Simultaneous Inference for Ratios of Location Parameters

Von der Naturwissenschaftlichen Fakultät

der Universität Hannover

zur Erlangung des akademischen Grades eines

Doktors der Gartenbauwissenschaften

-Dr.rer.hort.-

genehmigte

Dissertation

von

Gemechis Dilba Djira

geboren am 06.03.1968 in Äthiopien

2005

Referent:	Prof. Dr. Ludwig A. Hothorn
Korreferent:	PD Dr. Frank Bretz
Tag der Promotion:	13 Juli 2005

Zusammenfassung

Die Inferenz für Quotienten von Lageparametern oder Quotienten von Koeffizienten im verallgemeinerten linearen Modell ist in vielen Forschungsfeldern von Bedeutung. Die Verwendung solcher Quotienten umfasst: Schätzung der relativen Wirksamkeit bei multiplen Bioassays, Entscheidung über Äquivalenz, Nicht-Unterlegenheit und Überlegenheit von mehreren Behandlungen basierend auf relativen Schwellenwerten. Die vorliegende Arbeit behandelt Fragestellungen mit mehr als einem Quotienten von Lageparametern (oder Quotienten von Linearkombinationen von Lageparametern). Die spezifischen Probleme sind: (i) die Herleitung und Untersuchung von Güte und Fallzahlschätzung für einseitige simultane Tests quotientenbasierter Vergleiche zur Kontrolle, (ii) die Entwicklung neuer Methoden zur Konstruktion exakter simultaner Konfidenzmengen und approximativer simultaner Konfidenzintervalle für multiple Quotienten.

Im ersten Teil führen wir in die grundsätzliche Idee des quotientenbasierten, multiplen Testens ein und wenden uns dann dem Problem der Berechnung von Güte und Fallzahl für simultane Tests auf Nicht-Unterlegenheit zu. Hier betrachten wir den Fall des Vergleichs mehrerer experimenteller Behandlungen mit einer aktiven Kontrolle. Der Ansatz basiert auf den Quotienten zur Kontrolle mit gemeinsamer Nicht-Unterlegenheits-Grenze. Zwei Gütedefini-

tionen im multiplen Testen, complete power und minimal power, werden diskutiert. Weiterhin werden die notwendigen Fallzahlen für quotientenbasierte Inferenz mit den Fallzahlen verglichen, die bei Inferenz basierend auf Differenzen von Mittelwerten in vergleichbaren Situationen benötigt werden. In einer numerischen Studie wird gezeigt, dass die benötigte Fallzahl für quotientenbasierte Inferenz kleiner ist als die für Inferenz basierend auf Differenzen, wenn die relative Nicht-Unterlegenheits-Grenze kleiner als eins ist und wenn grosse Werte der Zielgrösse eine bessere Wirkung der Behandlung bedeuten. Das Design von Versuchen zum Nachweis quotientenbasierter Nicht-Unterlegenheit (oder Überlegenheit) ist anhand von Beispielen veranschaulicht.

Im zweiten Teil behandeln wir das Problem simultaner Konfidenzintervalle für mehrere Quotienten. Im einfachsten Fall eines einzelnen Quotienten kann anhand von Fiellers Theorem ein Konfidenzintervall für einen Quotienten berechnet werden. Für multiple Quotienten gibt es keine Methoden zur Konstruktion simultaner Konfidenzintervalle, die ein gegebenes versuchsbezogenes Konfidenzniveau exakt einhalten. Das grundlegende Problem ist hier, dass die gemeinsame Verteilung der Teststatistiken, die für die simultane Schätzung benötigt wird, von den unbekannten Quotienten abhängt. Die derzeit verfügbaren Methoden zur Konstruktion von simultanen Konfidenzintervallen sind konservativ, da sie auf Ungleichungen basieren. In dieser Arbeit

betrachten wir zuerst exakte simultane Konfidenzmengen, die auf der multivariaten t-Verteilung basieren. Zwei Ansätze zur Bestimmung dieser exakten Konfidenzmengen werden vorgeschlagen. Weiterhin werden approximative simultane Konfidenzintervalle, basierend auf der multivariaten t-Verteilung mit geschätzter Korrelationsmatrix, sowie ein resampling-Ansatz vorgeschlagen. Die Methoden werden auf Quotienten von Linearkombinationen der Mittelwerte (Parameter, Koeffizienten) von Einweg-Anlagen und Quotienten von Parameterkombinationen im verallgemeinerten linearen Modell angewendet. Umfangreiche Monte-Carlo-Simulationen werden durchgefhrt, um das Verhalten der verschiedenen Methoden in Hinsicht auf die Stabilität der geschätzten kritischen Werte und der coverage probability zu vergleichen. Eine vorgeschlagene Methode zur Konstruktion approximativer simultaner Konfidenzintervalle wird als plug-in-Ansatz bezeichnet. Bei dieser Methode werden die Maximum-Likelihood-Schätzer der Quotienten in die unbekannte Korrelationsmatrix der multivariaten t-Verteilung eingesetzt. Die Untersuchungen zeigen, dass die plug-in-Methode besser als alle anderen Methoden das nominelle versuchsbezogene Konfidenzniveau erreicht.

Schlagworte: multiple Quotienten, Nicht-Unterlegenheit, simultane Inferenz

Abstract

Inferences concerning ratios of location parameters or ratios of coefficients in the general linear model arise in many areas of research. Applications of such ratios include: relative potency estimation in multiple assays (direct, parallel-line, slope-ratio assays), inferences for equivalence, non-inferiority, and superiority of several treatments based on relative margins, and so on. This research focuses on problems involving more than one ratio of location parameters (or ratios of linear combinations of location parameters). The specific objectives are: (i) to derive and investigate power and sample size in one-sided simultaneous tests for comparisons with a control based on ratios, (ii) developing new methods of constructing exact simultaneous confidence sets and approximate simultaneous confidence intervals for multiple ratios.

In the first part, we introduce the general idea of multiple testing based on ratios and then address the special problem of calculating power and sample sizes associated with simultaneous tests for non-inferiority. We consider the case of comparing several experimental treatments with an active control. The approach is based on the ratio view, where the common non-inferiority margin is chosen to be some percentage of the mean of the control treatment. Two power definitions in multiple hypothesis testing, namely, complete power and minimal power, are used in the discussions. The sample sizes associated

with the ratio-based inferences are also compared with that of comparable inferences based on the difference of means for various scenarios. A numerical study revealed that the sample sizes required for ratio-based inferences are smaller than that of the difference-based inferences when the relative non-inferiority margin is less than one and when large response values indicate better treatment effects. The design of non-inferiority trials (or superiority trials) based on the ratio view are illustrated with examples.

In the second part, we deal with the problem of simultaneously estimating multiple ratios. In the simplest case of only one ratio parameter, Fieller's theorem provides a confidence interval for the single ratio. For multiple ratios, there is no method available to construct simultaneous confidence intervals that exactly satisfy a given family-wise confidence level. The key challenge here is that the joint distribution of the test statistics used for simultaneous estimation depends on the unknown ratios. The currently available methods for constructing simultaneous confidence intervals are conservative since they are based on probability inequalities. In this research, first we consider exact simultaneous confidence sets based on the multivariate t-distribution. Two approaches of determining the exact simultaneous confidence sets are proposed. Second, approximate simultaneous confidence intervals based on the multivariate t-distribution with estimated correlation matrix and a re-sampling approach are proposed. The methods are applied to ratios of linear

combinations of the means in the one-way layout and ratios of parameter combinations in the general linear model. Extensive Monte Carlo simulations are carried out to compare the performance of the various methods with respect to the stability of the estimated critical points and of the coverage probabilities. One of the methods proposed to construct approximate simultaneous confidence intervals is called the plug-in approach. This method works by plugging the maximum likelihood estimates of the ratios in the unknown correlation matrix of a multivariate t-distribution. It is observed that the plug-in method outperforms all other methods in terms of achieving the nominal family-wise confidence level.

Keywords: multiple ratios, non-inferiority, simultaneous inference

Notations

$\boldsymbol{\mu}$: vector of treatment means

$\boldsymbol{\gamma}$: vector of ratio parameters

ψ: threshold against which to perform tests

θ^*: clinically irrelevant percentage to be detected

α: type I familywise error rate in multiple testing

$1-\beta$: given power

$\mathbf{R}$: correlation matrix

$Mt_k(\nu, \mathbf{R})$: a central k-variate t-distribution with ν degrees of freedom and a correlation matrix $\mathbf{R}$

$Mt_k(\nu, \mathbf{R}, \boldsymbol{\delta})$: a non-central k-variate t-distribution with ν degrees of freedom, a correlation matrix $\mathbf{R}$, and a non-centrality vector $\boldsymbol{\delta}$

CV_0: coefficient of variation of the control group

$CV_{\overline{Y}_0}$: coefficient of variation of the mean of the control group

List of Figures

List of Tables

Contents

Chapter 1

General Introduction

1.1 Introduction

Inferences concerning ratios of location parameters or ratios of coefficients in the general linear model arise in many areas of research. Among others, ratios appear (i) in estimation of relative potencies in biological assays (bioassays) (e.g., Jensen, 1989; Sen, 1998), (ii) in ratio-based inference for equivalence and non-inferiority trials (Hsu *et al.*, 1994; Röhmel, 1998; Hauschke *et al.*, 1999a; Hauschke *et al.*, 1999b), (iii) in calibration (inverse regression) problems where interest lies in estimating value of an independent variable after fitting regression models (Fox, 1991; Lee, 1998), (iv) in estimating the abscissa of the point of intersection of two simple linear regressions, or estimating the point of extremum in a quadratic regression (Buonaccorsi and

Iyer, 1984), (v) in discriminant analysis (ratios of discriminant coefficients, e.g., see Chikuse, 1981), and (vi) in health economics to estimate incremental cost-effectiveness ratios (Gold *et al.*, 1996; Laska *et al.*, 1997). In these problems, either a single ratio of location parameters (or coefficients) is involved or there are multiple ratios. The central focus of this research is on the latter case where more than one ratio is of interest.

Inference for one ratio parameter is studied by several researchers in various contexts. Here we provide only few examples. In dose-response studies (with only one compound or insecticide), estimation of the dose level that produces 50% response (ED50) is often of interest. For example, Faraggi *et al.* (2003) compared various methods of estimating ED50. In parallel-line assay involving two preparations (standard and test preparations), the relative potency of the test preparation with respect to the standard is also expressed as a single ratio, the ratio of the difference of the intercepts to the common slope. The incremental cost-effectiveness ratio (ICER), which is the ratio of the average cost difference (of two treatments) to the average effect difference is also widely studied. For example, Polsky *et al.* (1997) and Briggs *et al.* (1999) compared various methods of constructing confidence intervals for ICER via simulation.

To further motivate our aim, we consider the one-way layout. Often statistical inferences about location parameters are done for the differences

between the treatment means (e.g., in comparisons with a control). However, for some biomedical problems, it is tenable to base inference on ratios of treatment means. In multiple tests for equivalence, non-inferiority, or superiority, the margins based on the ratio approach can easily be medically interpreted and can easily be defined as compared with inferences based on differences. For example, see Hauschke and Kieser (2001) for applications in non-inferiority trials where k treatments are compared with a control based on the ratio view. Ratio parameters have the notable advantage of being dimensionless (free of the unit of measurement of the endpoint) as compared with the difference of location parameters. In addition to interpretational convenience, more recently, Laster and Johnson (2003) showed that the ratio view is more powerful in test for non-inferiority of an experimental therapy as compared with a standard one.

1.2 Objectives

The kernel of this research is design in multiple testing based on ratios and simultaneous confidence intervals estimation of several ratios. For example, it might be of interest to compare not only two treatments but several treatments with a control treatment (positive or negative control). In this comparison, one encounters multiple ratios. Previously, this problem is ad-

dressed by Hauschke and Kieser (2001) in multiple tests for non-inferiority. Biesheuvel (2002) also discussed the issue of multiple testing for ratios as applied in stratified designs. Jensen (1989) discussed in detail simultaneous estimation in direct, parallel-line and slope-ratio assays based on Ŝidák (1967) inequality and compared the results with a method due to Scheffé (1970). In some circumstances, inferences based on probability inequalities and the projection method discussed by Scheffé (1970) can be quite conservative. Therefore, in this research, we propose two alternative simultaneous estimation procedures and investigate their performance by simulation. The main objectives are: (i) to derive power formulas and compute the sample sizes required in simultaneous tests based on ratios, (ii) to develop simultaneous confidence sets (SCS) and simultaneous confidence intervals (SCI) based on the joint distribution of the relevant test statistics which follows a multivariate t-distribution.

Due to the inherent multiplicity aspects, we need special statistical techniques for simultaneous inference. There exist numerous multiple comparison procedures in the statistical literature. See Miller (1977) for a detailed historical development and bibliography on these procedures. Standard textbooks on this subject include Miller (1981), Hochberg and Tamhane (1987), and Hsu (1996). Multiple comparison procedures are broadly classified as single-step and step-wise procedures. In single-step procedures, the critical

points are determined once and for all comparisons (e.g., the simple Bonferroni adjustment). Whereas, in step-wise procedures, the critical points are determined in a sequential manner (e.g., Bonferroni-Holm procedure). The methods used in this research are all single-step procedures.

