum1(n), tsum2(n))
allocate(tstat1(n), tstat2(n))

do sri=1, n
     tsum1(sri) = sum(snt2(n*(sri-1)+1:n*(sri-1)+n1))
     tsum2(sri) = sum(snt2(n*(sri-1)+n1+1:n*(sri-1)+2*n1))
end do


! Set the boundaries and test statistics:
if (method==1) then
     if (k==2) then
          if (alpha==0.1) then
               if (n1==20) then
                    b = 2.65
               else if (n1==50) then
                    b = 2.74
```

```
            end if
        else if (alpha==0.05) then
            if(n1==20) then
                b = 2.93
            else if (n1==50) then
                b = 3.02
            end if
        else if (alpha==0.01) then
            if (n1==20) then
                b = 3.48
            else if (n1==50) then
                b = 3.56
            end if
        end if

        do sri=1, n
            if (nsubs1(sri).gt.0.0) then
                tstat1(sri) = tsum1(sri)/sqrt(nsubs1(sri))
            else
                tstat1(sri) = 0.0
            end if
        end do
    else if (k==3) then
        if (alpha==0.1) then
            if (n1==20) then
                b = 2.92
            else if (n1==50) then
                b = 3.01
            end if
        else if (alpha==0.05) then
```

```
                if (n1==20) then
                      b = 3.18
                else if (n1==50) then
                      b = 3.26
                end if
           else if (alpha==0.01) then
                if (n1==20) then
                      b = 3.69
                else if (n1==50) then
                      b = 3.76
                end if
           end if

           do sri=1, n
                if (nsubs1(sri).gt.0.0) then
                      tstat1(sri) = tsum1(sri)/sqrt(nsubs1(sri))
                else
                      tstat1(sri) = 0.0
                end if

                if (nsubs2(sri).gt.0.0) then
                      tstat2(sri) = tsum2(sri)/sqrt(nsubs2(sri))
                else
                      tstat2(sri) = 0.0
                end if
           end do
      end if
else if (method==2) then
      if (k==2) then
           if (alpha==0.1) then
```

```
            b = 1.96
        else if (alpha==0.05) then
            b = 2.24
        else if (alpha==0.01) then
            b = 2.80
        end if

        tstat1(1:n) = tsum1(1:n)/sqrt(real(n1))
    else if (k==3) then
        if (alpha==0.1) then
            b = 2.23
        else if (alpha==0.05) then
            b = 2.49
        else if (alpha==0.01) then
            b = 3.00
        end if

        tstat1(1:n) = tsum1(1:n)/sqrt(real(n1))
        tstat2(1:n) = tsum2(1:n)/sqrt(real(n1))
    end if
else
    if (k==2) then
        if (alpha==0.1) then
            if (n1==10) then
                b = 6.0
            else if (n1==15) then
                b = 8.0
            else if (n1==20) then
                b = 9.0
            else if (n1==50) then
```

```
                b = 14.0
            end if
        else if (alpha==0.05) then
            if (n1==10) then
                b = 7.0
            else if (n1==15) then
                b = 9.0
            else if (n1==20) then
                b = 10.0
            else if (n1==50) then
                b = 16.0
            end if
        else if (alpha==0.01) then
            if (n1==10) then
                b = 8.0
            else if (n1==15) then
                b = 10.0
            else if (n1==20) then
                b = 12.0
            else if (n1==50) then
                b = 20.0
            end if
        end if

        tstat1(1:n) = tsum1(1:n)
    else if (k==3) then
        if (alpha==0.1) then
            if (n1==10) then
                b = 7.0
            else if (n1==15) then
```

```
            b = 9.0
      else if (n1==20) then
            b = 10.0
      else if (n1==50) then
            b = 16.0
      end if
else if (alpha==0.05) then
      if (n1==10) then
            b = 8.0
      else if (n1==15) then
            b = 10.0
      else if (n1==20) then
            b = 11.0
      else if (n1==50) then
            b = 18.0
      end if
else if (alpha==0.01) then
      if (n1==10) then
            b = 9.0
      else if (n1==15) then
            b = 12.0
      else if (n1==20) then
            b = 13.0
      else if (n1==50) then
            b = 21.0
      end if
end if

tstat1(1:n) = tsum1(1:n)
tstat2(1:n) = tsum2(1:n)
```

```
     end if
end if


! Monitor:
count = 1
if (k==2) then
     if (method.ne.3) then
          do sri=1, n
               if (abs(tstat1(sri)).gt.b) then
                    exit
               else
                    count = count + 1
               end if
          end do
     else
          do sri=1, n
               if (abs(tstat1(sri)).ge.b) then
                    exit
               else
                    count = count + 1
               end if
          end do
     end if
else if (k==3) then
     if(method.ne.3) then
          do sri=1, n
               if ((abs(tstat1(sri)).gt.b).or.(abs(tstat2(sri)).gt.b)) then
                    exit
               else
```

```
                    count = count + 1
               end if
          end do
     else
          do sri=1, n
               if ((abs(tstat1(sri)).ge.b).or.(abs(tstat2(sri)).ge.b)) then
                    exit
               else
                    count = count + 1
               end if
          end do
     end if
end if


! Set the vector of results:
if (count==(n+1)) then
     seqrank(1) = 0.0
     seqrank(2) = real(n)
else
     seqrank(1) = 1.0
     seqrank(2) = real(count)
end if

if (method.ne.3) then
     deallocate(C)
     deallocate(nsubs1,nsubs2)
end if

deallocate(R, M, Z)
```

```
      deallocate(snt, snt2)
      deallocate(tsum1, tsum2)
      deallocate(tstat1, tstat2)
      deallocate(rowobs, colobs, Orow1)

      return
end function seqrank

end program simseq
```