Accordingly, the thesis is organized as follows. In Chapter 2, we present various datasets to be analyzed later on. In Chapter 3, we describe the problem of multiple testing for ratios and thereby introduce the notations to be used in subsequent chapters. Chapter 4 deals with the special problem of calculating power and sample size in simultaneous tests for non-inferiority based on relative margins. Chapter 5 deals with the general problem of constructing simultaneous confidence sets and simultaneous confidence intervals for several ratios. In this chapter, two methods of constructing SCS and several methods of constructing SCI are discussed. Finally, Chapter 6 is devoted to the conclusions and proposals for further research.

Chapter 2

Data Examples

In this chapter, we describe four motivating data examples to be analyzed in later chapters. The first two datasets will be used to illustrate the construction of SCS and SCI, while the other two will be used for design purposes.

2.1 Chroma Value and Chlorophyll Content

Mutui *et al.* (2005) compared three *Pelargonium* cultivars ('Fire', 'Katinka', and 'Ganymed') and four levels of ethylene with respect to chroma value and total chlorophyll content of leaves. Of the three cultivars, here we consider only 'Katinka'. The control group is untreated (ethylene at 0 level). The summary statistics for the two response variables are provided in Table 2.1. The aim is to simultaneously compare the three active treatments with the

Table 2.1: Descriptive summary of leaf chroma value and chlorophyll content

	Chroma value			Chlorophyll content		
Ethe. ($\mu l/l$)	n	mean	sd	n	mean	sd
0	24	28.4	4.1	24	3.2	0.3
0.5	24	29.7	5.1	24	2.8	0.2
1	24	31.3	10.4	24	2.3	0.4
2	24	42.0	2.7	24	1.5	0.3

control for both leaf chroma value and leaf chlorophyll content based on the ratio of the treatment means to that of the control. In other words, we construct simultaneous confidence intervals for the three ratios to control and see how the ratios (percentages) change. The merit of this construction is that the SCI are dimensionless (percentages) though the two responses are in different units. Moreover, these type of confidence intervals would enable us to tell the percentage increase (decrease) in the mean of the treatments as compared with that of the control group.

Table 2.2: Body weight gain data

Treatment										
Control	107	91	115	90	133	95	112	115	117	91
Thyroxin	119	88	84	133	87	118	132			
Thiouracil	61	68	89	80	69	52	80	63	63	68

2.2 Body Weight Gain Data

Consider data on weight gains of rats treated with three treatments, namely, control, Thyroxin, and Thiouracil (Westfall and Young, 1993, p29). The number of observations under the three treatments are $n_0 = 10$, $n_1 = 7$ and $n_2 = 10$, respectively. The original data are longitudinal and weights of rats were recorded weekly at five time points. The primary variable of interest is the weight gain over the study period (the difference between the weight at the end of the study and the baseline). The data are shown in Table 2.2. The aim is to simultaneously compare the effect of Thyroxin and Thiouracil with the control based on the ratio of treatment means.

2.3 Osteoporosis Study

Consider the problem of determining the sample size in a simultaneous non-inferiority test of three intermittent administration schedules against a continuous administration of the same total dose of ibandronate in a long-term *in-vivo* study on osteoporosis (Hothorn and Bauss, 2004). The continuous administration represents the active control group and the endpoint is trabecular bone mass in tibiae (in %). According to previous results, it can be assumed that the coefficient of variation for the control group is 50% (mean and standard deviation of 10% and 5%, respectively). Suppose that the interest is to design a new confirmatory non-inferiority trial based on these previous study results with a relative non-inferiority margin of 0.70 (which was explicitly described for trabecular bone mass in osteoporosis trials (Kanis *et al.*, 2002)), a minimal power of 80% and overall type I error rate $\alpha = 0.05$. In particular, it is of interest to investigate the optimum sample size allocation across the four treatment arms when the total sample size is fixed.

2.4 Superiority in a Clinical Trial

Knapp *et al.* (2001) described a double-blind, placebo-controlled, multicentre trial with four arms on subjects with hypercholesterolaemia. The

study involved the comparisons of placebo with two doses of Simvastatin and a third treatment group with a fixed Simvastatin/Colesevelam dose combination. The primary outcome variable was serum low density lipoprotein (LDL) cholesterol level after 45 days. The goal of this study was to show superiority by a cholesterol level reduction of at least 10% over placebo in at least one treatment arm. Assuming that a follow-up confirmatory trial is planned, we are interested in calculating the necessary sample size for the proof of efficacy due to the superiority associated with a minimal power of 80% and $\alpha = 0.025$. From the previous study, it can be assumed that the coefficient of variation for the control group is 17% (mean and standard deviation of 177 and 30, respectively).

Chapter 3

Simultaneous Tests of Hypotheses

3.1 Simultaneous Tests

Consider the problem of simultaneously comparing k experimental treatments ($i = 1, \ldots, k$) with a control treatment ($i = 0$). Let Y_{ij} denote independent observations from a normal distribution with mean μ_i and common unknown variance σ^2, $i = 0, 1, \ldots, k$; $j = 1, \ldots, n_i$. Throughout, the variances are assumed to be homogeneous. The primary interest is to use a fixed *relative* threshold ψ (a dimensionless number) in constructing the tests.

Without loss of generality, we consider the case when the responses are nonnegative and when large response values indicate better treatment effects.

We wish to test

$$H_{0\ell} : \gamma_\ell \leq \psi \quad \text{against} \quad H_{1\ell} : \gamma_\ell > \psi, \quad \ell = 1, 2, \ldots, k, \tag{3.1}$$

where $\gamma_\ell = \mu_\ell / \mu_0$ denotes the ratio of the mean of the ℓ^{th} treatment to that of the control. For $\mu_i > 0$, the hypotheses in (3.1) can equivalently be stated as

$$H_{0\ell} : \mu_\ell - \psi\mu_0 \leq 0 \quad \text{against} \quad H_{1\ell} : \mu_\ell - \psi\mu_0 > 0, \quad \ell = 1, 2, \ldots, k,$$

which naturally lead us to the accompanying test statistics. Let $\overline{Y}_i$ and S^2 be the usual unbiased estimators of μ_i and σ^2, respectively. The pooled variance estimator S^2 has $\nu = \sum_{i=0}^{k}(n_i - 1)$ degrees of freedom. The likelihood ratio statistics to test the hypotheses in (3.1) are

$$T_\ell(\psi) = \frac{\overline{Y_\ell} - \psi\overline{Y_0}}{S\sqrt{\frac{1}{n_\ell} + \frac{\psi^2}{n_0}}} = \frac{Z_\ell}{S/\sigma}, \quad \ell = 1, 2, \ldots, k, \tag{3.2}$$

where

$$Z_\ell = \frac{\overline{Y_\ell} - \psi\overline{Y_0}}{\sigma\sqrt{\frac{1}{n_\ell} + \frac{\psi^2}{n_0}}}.$$

Under $H_{0\ell}$, each of the $T_\ell(\psi)$'s follows a central t-distribution with ν degrees of freedom. The random vector $\mathbf{T} = (T_1, \ldots, T_k)'$ of the test statistics in (3.2) then follows a central k-variate t-distribution with ν degrees of freedom and a correlation matrix $\mathbf{R}(\psi) = [\rho_{ij}(\psi)]$, where

$$\rho_{ij}(\psi) = \text{Corr}(Z_i, Z_j) = \frac{\psi}{\sqrt{\frac{n_0}{n_i} + \psi^2}} \frac{\psi}{\sqrt{\frac{n_0}{n_j} + \psi^2}}, \quad 1 \leq i \neq j \leq k. \tag{3.3}$$

This is based on the fact that $\mathbf{Z} = (Z_1, \ldots, Z_k)'$ follows a multivariate normal distribution with mean $\mathbf{0}$ and a correlation matrix $\mathbf{R}(\psi)$ (see Kotz and Nadarajah (2004) for various definitions of central and non-central multivariate t-distributions). We designate a central k-variate t-distribution with ν degrees of freedom and a correlation matrix $\mathbf{R}$ by $Mt_k(\nu, \mathbf{R})$. The $H_{0\ell}$ hypothesis will be rejected if the test statistic $T_\ell(\psi)$ is larger than a multiplicity adjusted critical point c (see Hochberg and Tamhane, 1987). The method in this section aims at controlling the family-wise error rate (FWE) (also called the experiment-wise error rate) in the strong sense. FWE is the chance of erroneously rejecting at least one true null hypothesis. That is, if $I_0 \subseteq I = \{1, \ldots, k\}$ denotes the index set of all true null hypotheses, then

$$\text{FWE} = 1 - P\{T_\ell(\psi) \leq c \text{ for all } \ell \in I_0\}.$$

In order to get FWE $\leq \alpha$ for any configuration of the true null hypotheses, we require FWE $= \alpha$ in the case of $I_0 = I$. The critical point c is seen to be an equi-coordinate percentage point of $\mathbf{T} \sim Mt_k(\nu, \mathbf{R}(\psi))$ satisfying

$$P\left\{T_1(\psi) \leq c, \ldots, T_k(\psi) \leq c\right\} = 1 - \alpha. \tag{3.4}$$

As shown in Figure 3.1, for two comparisons (i.e., $k = 2$), the critical point c is determined such that the volume above the shaded region is $1 - \alpha$. The correlation matrix $\mathbf{R}(\psi)$ in (3.3) has a product correlation structure, i.e., we can factorize $\rho_{ij}(\psi)$ as $\rho_{ij}(\psi) = \lambda_i \lambda_j$, where $\lambda_i = \left(n_i \psi^2 / (n_0 + n_i \psi^2)\right)^{1/2}$.

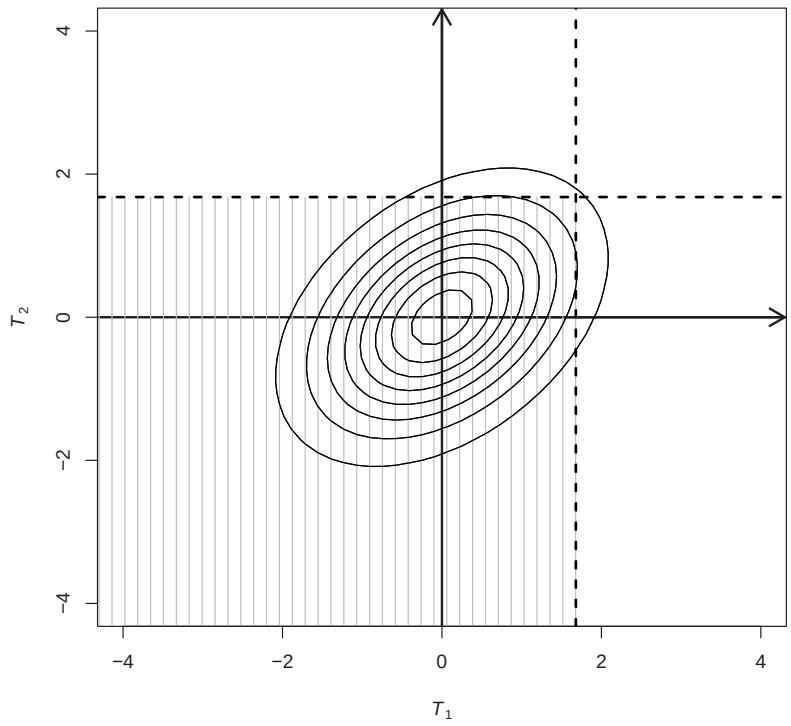

Figure 3.1: *Contour plot of a central bivariate t-distribution with* $\nu = 30$ *and* $\rho_{12} = 0.4$.

This property enables us to reduce the dimension of the multivariate normal integrals involved in the computation of multivariate t equi-coordinate percentage points (see, e.g., Tong (1990)). In general, for the computation of central/non-central multivariate t critical points and probabilities, we refer to numerical integration routines of Genz and Bretz (1999, 2002) which can directly be used for power calculations (Bretz, Genz and Hothorn, 2001). This algorithm is not restricted to special correlation structures.

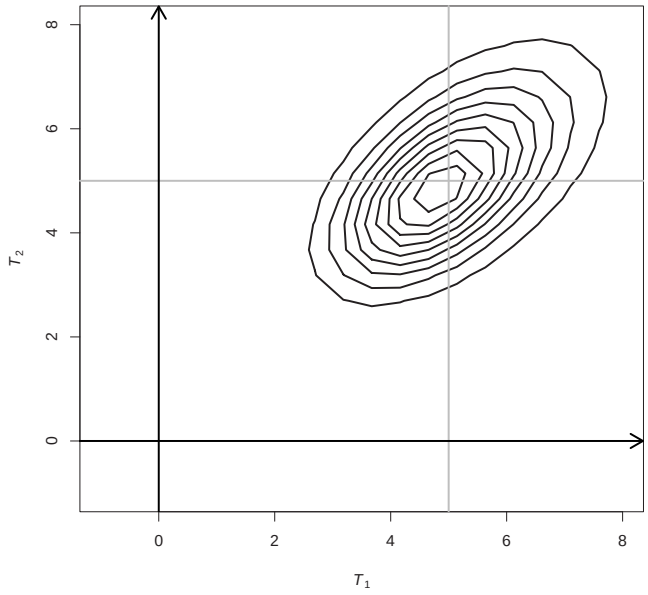

Figure 3.2: *Contour plot of a non-central bivariate t-distribution with* $\nu = 30$*,* $\rho_{12} = 0.4$*, and* $\delta_1^{\text{ratio}} = \delta_2^{\text{ratio}} = 5$.

3.2 Distribution Under the Alternatives

Let $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)'$ denote a vector of the true but unknown ratios. Note that some of the θ_ℓ's may be less than or equal to ψ (i.e., under $H_{0\ell}$). When some of the $H_{1\ell}$'s are true, the vector of test statistics $\mathbf{T}$ has a non-central k-variate t-distribution with ν degrees of freedom, a correlation matrix $\mathbf{R}(\psi)$, and non-centrality vector $\boldsymbol{\delta}^{\text{ratio}} = (\delta_1^{\text{ratio}}, \ldots, \delta_k^{\text{ratio}})'$, where

$$\delta_\ell^{\text{ratio}} = \frac{\theta_\ell - \psi}{CV_0\sqrt{\frac{1}{n_\ell} + \frac{\psi^2}{n_0}}}, \quad \ell = 1, \ldots, k,$$

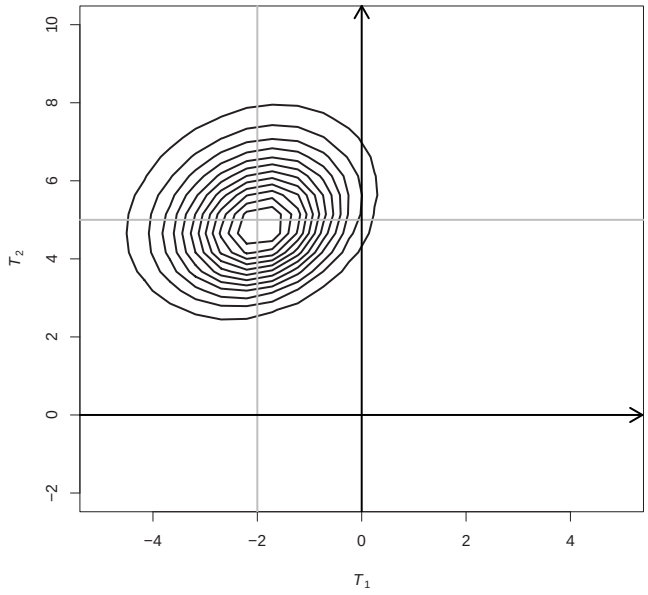

Figure 3.3: *Contour plot of a non-central bivariate t-distribution with* $\nu = 30$*,* $\rho_{12} = 0.4$*,* $\delta_1^{\text{ratio}} = -2$*, and* $\delta_2^{\text{ratio}} = 5$*.*

and $CV_0 = \sigma / \mu_0$ denotes the coefficient of variation of the control group. We denote this distribution by $Mt_k(\nu, \mathbf{R}, \boldsymbol{\delta}^{\text{ratio}})$. The non-centrality parameters are dimensionless as are the ratio parameters γ_ℓ. Power computations for the tests in (3.1) depend on this non-central multivariate t-distribution and it is a function of the threshold value ψ, the fixed ratios θ_ℓ, the coefficient of variation of the control group CV_0, and the sample sizes n_i, $i = 0, 1, \ldots, k$. Two examples of bivariate non-central t-distributions are shown in Figures 3.2 and 3.3. These are non-central t-distributions according to Kshirsagar (1961) (see also Kotz and Nadarajah, 2004, p. 87, for an explicit expression

of the probability density function). Since these probability density functions involve infinite sums, Figures 3.2 and 3.3 are obtained by taking only the first few terms. The figures show the effect of the non-centrality parameters on the shape of the distribution. In Figure 3.2, the non-centrality parameters are positive and the same, whereas in Figure 3.3 they are unequal and of different signs.

3.3 Ratios of Linear Combinations

Multiple testing concerning the ratios of linear combinations of the treatment means can be performed in much similar fashion as in the previous section. Suppose that we wish to test

$$H_{0\ell} : \frac{\mathbf{c}'_\ell \boldsymbol{\mu}}{\mathbf{d}'_\ell \boldsymbol{\mu}} = \psi \quad \text{against} \quad H_{1\ell} : \frac{\mathbf{c}'_\ell \boldsymbol{\mu}}{\mathbf{d}'_\ell \boldsymbol{\mu}} \neq \psi, \quad \ell = 1, \ldots, r, \tag{3.5}$$

where r is the number of ratio parameters. The vectors $\mathbf{c}_\ell = (c_{\ell 0}, \ldots, c_{\ell k})'$ and $\mathbf{d}_\ell = (d_{\ell 0}, \ldots, d_{\ell k})'$ are known vectors of real constants associated with the ℓ^{th} ratio, and $\boldsymbol{\mu} = (\mu_0, \mu_1, \ldots, \mu_k)'$. The likelihood ratio statistics to test the set of hypotheses in (3.5) are given by

$$T_\ell(\psi) = \frac{\mathbf{c}'_\ell \overline{\mathbf{Y}} - \psi \mathbf{d}'_\ell \overline{\mathbf{Y}}}{S[\psi^2 \mathbf{d}'_\ell \mathbf{M} \mathbf{d}_\ell - 2\psi \mathbf{c}'_\ell \mathbf{M} \mathbf{d}_\ell + \mathbf{c}'_\ell \mathbf{M} \mathbf{c}_\ell]^{\frac{1}{2}}} \sim t(\nu), \quad \ell = 1, \ldots, r, \tag{3.6}$$

where $\overline{\mathbf{Y}}$ is the maximum likelihood estimator of $\boldsymbol{\mu}$ and $\mathbf{M}$ is a diagonal matrix containing the reciprocals of the sample sizes. Jointly, the statistics

in (3.6) follows $Mt_r(\nu, \mathbf{R})$, where the elements of the correlation matrix are given by

$$\rho_{ij} = \frac{(\psi \mathbf{d}_i - \mathbf{c}_i)' \mathbf{M} (\psi \mathbf{d}_j - \mathbf{c}_j)}{\sqrt{(\psi \mathbf{d}_i - \mathbf{c}_i)' \mathbf{M} (\psi \mathbf{d}_i - \mathbf{c}_i)} \sqrt{(\psi \mathbf{d}_j - \mathbf{c}_j)' \mathbf{M} (\psi \mathbf{d}_j - \mathbf{c}_j)}}, \tag{3.7}$$

$1 \leq i \neq j \leq r$. Therefore, decisions about the tests in (3.5) can be based on the equicoordinate percentage point obtained from the above multivariate t-distribution.

Example 1. A three-arm non-inferiority trial discussed by Pigeot *et al.* (2003) is a one-sided test about a single ratio parameter $(\mu_2 - \mu_0)/(\mu_1 - \mu_0)$ which can be tested by using $r = 1$, $\boldsymbol{\mu} = (\mu_0, \mu_1, \mu_2)'$, $\mathbf{c}_\ell = (-1, 0, 1)'$, and $\mathbf{d}_\ell = (-1, 1, 0)'$ in the above general formulation.

Example 2. In comparisons with a control, the simultaneous tests described in (3.1) is a one-sided special case of the above formulation with $k = 3$, $\mathbf{d}_1 = \mathbf{d}_2 = \mathbf{d}_3 = (1, 0, 0, 0)'$, $\mathbf{c}_1 = (0, 1, 0, 0)'$, $\mathbf{c}_2 = (0, 0, 1, 0)'$, and $\mathbf{c}_3 = (0, 0, 0, 1)'$.

Chapter 4

Power and Sample Size Computations in Simultaneous Tests for Non-inferiority

4.1 Introduction

The goal of a non-inferiority trial is to test whether a given test treatment is no worse than a standard treatment in its efficacy. Such trials are becoming increasingly more popular as an alternative to placebo controlled clinical trials. If the use of a placebo group is unethical, an active competitor may be included against which non-inferiority has to be claimed. Such approach replaces the traditional superiority trials including a placebo arm

if well-established active competitors are available. In many therapeutic areas, such active competitors do exist and it may, therefore, not be sufficient to solely show superiority of the newly developed experimental treatment over placebo since its relative performance to the competitors on the market would remain uninvestigated. Another reason to conduct non-inferiority trials is the increased number of studies claiming a better safety profile of the experimental treatment over competing treatments while not being inferior in efficacy. In such instances, the new experimental treatment is shown to be safer than its competitors in which case a proven non-inferior efficacy justifies its potential release on the market.

In a special issue of Statistics in Medicine (2003, volume 22, issue 2), several statistical problems related to non-inferiority trials were discussed. Among others, D'Agostino *et al.* (2003) addressed the design concepts while Rashid (2003) dealt with non-parametric analysis, and Laster and Johnson (2003) discussed the use of ratio hypotheses. Ratio-based tests reformulate the standard hypotheses of differences, say for the efficacy parameters, in terms of relative effects which are particularly appealing in non-inferiority trials. The primary merit of this approach is that margins of non-inferiority (or equivalence) or superiority can be easily defined as a percentage of the unknown mean of the control treatment, particularly when the definition of the non-inferiority margin in an absolute term is difficult (Röhmel, 1998;

Hwang and Morikawa, 1999). Hauschke *et al.* (1999a) dealt with sample size calculations in test of equivalence based on ratio. In this problem, a single ratio is involved, but the nature of the problem leads to the computation of percentage points of a non-central bivariate t-distribution. Pigeot *et al.* (2003) considered the problem of comparing an experimental treatment for non-inferiority with a reference including a placebo arm. This problem is succinctly formulated as inference for a single ratio of linear combinations of the treatment means. They also derived formulas for determining power and sample sizes. Laster and Johnson (2003) described the ratio-based inference in detail and compared it with the classical testing approach due to Blackwelder (1982) which is based on the difference of means . In terms of the sample size, they conclude that under certain conditions testing for ratio of means is more efficient than testing the associated difference of means.

In this chapter, we extend the results from Laster and Johnson (2003) to the case of testing multiple treatments for non-inferiority against an active competitor. The problem of multiple ratios occurs, for example when several dose levels of a certain compound are assessed for their efficacy in comparison to an established treatment. Multiple testing for non-inferiority based on ratios was first addressed by Hauschke and Kieser (2001). In dose finding studies, Bretz *et al.* (2003) implemented a step-wise test procedure with associated confidence intervals to identify effective and/or safe doses based

on ratios. Tamhane *et al.* discussed determining sample sizes in the context of estimating the maximum safe dose.

Power or sample size formulas related to simultaneous testing of non-inferiority of several treatments against a common control, without prioritizing the hypothesis, are yet not available. This chapter aims at bridging this gap and we derive power and sample size formulas associated with the single-step multiple test procedure. Due to the inherent multiplicity aspect, different power concepts are available. They are thoroughly discussed in the context of non-inferiority testing and advice is given on how to proceed in practical situations. The inverse problem of determining the necessary sample size for a given power is also addressed and numerical comparisons related to the different power definitions are performed. In particular, we investigate how the required sample sizes for the ratio-based inference compare with that of the inference based on differences (or absolute margin) for various scenarios.

4.2 Simultaneous Tests for Non-inferiority

In this section, we consider a special case of the test described in the previous chapter with a restriction on the margin ψ. Let $\psi < 1$ denote the relative non-inferiority margin (in the case of large response values indicating better

treatment effects). Here we are interested in the one-sided tests

$$H_{0\ell} : \gamma_\ell \leq \psi \quad \text{against} \quad H_{1\ell} : \gamma_\ell > \psi, \quad \ell = 1, 2, \ldots, k. \tag{4.1}$$

The procedures for computing the critical point c and the decision rules are exactly as described in the previous chapter.

Some remarks on other cases or possible generalizations of the setup above are as follows. First, in the situation above where large response values indicate better treatment effects, the choice $\psi \geq 1$ corresponds to a test for superiority. Secondly, for the tests in (4.1), one may also use unequal non-inferiority margins ψ_ℓ, $\ell = 1, \ldots, k$, if such is of interest. In all cases, the derivations in the subsequent sections are equally applicable. We defer a detailed discussion for the case when small metrics indicate better treatment effects to Section 4.8.

4.3 Power Formulas

A major task in the design phase of a clinical study is that of determining sample sizes which guarantee a pre-specified power. In this section, we provide the power associated with the tests described in (4.1).

Let θ^* ($> \psi$) denote the greatest clinically (or biologically) irrelevant percentage of the control mean which is to be detected. Define the set of indices

$I(\theta^*) = \{\ell |\ \theta_\ell > \theta^*\} = \{\ell_1, \ldots, \ell_m\}$, $1 \leq m \leq k$. All treatments with θ_ℓ values greater than θ^* are non-inferior to the control. We consider the following two power definitions.

(i) Complete power. Suppose that the interest is to detect *all* non-inferior treatments with a given power of $1 - \beta$, where β is the size of type II error. The power associated with this problem is called complete (or all-pairs) power and it is given by

$$\begin{aligned} \pi_{\text{Com}}(\boldsymbol{\theta}, \theta^*) &= P\left\{T_\ell > c, \quad \text{for all } \ell \in I(\theta^*)\right\} \\ &= \int_0^\infty \int_{-\infty}^\infty \prod_{\ell \in I(\theta^*)} \Phi\left(-\frac{c\eta - \delta_\ell^{\text{ratio}} + \lambda_\ell z}{\sqrt{1-\lambda_\ell^2}}\right) \phi(z)\varphi(\eta) dz d\eta, \end{aligned} \tag{4.2}$$

where $\Phi(.)$ and $\phi(.)$, respectively denote the cumulative density function and the density function of the univariate standard normal distribution, and $\varphi(.)$ is the density function of $\left(\chi_\nu^2 / \nu\right)^{1/2}$.

(ii) Minimal power. Suppose that the interest is to detect *at least one* non-inferior treatment with a given power of $1 - \beta$. This is called minimal (or any-pair) power. The power of this test is given by

$$\begin{aligned} \pi_{\text{Min}}(\boldsymbol{\theta}, \theta^*) &= P\left\{T_\ell > c, \quad \text{for some } \ell \in I(\theta^*)\right\} \\ &= 1 - \int_0^\infty \int_{-\infty}^\infty \prod_{\ell \in I(\theta^*)} \Phi\left(\frac{c\eta - \delta_\ell^{\text{ratio}} + \lambda_\ell z}{\sqrt{1-\lambda_\ell^2}}\right) \phi(z)\varphi(\eta) dz d\eta. \end{aligned} \tag{4.3}$$

For other definitions of power in simultaneous testing (e.g., individual power and proportional power), we refer to Horn and Vollandt (1998) and Westfall *et al.* (1999).

We now return to the practical problem of determining the sample size associated with a given lower bound $1-\beta$ of the power. Note that all parameters of the distribution of $\mathbf{T}$ depend on the sample sizes for each treatment. For simplicity, we consider a balanced design with n observations per treatment. The required size n is determined iteratively by starting with a given sample size and search until the power condition is satisfied. That is, for minimal power, we look for the smallest n such that $\pi_{\text{Min}}(\boldsymbol{\theta}, \theta^*) \geq 1-\beta$ and similarly for complete power the smallest n for which $\pi_{\text{Com}}(\boldsymbol{\theta}, \theta^*) \geq 1-\beta$. A program that computes the necessary sample size to achieve a pre-specified power is given in Appendix A.

A more practical allocation is to consider the case where a different number of subjects is allocated to the control group than to the other treatment groups, i.e., $n_0 \neq n_1 = \ldots = n_k$. It is well known that in case of testing for hypotheses based on differences of means, the square-root allocation rule $n_0/n_\ell = \sqrt{k}$, $\ell = 1, \ldots, k$, is nearly optimal (Dunnett, 1955; Hochberg and Tamhane, 1987). If we use the same idea of minimizing $\text{Var}(\overline{Y_\ell} - \psi\overline{Y_0})$, $\ell = 1, \ldots, k$ subject to a fixed total sample size $N = \sum_{i=0}^{k} n_i$, then we get the solution $n_0 = \psi\sqrt{k}n_\ell$. The power behaviour of this allocation will be

discussed in later sections.

4.4 Comparison with Difference-Based Tests

In this section, we investigate the advantage of ratio-based inference over the classical difference-based inferences in a multiple testing situation. In case of a single ratio (comparing two treatments), Laster and Johnson (2003) showed that the ratio-based inference is more efficient as long as $\psi < 1$. That is, the sample size associated with the ratio-based inference is smaller than that of a comparable inference based on the difference of means in tests for non-inferiority. In multiple testing, the results based on the ratio view can also be compared with tests based on difference. To do this, we reformulate the tests in (4.1) as

$$H_{0\ell} : \mu_\ell - \mu_0 \leq \Delta_0 \quad \text{against} \quad H_{1\ell} : \mu_\ell - \mu_0 > \Delta_0, \quad \ell = 1, 2, \ldots, k, \quad (4.4)$$

where Δ_0 is the *absolute non-inferiority margin* and it is fixed as $\Delta_0 = (\psi - 1)\mu_0$. The likelihood ratio statistics to test the hypotheses in (4.4) are

$$T_\ell(\Delta_0) = \frac{\overline{Y_\ell} - \overline{Y_0} - \Delta_0}{S\sqrt{\frac{1}{n_0} + \frac{1}{n_\ell}}}, \quad \ell = 1, 2, \ldots, k.$$

Under the null hypotheses in (4.4), $T_\ell(\Delta_0)$ follows a central t-distribution with ν degrees of freedom. Jointly, $\mathbf{T}(\Delta_0) = (T_1(\Delta_0), \ldots, T_k(\Delta_0))'$ is distributed as a central k-variate t-distribution with ν degrees of freedom and

a correlation matrix $\mathbf{R}(\Delta_0) = [\rho_{ij}(\Delta_0)]$. The correlation $\rho_{ij}(\Delta_0)$ has also a product correlation structure and can be written as $\rho_{ij}(\Delta_0) = \lambda_i\lambda_j$, where $\lambda_i = (n_i/(n_0 + n_i))^{1/2}$. Unlike in the case of the ratio-based inference, note that the correlation matrix for the inference based on difference does not depend on the non-inferiority margin. If the design is balanced, we have $\rho_{ij}(\Delta_0) = 0.5$, $1 \leq i \neq j \leq k$. The $H_{0\ell}$ hypotheses in (4.4) will be rejected if $T(\Delta_0) > c$, where c is an equi-coordinate percentage point of $\mathbf{T}(\Delta_0)$. Now, to obtain a comparable power with that of the ratio-based inference, we set the vector of mean differences $(\mu_1 - \mu_0, \ldots, \mu_k - \mu_0)'$ to $\Delta = (\Delta_1, \ldots, \Delta_k)'$, where $\Delta_\ell = (\theta_\ell - 1)\mu_0$, $\ell = 1, \ldots, k$. Under the alternative hypotheses, $\mathbf{T}(\Delta_0)$ is distributed as $Mt_k(\nu, \mathbf{R}(\Delta_0), \boldsymbol{\delta}^{\text{diff}})$, where the elements of the non-centrality vector are given by

$$\delta_\ell^{\text{diff}} = \frac{\Delta_\ell - \Delta_0}{\sigma\sqrt{\frac{1}{n_\ell} + \frac{1}{n_0}}} = \frac{\theta_\ell - \psi}{CV_0\sqrt{\frac{1}{n_\ell} + \frac{1}{n_0}}}, \quad \ell = 1, \ldots, k.$$

Consider the minimal power $\pi_{\text{Min}}(\Delta, \Delta^*)$, where $\Delta^* = (\theta^* - 1)\mu_0$. The sample size n required per treatment (in a balanced design) is the smallest n such that $\pi_{\text{Min}}(\Delta, \Delta^*) \geq 1 - \beta$. It is observed that the power function increases as both the elements of the correlation matrix and the non-centrality parameters increase, and vice versa. For a given $\psi < 1$, CV_0 and $\boldsymbol{\theta}$, note that $\rho_{ij}(\Delta_0) > \rho_{ij}(\psi)$ but $\delta_\ell^{\text{diff}} < \delta_\ell^{\text{ratio}}$. Therefore, there is no easy analytical way of comparing the power functions of the ratio-based and difference-based

inferences as in the single-ratio case. However, from the plot of the power against the elements of the correlation matrix and the non-centrality parameters (not shown here), it is observed that the elements of the correlation matrix have little impact on the power compared with the impact of the non-centrality parameters. Thus, the difference in powers of the two approaches is mainly due to the differences in the non-centrality parameters. Let n_{ratio} and n_{diff} denote the number of observations required per treatment by the ratio and difference approaches, respectively. In Section 4.7, we show by various numerical examples that $n_{\text{ratio}} \leq n_{\text{diff}}$ if we have increasing effect (i.e., for the hypotheses in (4.1)) in tests for non-inferiority with $\psi < 1$. In the next section, we introduce least favourable configuration (LFC) and then compute the relative efficiency of the ratio-based test at LFC.

4.5 Least Favourable Configuration

As noted in Section 4.3, power computation in multiple testing relies on the knowledge about the configuration of the m true alternative hypotheses with $\theta_\ell > \theta^*$. Typically, the number m of true alternatives is not known in advance. One possibility is to evaluate the power at the least favorable configuration, i.e., at the parameter configuration under the alternative hypotheses at which the smallest power is attained. For multiple testing based

on the difference of location parameters, Horn and Vollandt (1998) derived LFCs for various power definitions. Along a similar line, we obtain LFCs associated with the ratio-based inference for complete and minimal power in one-sided tests for non-inferiority (or superiority). Suppose that a priori one knows upper and lower bounds on m, i.e., $m_1 \leq m \leq m_2$, where m_1 and m_2 are integers such that $1 \leq m_1 \leq m_2 \leq k$. We consider the case $n_0 \neq n_1 = \ldots = n_k = n$, that is $\lambda_\ell = \lambda$. From the expression for the complete power in (4.2), we see that

$$\begin{aligned}
\pi_{\mathrm{Com}}(\boldsymbol{\theta}, \theta^*) &= \int_0^\infty \int_{-\infty}^\infty \prod_{\ell \in I(\theta^*)} \Phi\left(\frac{(\theta_\ell - \psi)\mu_0}{\sigma}\sqrt{n} - \frac{c\eta + \lambda_\ell z}{\sqrt{1-\lambda_\ell^2}}\right) \phi(z)\varphi(\eta) dz d\eta, \\
&> \int_0^\infty \int_{-\infty}^\infty \Phi^m\left(\frac{(\theta^* - \psi)\mu_0}{\sigma}\sqrt{n} - \frac{c\eta + \lambda z}{\sqrt{1-\lambda^2}}\right) \phi(z)\varphi(\eta) dz d\eta, \\
&\geq \int_0^\infty \int_{-\infty}^\infty \Phi^{m_2}\left(\frac{(\theta^* - \psi)\mu_0}{\sigma}\sqrt{n} - \frac{c\eta + \lambda z}{\sqrt{1-\lambda^2}}\right) \phi(z)\varphi(\eta) dz d\eta, \\
&= P\left\{T_1 > c, \ldots, T_{m_2} > c \,|\, \theta_1 = \ldots = \theta_{m_2} = \theta^*\right\}.
\end{aligned}$$

Thus, if $m_1 \leq m \leq m_2$, a LFC for the complete power is $\theta_1 = \ldots = \theta_{m_2} = \theta^*$, and $\theta_\ell < \theta^*$ for $\ell > m_2$, . That is, if we compute the power at a LFC, for any other configuration of $\boldsymbol{\theta}$, the resultant power is larger than $1 - \beta$. Note that $\theta_{k-m_2+1} = \ldots = \theta_{k-1} = \theta_k = \theta^*$, and $\theta_\ell < \theta^*$ for $\ell \leq k - m_2$ is also a LFC. Therefore, the LFCs are permutation invariant. When there is no prior information about m (i.e., $m_1 = 1$ and $m_2 = k$), $\theta_1 = \ldots = \theta_k = \theta^*$ is a LFC.

In a similar manner, for the minimal power, it can be shown from (4.3) that

$$\pi_{\text{Min}}(\boldsymbol{\theta}, \theta^*) \geq 1 - P\left\{T_1 < c, \ldots, T_{m_1} < c \,|\, \theta_1 = \ldots = \theta_{m_1} = \theta^*\right\}.$$

Therefore, if a priori $m_1 \leq m \leq m_2$, then $\theta_1 = \ldots = \theta_{m_1} = \theta^*$, and $\theta_\ell < \theta^*$ for $\ell > m_1$, constitutes a LFC for the minimal power, and if there is no prior information about m, $\theta_1 = \theta^*$, and $\theta_\ell < \theta^*$ for $\ell > 1$ is a LFC .

4.6 Relative Efficiency of Ratio-Based Tests

The relative efficiency of ratio-based test is defined as the ratio of the sample size required by the ratio approach to the sample size required by a comparable difference-based test. Sample size can be determined either by using the exact central and non-cental t-distributions or by normal approximations to the t-distributions. In a balanced design, since both the number of degrees of freedom and the non-centrality parameter(s) depend on the sample size, the exact method involves a step-by-step calculation until the power condition is satisfied. The problem of this approach is that it can be time consuming since one has to compute and check the power for a wide range of sample sizes. An alternative and efficient approach is to use the corresponding normal approximations to the central and non-central t-distributions. Normal approximation also enable us to explicitly write an expression for sample size which is useful in practical calculations and it also facilitates the computation

of relative efficiencies.

Let $c_{d,\mathbf{R}(),1-\alpha} > 0$ denote a one-sided equi-coordinate percentage point of a central multivariate normal distribution of dimension d, a correlation matrix $\mathbf{R}()$, and a given familywise error rate of α. If there is no prior information about m (i.e., $1 \leq m \leq k$), given a minimal power of $1-\beta$, ψ, θ^* and CV_0, the required sample size by the ratio approach is

$$n_{\text{ratio}} = \left(c_{k,\mathbf{R}(\psi),1-\alpha} - Z_\beta\right)^2 \frac{1+\psi^2}{(\theta^* - \psi)^2} CV_0^2$$

where Z_β is the β^{th} quantile point of a univariate standard normal density. The corresponding sample size for the difference-based test is

$$n_{\text{diff}} = \left(c_{k,\mathbf{R}(\Delta_0),1-\alpha} - Z_\beta\right)^2 \frac{2}{(\theta^* - \psi)^2} CV_0^2.$$

Therefore, the relative efficiency of the ratio-based test is given by

$$\text{Rel.Eff.} = \frac{n_{\text{ratio}}}{n_{\text{diff}}} = \left(\frac{c_{k,\mathbf{R}(\psi),1-\alpha} - Z_\beta}{c_{k,\mathbf{R}(\Delta_0),1-\alpha} - Z_\beta}\right)^2 \frac{1+\psi^2}{2}. \tag{4.5}$$

In two-sample problems, Laster and Johnson (2003) has shown that Rel.Eff $= (1+\psi^2)/2$. This is in line with the above result for $k = 1$. In (4.5), for $2 \leq k \leq 5$, $1-\beta = 0.8$ and $\alpha = 0.05$, it is observed that

$$\left(\frac{c_{k,\mathbf{R}(\psi),1-\alpha} - Z_\beta}{c_{k,\mathbf{R}(\Delta_0),1-\alpha} - Z_\beta}\right)^2 \approx 1.$$

Thus, in these cases, the relative efficiency of the ratio-based approach is roughly $0.5+0.5\psi^2$. This means that in simultaneous tests for non-inferiority

with large response values indicating better treatment effects, the ratio-based inference requires smaller sample size if $\psi < 1$.

When controlling the minimal power, if a priori it is known that $m_1 \leq m \leq m_2$, then

$$\text{Rel.Eff.} = \left(\frac{c_{k,\mathbf{R}(\psi),1-\alpha} - c_{m_1,\mathbf{R}(\psi),\beta}}{c_{k,\mathbf{R}(\Delta_0),1-\alpha} - c_{m_1,\mathbf{R}(\Delta_0),\beta}} \right)^2 \frac{1+\psi^2}{2}.$$

Likewise, the relative efficiency for the complete power can be derived.

4.7 Numerical Study

In this section, we investigate the power and the associated sample sizes for both the ratio-based and difference-based inferences. Various scenarios of the coefficient of variation for the control group (CV_0) and specific ratio parameter configurations (θ) under the alternative hypotheses are considered. Suppose that large response values indicate better treatment effect. Figure 4.1 shows the minimal power function at LFC for the ratio-based test with three comparisons ($k = 3$) and $\psi = 0.8$. The figure compares the power functions for various sample sizes n in a balanced design. As one would expect, the power is an increasing function of the clinically irrelevant percentage θ^* and larger sample sizes lead to larger power values. The power at $\theta^* = \psi$ is α.

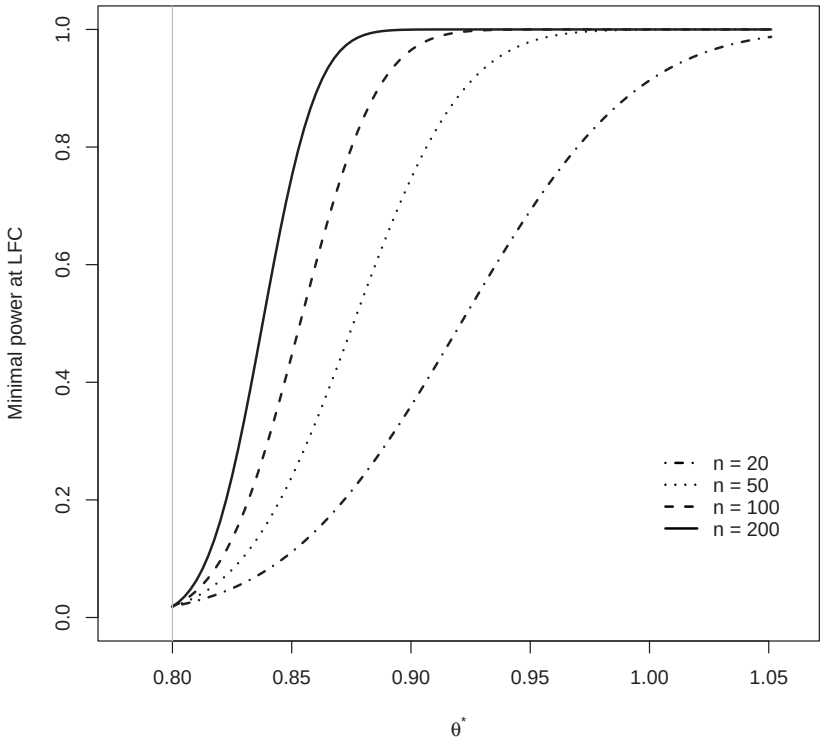

Figure 4.1: *Minimal power at LFC for various balanced sample sizes n when $k = 3$, $m_1 = 1$, $\psi = 0.8$, $CV_0 = 0.2$, and $\alpha = 0.05$.*

Figure 4.2 shows the minimal power differences at LFC between the ratio and difference approach $\pi_{\text{Min}}(\boldsymbol{\theta}, \theta^*) - \pi_{\text{Min}}(\Delta, \Delta^*)$ when there is no prior information about the correct configuration of the θ_ℓs for different number of comparisons k. From this figure, we see that the ratio-based is more powerful than the difference-based in tests for non-inferiority with $\psi < 1$ (assuming that large response values correspond to better treatment effects). This result

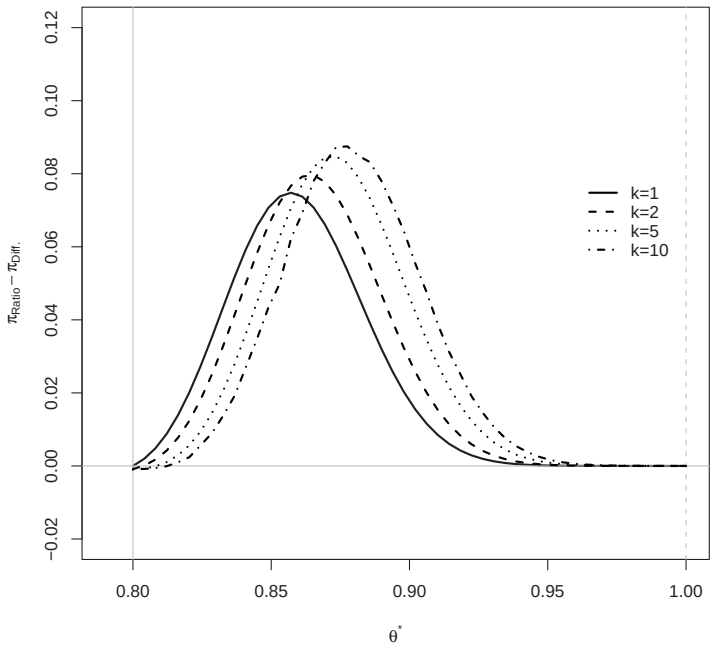

Figure 4.2: *Comparisons of the differences between the minimal power at LFC for the ratio-based and the difference-based tests when* $m_1 = 1$, $\psi = 0.80$, $CV_0 = 0.2$, *and* $n = 100$ *(balanced design).*

is in line with the results obtained by Laster and Johnson (2003) for a single ratio ($k = 1$). Figure 4.2, thus, indicates that this result also holds true for multiple ratios ($k > 1$). If the θ values fall far to the right of ψ, the power functions for both ratio and difference-based tests are close to one and the two approaches practically do not differ, i.e., the power difference is close to zero. Note that for the situation described above, when the non-inferiority margin is greater than 1, $\pi_{\text{Min}}(\boldsymbol{\theta}, \theta^*) < \pi_{\text{Min}}(\Delta, \Delta^*)$ for $\psi < \theta$, therefore,

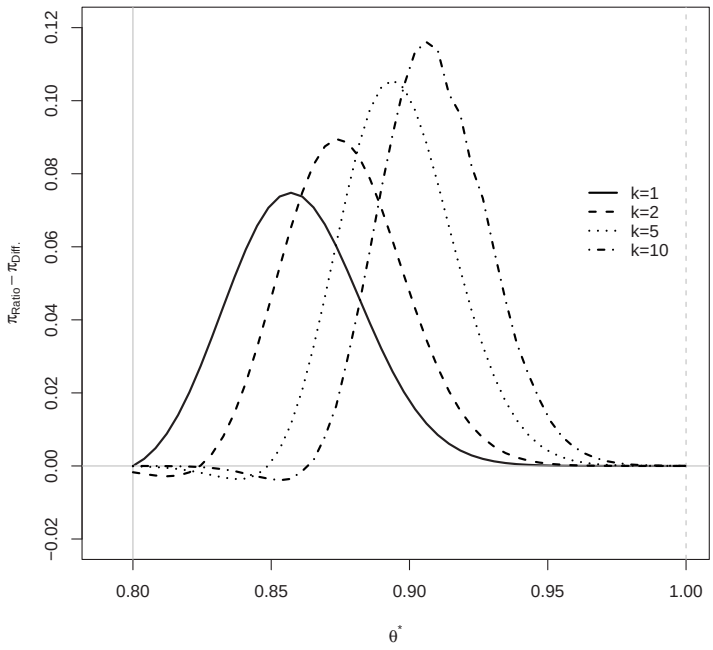

Figure 4.3: *Comparisons of the differences between the complete power at LFC for the ratio-based and the difference-based tests when* $m_2 = k$, $\psi = 0.80$, $CV_0 = 0.2$, *and* $n = 100$ *(balanced design).*

difference-based inferences lead to higher power when testing for superiority.

When interest lies in controlling the complete power, we have a slightly different situation in the power differences $\pi_{\text{Com}}(\boldsymbol{\theta}, \theta^*) - \pi_{\text{Com}}(\Delta, \Delta^*)$ at LFCs. As shown in Figure 4.3, the power of the difference-based testing is slightly greater than that of the ratio for θ^* values near the non-inferiority margin ψ. In other words, for a fixed very small complete power (often of no practical importance), n_{diff} can be smaller than n_{ratio}, even if $\psi < 1$.

Let us now consider the impact of different parameter constellations on the resulting sample sizes. Sample sizes are determined for LFC when no prior information concerning m is available (i.e., $m_1 = 1$ when controlling the minimal power and $m_2 = r$ for complete power). The computations are based on the exact non-central t-distribution, but normal approximation also yields very similar results (with a difference of 2 or 1 observations in few cases). Tables 4.1 and 4.2 consist of the sample sizes required for a given minimal power and complete power, respectively. From the tables, it can be seen that the sample size required for the minimal power (Table 4.1) is substantially smaller than that of the complete power (Table 4.2) as expected. In the tables, we compare the sample sizes required by the ratio-based inference (n_{ratio}) with that of the inferences based on difference of means (n_{diff}). To this end, the mean of the control group is fixed at an arbitrary value of μ_0 and the common standard deviation is fixed at $\sigma = CV_0\mu_0$. Tables 4.1 and 4.2 show that smaller sample sizes are associated with the ratio-based inference for the cases under investigation. Table 4.3 consists of the sample sizes required in test of superiority for the problem described above. The superiority margin is chosen to be $\psi = 1.2$. In this case, the sample size required for the ratio-based test is larger than that of the difference.

Comparing Tables 4.1 and 4.3, one also discerns the symmetry in the sample size required by the inference based on difference (n_{diff} is the same

Table 4.1: Test for non-inferiority: Sample size $n_{\text{ratio}}(n_{\text{diff}})$ based on minimal power for increasing effect ($k = 3$, $m_1 = 1$, $\psi = 0.80$, $\alpha = 0.05$)

		θ^*				
$CV_0(\%)$	$1-\beta$	0.85	0.90	0.95	1	1.10
10	0.75	51 (61)	14 (16)	7 (8)	4 (5)	3 (3)
	0.80	57 (68)	15 (18)	7 (9)	5 (5)	3 (3)
	0.90	75 (90)	20 (23)	9 (11)	6 (7)	3 (4)
	0.95	92 (111)	24 (28)	11 (13)	7 (8)	4 (4)
20	0.75	201 (241)	51 (61)	23 (28)	14 (16)	7 (8)
	0.80	226 (271)	57 (68)	26 (31)	15 (18)	7 (9)
	0.90	298 (359)	75 (90)	34 (41)	20 (23)	9 (11)
	0.95	366 (441)	92 (111)	42 (50)	24 (28)	11 (13)
50	0.75	1249 (1499)	313 (375)	140 (167)	79 (95)	36 (43)
	0.80	1404 (1687)	352 (423)	157 (188)	89 (106)	40 (48)
	0.90	1858 (2237)	465 (560)	207 (249)	117 (141)	53 (63)
	0.95	2281 (2749)	571 (688)	254 (306)	144 (173)	64 (77)

Table 4.2: Test for non-inferiority: Sample size $n_{\text{ratio}}(n_{\text{diff}})$ based on complete power for increasing effect ($k = 3$, $m_2 = 3$, $\psi = 0.80$, $\alpha = 0.05$)

		θ^*				
$CV_0(\%)$	$1-\beta$	0.85	0.90	0.95	1	1.10
10	0.75	73 (85)	19 (22)	9 (10)	6 (6)	3 (3)
	0.80	79 (93)	21 (24)	10 (11)	6 (7)	3 (4)
	0.90	98 (116)	25 (30)	12 (14)	7 (8)	4 (5)
	0.95	115 (137)	29 (35)	14 (16)	8 (10)	4 (5)
20	0.75	289 (339)	73 (85)	33 (38)	19 (22)	9 (10)
	0.80	315 (371)	79 (93)	36 (42)	21 (24)	10 (11)
	0.90	388 (461)	98 (116)	44 (52)	25 (30)	12 (14)
	0.95	456 (544)	115 (137)	52 (61)	29 (35)	14 (16)
50	0.75	1801 (2114)	451 (529)	201 (236)	113 (133)	51 (60)
	0.80	1962 (2312)	491 (579)	219 (258)	124 (145)	55 (65)
	0.90	2423 (2879)	606 (720)	270 (321)	152 (181)	68 (81)
	0.95	2843 (3395)	711 (849)	317 (378)	179 (213)	80 (95)

Table 4.3: Test for superiority: Sample size $n_{\text{ratio}}(n_{\text{diff}})$ based on minimal power for increasing effect ($k = 3$, $m_1 = 1$, $\psi = 1.20$, $\alpha = 0.05$)

		θ^*				
$CV_0(\%)$	$1-\beta$	1.25	1.30	1.35	1.40	1.50
10	0.75	73 (61)	19 (16)	9 (8)	6 (5)	3 (3)
	0.80	82 (68)	21 (18)	10 (9)	6 (5)	3 (3)
	0.90	109 (90)	28 (23)	13 (11)	8 (7)	4 (4)
	0.95	133 (111)	34 (28)	16 (13)	9 (8)	5 (4)
20	0.75	288 (241)	73 (61)	33 (28)	19 (16)	9 (8)
	0.80	325 (271)	82 (68)	37 (31)	21 (18)	10 (9)
	0.90	431 (359)	109 (90)	49 (41)	28 (23)	13 (11)
	0.95	531 (441)	133 (111)	60 (50)	34 (28)	16 (13)
50	0.75	1797 (1499)	450 (375)	201 (167)	113 (95)	51 (43)
	0.80	2025 (1687)	507 (423)	226 (188)	127 (106)	57 (48)
	0.90	2691 (2237)	673 (560)	300 (249)	169 (141)	76 (63)
	0.95	3312 (2749)	829 (688)	369 (306)	208 (173)	93 (77)

in Tables 4.1 and 4.3). This is not only a numerical finding but it is also theoretically expected (since the correlation matrix and the non-centrality parameters for the two tables are identical). For the ratio approach, this kind of symmetry does not hold true. Further, it may seem that one gets the same n_{ratio} in Table 4.3 as in Table 4.1 if the margin in Table 4.3 is chosen to be $1/0.8 = 1.25$. But this is again not the case as can be seen from the following example. In Table 4.1, $n_{\text{ratio}} = 75$ when $CV_0 = 0.2$, $1 - \beta = 0.90$, and $\theta^* = 0.90$. In test for superiority (like in Table 4.3) with $\psi = 1.25$, $CV_0 = 0.2$, $1 - \beta = 0.90$, and $\theta^* = 1.30$, the sample size required for the ratio-based is $n_{\text{ratio}} = 451$, which is quite different from 75.

In summary, for the ratio-based inference, a smaller sample size is required in tests for non-inferiority with $\psi < 1$ compared with the common difference-based inference.

4.8 Small Metrics Indicating Better Treatment Effects

In the case of small response values indicating better treatment effects, the choice of $\psi < 1$ leads to test for superiority while $\psi > 1$ is test for non-inferiority. Therefore, in simultaneous tests for non-inferiority with small

response values indicating better treatment benefits, the hypotheses to be tested are

$$H_{0\ell} : \gamma_\ell \geq \psi \quad \text{against} \quad H_{1\ell} : \gamma_\ell < \psi, \quad \ell = 1, 2, \ldots, r, \tag{4.6}$$

where now the non-inferiority margin $\psi > 1$. In this case, we decide the ℓ^{th} treatment to be non-inferior to the control if $T_\ell(\psi) < -c$. The critical point c is computed as in equation (3.4) by using the value of ψ fixed in (4.6). From the symmetry of the distribution of $T_\ell(\psi)$ under the null, it can be shown that the power behaviors of the problem in (4.6) are exactly the same as the power behaviors of simultaneous tests for superiority when large response values indicate better treatment effect and $\psi > 1$. The sample sizes associated with the latter scenario is given in Table 4.3. Thus, there is no sample size advantage in simultaneous tests for non-inferiority with $\psi > 1$. In order to maintain the sample size advantage of the ratio view, if it gives sense, one may make inference about the ratio of control mean to that of the test treatments (as suggested by Laster and Johnson (2003) for a two-sample problem). Thus, if we invert the ratios in (4.6), the hypotheses to be tested are

$$H_{0\ell} : \mu_0 / \mu_\ell \leq \psi_1 \quad \text{against} \quad H_{1\ell} : \mu_0 / \mu_\ell > \psi_1, \quad \ell = 1, 2, \ldots, r, \tag{4.7}$$

where $\psi_1 < 1$. The associated joint distribution of the test statistics to test the hypotheses in (4.7) has a multivariate t-distribution with off-diagonal

elements of the correlation matrix given by

$$\rho_{ij}(\psi_1) = \frac{1}{\sqrt{1+\frac{n_0}{n_i}\psi_1^2}}\frac{1}{\sqrt{1+\frac{n_0}{n_j}\psi_1^2}}, \quad 1 \leq i \neq j \leq r.$$

For power computations, the corresponding non-centrality parameters are given by

$$\delta_\ell^{\text{ratio}}(\psi_1) = \frac{1-\psi_1(\mu_0\,/\mu_\ell)^{-1}}{CV_0\sqrt{\frac{1}{n_0}+\frac{1}{n_\ell}\psi_1^2}}.$$

Now, for comparing the power associated with the tests in (4.7) with that of the difference-based test, we consider the hypotheses

$$H_{0\ell}: \mu_0 - \mu_\ell \leq \Delta_\ell \quad \text{against} \quad H_{1\ell}: \mu_0 - \mu_\ell > \Delta_\ell, \quad \ell = 1, 2, \ldots, r,$$

where $\Delta_\ell = (\psi_1 - 1)\mu_\ell$. Here, we encounter *varying absolute non-inferiority margins* which depend on the mean of the new treatments. For these choices of the delta-margins, the corresponding ratio-formatted tests are more powerful. It might be more desirable to have identical absolute delta-margins across the comparisons. In this case, from the relationship between μ_0 and μ_ℓ on the boundaries of the $H_{0\ell}$ hypotheses in (4.7), we can write $\mu_\ell = \mu_0\,/\psi_1$. Substituting this in $\Delta_\ell = (\psi_1 - 1)\mu_\ell$, we get another delta-margin of $(1 - 1\,/\psi_1)\mu_0$. Often, there exists more prior information about the standard treatment (the control) than the new treatments. Thus, the latter approach seems to be a more practical way of choosing the delta-margins. However, for this second choice of the delta-margins, the difference-based test is more efficient.

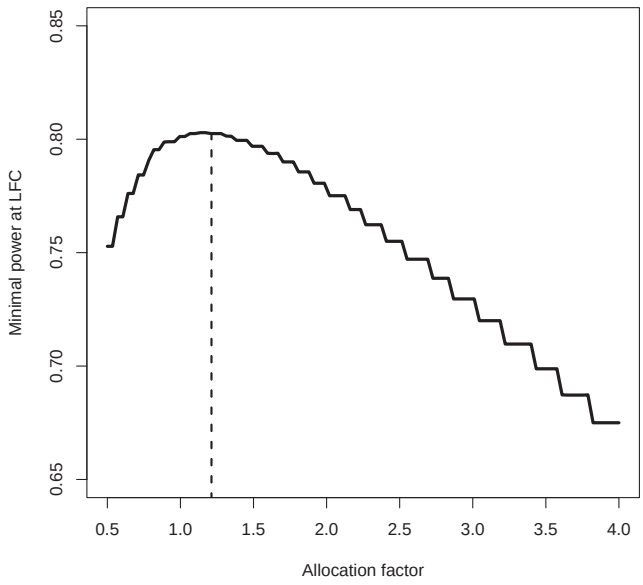

Figure 4.4: *Minimal power at LFC versus allocation factor* (n_0/n_ℓ) *for a fixed total sample size* $N = 208$. *The vertical dotted line is at the allocation factor* $\psi\sqrt{k} = 1.212$.

4.9 Examples

Now, let us determine the sample sizes required for two of the data examples in Chapter 2. All computations are carried out for the least favourable configuration.

Example 1. For the osteoporosis study described in Section 2.3, we compute the sample size associated with a non-inferiority margin of $\psi = 0.70$,

a minimal power of 80%, $\alpha = 0.05$, and $CV_0 = 0.50$. We further assume that the clinically irrelevant percentage of the control mean which is to be detected is $\theta^* = 0.95$. Since there is no prior information about m, a LFC is $\theta_1 = 0.95$, $\theta_2, \theta_3 < 0.95$. Under these conditions, for the one-sided simultaneous test of three treatment schedules versus an active control, the required number of observations per treatment is $n_{\text{ratio}} = 52$ ($n_{\text{diff}} = 68$) in a balanced design. For the ratio-based testing, if we use the allocation rule $n_0 = \psi\sqrt{k}n_\ell$ with a fixed total sample size of $N = 4 \times 52 = 208$, the number of observations required per treatment are $n_0 = 60$ and $n_1 = n_2 = n_3 = 50$. The power of this allocation is 0.807. The graph of power versus other allocation factors (n_0/n_ℓ) is shown in Figure 4.4. The power attains its maximum at the allocation factor $0.7 \times \sqrt{3}$ (the vertical dotted line in Figure 4.4). Note that this holds true when $m_1 = 1$. If $m_1 > 1$ or if the interest is to control the complete power (with $m_2 = 3$), the maximum power over all possible allocations is slightly greater than the power at the allocation $0.7 \times \sqrt{3}$. For the difference-based inference, using the allocation rule $n_0 = \sqrt{k}n_\ell$ with a fixed total sample size of $N = 4 \times 68 = 272$, the number of observations required per treatment are $n_0 = 100$ and $n_1 = n_2 = n_3 = 58$.

Example 2. The clinical data example described in Section 2.4 is the case when small response values indicate better treatment effect. The superior-

ity margin is $\psi = 0.90$, the given minimal power is 0.80, $\alpha = 0.025$, and $CV_0 = 0.17$. Let the clinically relevant percentage of the control to be detected be $\theta^* = 0.85$. Since there is no prior information about m, a LFC is $\theta_1 = 0.85$, $\theta_2, \theta_3 > 0.85$. Thus, the number of observations required per treatment is $n_{\text{ratio}} = 215$ ($n_{\text{diff}} = 237$) in a balanced design. Therefore, in terms of sample size, the ratio approach is more efficient in designing a new confirmatory clinical trial. For the same trial, if one wishes to control the complete power, the LFC is $\theta_1 = 0.85$, $\theta_2 = 0.85$, $\theta_3 = 0.85$, and the required sample size for each treatment is $n_{\text{ratio}} = 290$ ($n_{\text{diff}} = 315$).

We remark that when controlling minimal power with the prior information that $m_1 > 1$, the power at the allocation factor $\psi\sqrt{k}$ is slightly smaller than the maximum power over all possible allocations. In this case, it is again possible to determine the optimum sample sizes (associated with the maximum power) iteratively by first finding the optimum allocation factor.

4.10 Discussion

In this chapter, we considered the problem of sample size and power computations in simultaneous tests for non-inferiority based on the ratio view. The efficiency of this approach is also compared with that of tests based on the difference of location parameters. From the various numerical stud-

ies, the following results are observed. The ratio approach has advantage in one-sided tests, namely, (i) in tests for non-inferiority with relative margin less than one and when large response values indicate better treatment effect and (ii) in tests for superiority with the superiority margin less than one and when small response values indicate better treatment effect. The latter case is illustrated using a clinical data example (Example 2). It is not directly investigated but it can be shown analogously to test for non-inferiority. This generalizes the results of Laster and Johnson (2003) from two-sample to multiple-sample designs. In some cases, the reduction in sample size can be clinically relevant. For instance, for the osteoporosis study, we have about 25% reduction in the number of observations per treatment by applying the ratio approach.

Therefore, from the perspective of higher power and problem-adequate interpretation, ratio-based multiple testing can be recommended for selected non-inferiority (or superiority) trials when the interest is to control the minimal power. The related R code for the design is provided in the Appendix.

Chapter 5

Simultaneous Confidence Sets and Confidence Intervals

5.1 Introduction

A well-known theorem by Fieller (1954) provides a method for constructing confidence interval for ratio of bivariate normal means. Various extensions and the characteristics of the Fieller solution are studied by Cox (1967), Steffens (1971), Buonaccorsi (1985), Koschat (1987), and others. The focus of this chapter is that of constructing simultaneous confidence sets (SCS) and simultaneous confidence intervals (SCI) for multiple ratios. In the statistical literature, this problem is also addressed by a number of researchers. Scheffé (1970) showed that estimation of ratios is related to estimation of an

unoriented direction of a vector, and thereby proposed a method for joint estimation of multiple ratios based on the projection method. Zerbe *et al.* (1982) applied the Scheffé method to ratios of linear combinations of the coefficients in the general linear model. Young *et al.* (1997) further extended the Scheffé method to the case of ratios of parameters of linear and non-linear mixed models. Bennett (1961) dealt with the problem of constructing confidence intervals for the common ratio of means of several bivariate normal distributions. Malley (1982) dealt with the case of multivariate observations and several ratios and provides a method for constructing SCI for multiple ratios for various scenarios of the covariance matrix. Among others, Malley (1982) also proposed the use of a Bonferroni correction for the simultaneous estimation of ratios. Jensen (1989) used a critical point which is derived on the basis of Šidák (1967) inequality, and discussed its applications in direct, parallel-line and slope-ratio assays. In sequential analysis, Hwang and Liu (1990) proposed confidence sequences for multiple ratios which are analogous to Scheffé's simultaneous confidence intervals. They also give sharper confidence sequences. The primary objective is to propose some alternatives to the existing conservative procedures in non-sequential setting.

In this chapter, first we describe a new method of constructing SCS based on the multivariate t-distribution which controls the confidence level exactly. Secondly, we review the existing methods for constructing SCI (Bonferroni,

Šidák and Scheffé) which approximately control the confidence level and then propose two other methods (plug-in and resampling). The methods are compared for a variety of data sets. Monte Carlo simulation is also used to compare the performance of the methods with respect to estimates of the critical point and the coverage probabilities. Accordingly, the chapter is organized as follows. In Section 5.2, we define the problem as applied to the one-way ANOVA model. The general problem of constructing SCS and SCI for multiple ratios are introduced in Sections 5.3 and 5.4. Section 5.5 deals with a specific area of application, the many-to-one comparisons. Various numerical examples are also provided in this section. A simulation study based on the many-to-one comparisons is given in Section 5.6. Finally, Section 5.7 is devoted to concluding remarks.

5.2 Multiple Ratios

Suppose that we have k treatments. Let Y_{ij} denote the j^{th} independent observation under the i^{th} treatment, $i = 1, \ldots, k$; $j = 1, \ldots, n_i$. We consider the one-way ANOVA model with $Y_{ij} \sim \mathcal{N}(\mu_i, \sigma^2)$. Let $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_k)'$ be the vector of treatment means. The aim is to develop simultaneous confidence sets and confidence intervals for ratios of linear combinations of $\boldsymbol{\mu}$. We are

interested in the vector of parameters $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_r)'$, where

$$\gamma_\ell = \frac{\sum_{i=1}^{k} c_{\ell i}\mu_i}{\sum_{i=1}^{k} d_{\ell i}\mu_i} = \frac{\mathbf{c}_\ell'\boldsymbol{\mu}}{\mathbf{d}_\ell'\boldsymbol{\mu}}, \quad \ell = 1, \ldots, r, \tag{5.1}$$

and r is the number of ratios. The vectors $\mathbf{c}_\ell = (c_{\ell 1}, \ldots, c_{\ell k})'$ and $\mathbf{d}_\ell = (d_{\ell 1}, \ldots, d_{\ell k})'$ are known vectors of real constants associated with the ℓ^{th} ratio.

A key step in the derivation of confidence intervals for ratios is expressing the ratio problem as a linear form $L_\ell = (\gamma_\ell \mathbf{d}_\ell - \mathbf{c}_\ell)'\overline{\mathbf{Y}}$, $\ell = 1, \ldots, r$, where $\overline{\mathbf{Y}} = (\overline{Y}_1, \ldots, \overline{Y}_k)'$ is the maximum likelihood estimator of $\boldsymbol{\mu}$. More details on this approach can also be found in Fieller (1954) and Zerbe *et al.* (1982). Clearly, L_ℓ is distributed as $\mathcal{N}(0, \sigma^2_{L_\ell})$, where

$$\sigma^2_{L_\ell} = \text{Var}(L_\ell) = \sigma^2(\gamma_\ell \mathbf{d}_\ell - \mathbf{c}_\ell)'\mathbf{M}(\gamma_\ell \mathbf{d}_\ell - \mathbf{c}_\ell),$$

and $\text{Var}(\overline{\mathbf{Y}}) = \sigma^2\mathbf{M}$, with $\mathbf{M}$ being a diagonal matrix containing the reciprocals of the sample sizes. Let S^2 be the unbiased pooled variance estimator of the common variance σ^2 based on $\nu = \sum_{i=1}^{k}(n_i - 1)$ degrees of freedom. Denote an estimator of $\sigma^2_{L_\ell}$ by $S^2_{L_\ell}$, where $S^2_{L_\ell}$ is obtained from $\sigma^2_{L_\ell}$ by replacing σ^2 by S^2. Since S_{L_ℓ} is distributed as $\left(\sigma^2_{L_\ell}\nu^{-1}\chi^2(\nu)\right)^{\frac{1}{2}}$ independent of L_ℓ, the statistic $T_\ell(\gamma_\ell) = L_\ell / S_{L_\ell}$ follows a t-distribution with ν degrees of freedom. For notational convenience, note that this test statistic is chosen to be the negative of the one defined in Chapter 3 (compare with Equation 3.6). Jointly, the random vector $\mathbf{T} = (T_1, \ldots, T_r)'$ follows $Mt_r(\nu, \mathbf{R})$, where

the elements of $\mathbf{R}$ are given by

$$\rho_{ij} = \frac{(\gamma_i \mathbf{d}_i - \mathbf{c}_i)' \mathbf{M} (\gamma_j \mathbf{d}_j - \mathbf{c}_j)}{\sqrt{(\gamma_i \mathbf{d}_i - \mathbf{c}_i)' \mathbf{M} (\gamma_i \mathbf{d}_i - \mathbf{c}_i)} \sqrt{(\gamma_j \mathbf{d}_j - \mathbf{c}_j)' \mathbf{M} (\gamma_j \mathbf{d}_j - \mathbf{c}_j)}}, \tag{5.2}$$

$1 \leq i \neq j \leq r$. It can be shown that $\text{Corr}(Z_i, Z_j) = \rho_{ij}$, where $Z_i = L_i / \sigma_{L_i}$. In Section 5.3, we shall use the distribution of $\mathbf{T}$ as a basis for deriving confidence sets for $\boldsymbol{\gamma}$. One particular challenge is the dependence of $\mathbf{R}$ on $\boldsymbol{\gamma}$ which is the object of estimation. It is seen from (5.2) that the correlation ρ_{ij} is a function of the *unknown* ratios γ_i and γ_j, say $\rho_{ij} = h(\gamma_i, \gamma_j)$.

Before taking up the simultaneous inference for multiple ratios, we remark on some special cases of the above problem. For $r = 1$, we have a single ratio, say $\gamma = \mathbf{c}'\boldsymbol{\mu} / \mathbf{d}'\boldsymbol{\mu}$. A confidence interval for γ can be constructed using Fieller's theorem. A two-sided $(1 - \alpha)100\%$ confidence interval for γ is the solution in γ of the inequality

$$|T(\gamma)| = \frac{|(\gamma \mathbf{d} - \mathbf{c})' \overline{\mathbf{Y}}|}{S[\gamma^2 \mathbf{d}' \mathbf{M} \mathbf{d} - 2\gamma \mathbf{c}' \mathbf{M} \mathbf{d} + \mathbf{c}' \mathbf{M} \mathbf{c}]^{\frac{1}{2}}} \leq t_{1-\frac{\alpha}{2}}(\nu), \tag{5.3}$$

where $t_{1-\frac{\alpha}{2}}(\nu)$ is the $(1-\frac{\alpha}{2})^{\text{th}}$ quantile point of a t-distribution with ν degrees of freedom. The inequality in (5.3) can be expressed as a quadratic inequality in γ

$$A\gamma^2 + B\gamma + C \leq 0, \tag{5.4}$$

where $A = (\mathbf{d}'\overline{\mathbf{Y}})^2 - t^2 S^2 \mathbf{d}' \mathbf{M} \mathbf{d}$, $B = -2\left[(\mathbf{c}'\overline{\mathbf{Y}})(\mathbf{d}'\overline{\mathbf{Y}}) - t^2 S^2 \mathbf{c}' \mathbf{M} \mathbf{d})\right]$, $C = (\mathbf{c}'\overline{\mathbf{Y}})^2 - t^2 S^2 \mathbf{c}' \mathbf{M} \mathbf{c}$ and $t = t_{1-\frac{\alpha}{2}}(\nu)$. Depending on the values of the leading

coefficient A and the discriminant $B^2 - 4AC$, there are three possible solutions to the inequality in (5.4)(see, for example, Buonaccorsi and Iyer, 1984; Kendall (1999)). If $A > 0$, then it can be shown, that also $B^2 - 4AC > 0$, and the solution is a finite interval lying between the two roots. This is the most desirable situation. The other two cases result in either a region containing all values lying outside the finite interval defined by the two roots or even the entire γ-axis. If $\mathbf{d}'\boldsymbol{\mu}$ is significantly different from 0, the last two cases occur only with small probability. The condition $A > 0$ can equivalently be expressed as $g < 1$, where

$$g = \frac{t^2 S^2 \mathbf{d}'\mathbf{M}\mathbf{d}}{(\mathbf{d}'\overline{\mathbf{Y}})^2}. \tag{5.5}$$

Note that $S\sqrt{\mathbf{d}'\mathbf{M}\mathbf{d}}\,/\mathbf{d}'\overline{\mathbf{Y}}$ is an estimator of $\sigma\sqrt{\mathbf{d}'\mathbf{M}\mathbf{d}}\,/\mathbf{d}'\boldsymbol{\mu}$, the coefficient of variation of $\mathbf{d}'\overline{\mathbf{Y}}$.

Now for the general ratio problem in (5.1), in order to guarantee $A_\ell > 0$ (for the ℓ^{th} ratio) with high probability, we should require that

$$0 < \frac{q\sigma\sqrt{\mathbf{d}_\ell'\mathbf{M}\mathbf{d}_\ell}}{\mathbf{d}_\ell'\boldsymbol{\mu}} \ll 1 \quad \text{or} \quad \frac{\mathbf{d}_\ell'\boldsymbol{\mu}}{q\sigma\sqrt{\mathbf{d}_\ell'\mathbf{M}\mathbf{d}_\ell}} \gg 1, \tag{5.6}$$

where q is some relevant critical point. The direction of the inequality in (5.6) follows from the assumption that $\mathbf{d}_\ell'\boldsymbol{\mu} > 0$. As explained after (5.4), for a single ratio, there are three types of solutions depending on the value of g and the discriminant. For multiple ratios, many combinations of these types exist and it is difficult to fully describe the geometry of all such regions. Some

examples of unbounded two-sided SCS are shown in Appendix C. This is the reason for demanding the constraint in (5.6). In this case, the probability of the event $g_\ell > 1$ is small and ignorable. Later on, we shall see that $q\sigma\sqrt{\mathbf{d}_\ell'\mathbf{M}\mathbf{d}_\ell}\,/\mathbf{d}_\ell'\boldsymbol{\mu}$ is a key quantity dictating the geometric form of SCS for multiple ratios.

A major challenge in ratio estimations is the case of non-significant denominators ($g > 1$). If the probability of getting $g > 1$ is high, one may (i) change the design of the experiment such that $\mathbf{d}_\ell'\mathbf{M}\mathbf{d}_\ell$ will be smaller, (ii) estimate the reciprocal of the ratio if it gives sense (with appropriate inverse interpretation), or (iii) if feasible, apply Bayesian methods as proposed by Buonaccorsi and Gatsonis (1988).

5.3 Simultaneous Confidence Sets

Let $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_r)'$ be a point in the parameter space of $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_r)'$. Suppose that the interest is to construct a lower $(1-\alpha)100\%$ SCS for $\boldsymbol{\gamma}$. The usual way of determining a confidence set is to consider the test problems

$$H_{0\ell} : \gamma_\ell = \theta_\ell \quad \text{against} \quad H_{1\ell} : \gamma_\ell < \theta_\ell \tag{5.7}$$

with respect to the unknown parameters γ_ℓ and the constants $\theta_\ell, \ell = 1, \ldots, r$. If inequality (5.6) is true, then it is reasonable to apply

$$T_\ell(\theta_\ell) = \frac{\mathbf{c}_\ell'\overline{\mathbf{Y}} - \theta_\ell \mathbf{d}_\ell'\overline{\mathbf{Y}}}{S[\theta_\ell^2 \mathbf{d}_\ell'\mathbf{M}\mathbf{d}_\ell - 2\theta_\ell \mathbf{c}_\ell'\mathbf{M}\mathbf{d}_\ell + \mathbf{c}_\ell'\mathbf{M}\mathbf{c}_\ell]^{\frac{1}{2}}}$$

as a test statistic for testing (5.7). This means that $H_{0\ell}$ will be rejected if $T(\theta_\ell)$ exceeds a suitable significance threshold. Now a vector $\boldsymbol{\theta}$ belongs to the confidence set if and only if for this vector the null hypothesis $H_0 : \bigcap_{\ell=1}^{r} H_{0\ell}$ is accepted. Therefore, for a given sample, the collection of all such points constitutes a SCS for $\boldsymbol{\gamma}$. To test H_0 against the alternative hypothesis $H_1 : \bigcup_{\ell=1}^{r} H_{1\ell}$, we employ the union-intersection principle due to Roy (1953) which accepts H_0 if all $H_{0\ell}$'s are accepted. Thus, using the test statistics $T(\theta_\ell)$, for a lower $(1-\alpha)100\%$ SCS, we get the following definition.

Definition (*Simultaneous confidence sets*). Let $c_{1-\alpha}(\mathbf{R}(\boldsymbol{\theta}))$ denote a one-sided equicoordinate critical point of $Mt_r(\nu, \mathbf{R}(\boldsymbol{\theta}))$. A $(1-\alpha)100\%$ SCS for $\boldsymbol{\gamma}$ is defined by the set

$$\{\boldsymbol{\theta} : \quad T_\ell(\theta_\ell) \leq c_{1-\alpha}(\mathbf{R}(\boldsymbol{\theta})), \quad \ell = 1, \ldots, r\}. \tag{5.8}$$

We write $\mathbf{R}(\boldsymbol{\theta})$ to indicate the dependence of $\mathbf{R}$ on $\boldsymbol{\theta}$. Under H_0, $c_{1-\alpha}(\mathbf{R}(\boldsymbol{\theta}))$ satisfies

$$P\{T_\ell(\theta_\ell) \leq c_{1-\alpha}(\mathbf{R}(\boldsymbol{\theta})), \quad \ell = 1, \ldots, r\} = 1 - \alpha.$$

We refer to this confidence set as exact SCS in the sense that exact probability equality is inverted to obtain the set. In the same manner, a two-sided $(1-\alpha)100\%$ SCS for $\boldsymbol{\gamma}$ is defined as

$$\{\boldsymbol{\theta}: \quad |T_\ell(\theta_\ell)| \leq c'_{1-\alpha}(\mathbf{R}(\boldsymbol{\theta})), \quad \ell = 1, \ldots, r\}, \tag{5.9}$$

where $c'_{1-\alpha}(\mathbf{R}(\boldsymbol{\theta}))$ denotes the two-sided equicoordinate critical point of $Mt_r(\nu, \mathbf{R}(\boldsymbol{\theta}))$.

From (5.8) and (5.9), it is apparent that the critical points of $\mathbf{T}$ depend on $\boldsymbol{\theta}$ through the correlation matrix $\mathbf{R}$. Therefore, there is no easy way of explicitly determining the boundary of the confidence set. In the following sections, two approaches of constructing SCS are discussed.

5.3.1 Iterative Approach

In general, the iterative approach consists of a step-by-step method of estimating the unknown parameters (one conditioned over the other in turn) and determining the boundaries of the confidence set by starting from some initial values. We consider the simplest case of $r = 2$ to describe the method. The algorithm below is developed from the discussions of Guiard (2002) in comparisons with a control. Now, suppose that the interest is to construct the lower $(1-\alpha)100\%$ SCS for $\boldsymbol{\gamma} = (\gamma_1, \gamma_2)' = (\mathbf{c}'_1\boldsymbol{\mu} / \mathbf{d}'_1\boldsymbol{\mu}\,, \mathbf{c}'_2\boldsymbol{\mu} / \mathbf{d}'_2\boldsymbol{\mu})'$. The steps are:

Case 1. Estimating upper confidence limits for γ_1 conditional on γ_2 values.

Step 1 Initiate the parameters: Let $\gamma_1^{(0)}$ be the initial value for γ_1 and let Γ_2 be a set of fine grid points for γ_2.

Step 2 For a given $\theta_2 \in \Gamma_2$, compute the correlation and hence the quantile of $\mathbf{T}$. The correlation matrix at the j^{th} iteration is given by

$$\mathbf{R}^{(j)} = \begin{bmatrix} 1 & \rho_{12}^{(j)} \\ \rho_{12}^{(j)} & 1 \end{bmatrix}$$

where $\rho_{12}^{(j)} = h(\gamma_1^{(j)}, \theta_2)$. Compute the quantile $c_{1-\alpha}^{(j)}$ of $\mathbf{T}$ such that

$$P\left\{T_1 \leq c_{1-\alpha}^{(j)}, T_2 \leq c_{1-\alpha}^{(j)}\right\} = 1 - \alpha.$$

Step 3 Compute the boundary (upper confidence limit for γ_1) as

$$\gamma_1^{(j+1)} = \frac{-B_1^{(j)} + \sqrt{\left(B_1^{(j)}\right)^2 - 4A_1^{(j)}C_1^{(j)}}}{2A_1^{(j)}},$$

where

$$\begin{aligned} A_1^{(j)} &= (\mathbf{d}_1'\overline{\mathbf{Y}})^2 - (c_{1-\alpha}^{(j)})^2 S^2 \mathbf{d}_1'\mathbf{M}\mathbf{d}_1, \\ B_1^{(j)} &= -2\left[(\mathbf{c}_1'\overline{\mathbf{Y}})(\mathbf{d}_1'\overline{\mathbf{Y}}) - (c_{1-\alpha}^{(j)})^2 S^2 \mathbf{c}_1'\mathbf{M}\mathbf{d}_1)\right], \\ C_1^{(j)} &= (\mathbf{c}_1'\overline{\mathbf{Y}})^2 - (c_{1-\alpha}^{(j)})^2 S^2 \mathbf{c}_1'\mathbf{M}\mathbf{c}_1. \end{aligned}$$

Step 4 Repeat steps 2 and 3 until convergence, i.e., $\left|c_{1-\alpha}^{(j+1)} - c_{1-\alpha}^{(j)}\right| < \epsilon$, for some pre-specified accuracy $\epsilon > 0, \quad j = 0, 1, 2, \ldots$

Step 5 Do steps 2 to 4 for all $\theta_2 \in \Gamma_2$. If $\gamma_1(\theta_2)$ is the confidence limit for γ_1 given θ_2, then use this limit as an initial value if $\gamma_1(\theta_2')$ will be searched for a neighbouring value $\theta_2' \in \Gamma_2$.

Step 6 Sketch the upper confidence limits for γ_1 at the points of convergence versus Γ_2.

Case 2. Estimating upper confidence limits for γ_2 conditional on γ_1 values.

This case is exactly the same as Case 1, except that we interchange the role of γ_1 and γ_2. Thus, the correlation and the upper confidence limits are going to be updated as

$$\rho_{12}^{(j)} = h(\theta_1, \gamma_2^{(j)})$$

$$\gamma_2^{(j+1)} = \frac{-B_2^{(j)} + \sqrt{(B_2^{(j)})^2 - 4A_2^{(j)}C_2^{(j)}}}{2A_2^{(j)}}, \quad j = 0, 1, 2, \ldots$$

until convergence. The quantities $A_2^{(j)}$, $B_2^{(j)}$ and $C_2^{(j)}$ are the coefficients of the quadratic equation associated with γ_2 at the j^{th} iteration.

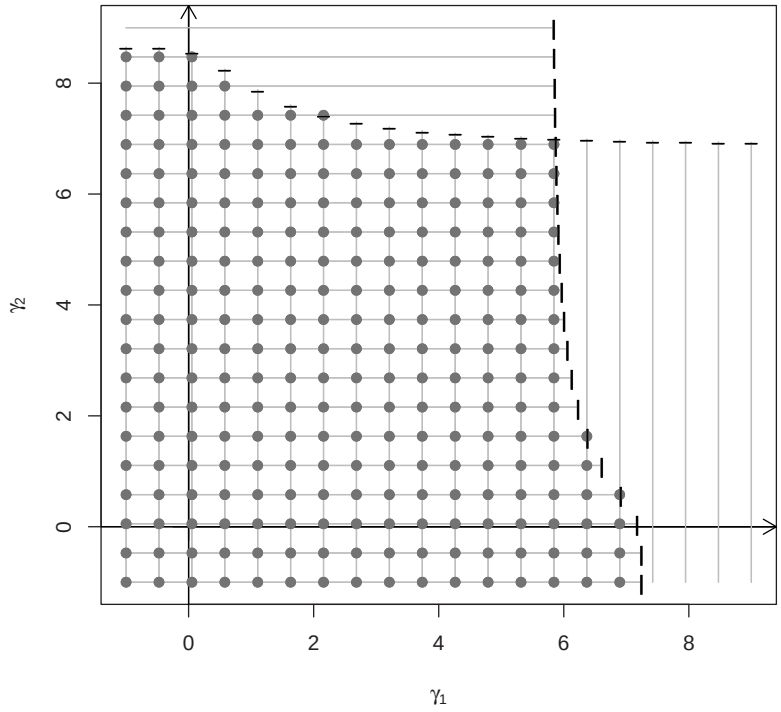

Figure 5.1: *Comparison of the iterative and pointwise testing for constructing SCS.*

Finally, the desired SCS is the set of points in the $\gamma_1\gamma_2$-plane for which $\gamma_1 \leq \gamma_1^*(\theta_2)$ and $\gamma_2 \leq \gamma_2^*(\theta_1)$, where $\gamma_1^*(\theta_2)$ and $\gamma_2^*(\theta_1)$ are the points of convergences in Cases 1 and 2, respectively. In Figure 5.1, these points are the same as the intersection points of the vertical and the horizontal lines. An alternative way of obtaining the SCS is described in the next section. As to the rate of convergence, few iterations (like 2 and 3) are required when $\epsilon = 0.001$ for the examples considered this report.

5.3.2 Point-wise Testing

Suppose that the interest is to construct a lower $(1-\alpha)100\%$ confidence set for $\boldsymbol{\gamma}$ as in (5.8). From (5.8) it follows that the point $\boldsymbol{\theta}$ belongs to this confidence set if it fulfils the condition $T_\ell(\theta_\ell) \leq c_{1-\alpha}(\mathbf{R}(\boldsymbol{\theta})), \quad \ell = 1, \ldots, r$. Therefore, for every point $\boldsymbol{\theta}$ we can directly decide whether it belongs to the confidence set or not. In Figure 5.1, the accepted points are shown by dots. Therefore, both approaches lead to the same region. Using (5.9), the two-sided SCS can be constructed analogously.

The difference between the two approaches is that the iterative approach initializes the unknown parameters and the iteration is repeated until convergence to boundary points, whereas the second approach consists of deciding for every point whether it belongs to the confidence set or not. The iterative approach works best for $r = 2$. The approach based on pointwise testing works for any r, although it takes more time since it has to go over all the grid points.

5.4 Simultaneous Confidence Intervals

In this section, we discuss useful approximations to the exact SCS described in the previous section. In effect, we replace the critical point $c_{1-\alpha}(\boldsymbol{\gamma})$ which depends on the vector of the unknown ratios by some constant(s) which is

free of $\boldsymbol{\gamma}$. By doing so we get conservative or approximate SCI. As discussed in Section 5.3, SCS can be determined precisely; but often, they have strange and irregular shapes which make the interpretations harder (See, e.g., Chikuse (1981), where confidence sets are constructed for ratios of the discriminant coefficients.) On the other hand, SCI are rectangular in shape and easily interpretable. However, it is not possible to construct SCI which satisfy the pre-specified familywise confidence level exactly. Thus, there is a trade-off between exact SCS and SCI for ratios.

Let $Q_\ell(\gamma_\ell, c_{1-\alpha}) = A_\ell\gamma_\ell^2 + B_\ell\gamma_\ell + C_\ell$ denote a quadratic function in the ratio parameter γ_ℓ derived on the basis of the critical point $c_{1-\alpha}$, and by solving inequality of the type in (5.3). The SCI are determined by solving inequalities of the type in (5.4) for each ratio separately.

5.4.1 Probability Inequalities

Here we present three basic probability inequalities used in multiple comparison procedures.

a) Bonferroni method. The Bonferroni simultaneous confidence limits can be obtained by using the usual Bonferroni adjusted critical point in Fieller intervals (see, e.g., Malley (1982)). For r ratios, the critical point for a two-sided SCI is $t_{1-\frac{\alpha}{2r}}(\nu)$. Therefore, the Bonferroni SCI for $\boldsymbol{\gamma}$ are the solutions

of the inequalities

$$Q_\ell(\gamma_\ell, t_{1-\frac{\alpha}{2r}}(\nu)) \leq 0, \quad \ell = 1, \ldots, r.$$

Note that from the general Bonferroni inequality it follows that,

$$P\{|T_\ell| < c'_{1-\alpha}(\mathbf{R}(\boldsymbol{\gamma})), \quad \ell = 1, \ldots, r\} \geq 1 - r\left[1 - P\{|T_1| < c'_{1-\alpha}(\mathbf{R}(\boldsymbol{\gamma}))\}\right].$$

And since $1 - r\left[1 - P\{|T_1| < t_{1-\frac{\alpha}{2r}}(\nu)\}\right] = 1 - \alpha$, we have that

$$c'_{1-\alpha}(\mathbf{R}(\boldsymbol{\gamma})) \leq t_{1-\frac{\alpha}{2r}}(\nu), \tag{5.10}$$

for every correlation matrix $\mathbf{R}(\boldsymbol{\gamma})$. Therefore, the SCS are always bounded by the Bonferroni SCI. Similarly, the inequality in (5.10) can be written for the one-sided case, $c_{1-\alpha}(\mathbf{R}(\boldsymbol{\gamma})) \leq t_{1-\frac{\alpha}{r}}(\nu)$.

b) $Mt_r(\nu, \mathbf{I}_r)$ method. For a two-sided SCI, according to Jensen (1989), we apply an inequality due to Šidák (1967) for multivariate normal distributions which can be generalized for the multivariate t-distribution. For a detailed account of this inequality, we refer to Hochberg and Tamhane (1987). The correlation matrix $\mathbf{R} = [\rho_{ij}]$ in (5.2) can be written in the form $\rho_{ij} = \delta_i \delta_j c_{ij}, \quad \delta_i, \delta_j \in [0, 1], \quad i \neq j$ and $\rho_{ii} = 1$, where $\mathbf{C} = \{c_{ij}\}$ is a positive (semi) definite correlation matrix. According to Šidák (1967), $c'_{1-\alpha}(\mathbf{R}(\boldsymbol{\theta}))$ is decreasing in all δ_i. Therefore, for all vectors $\boldsymbol{\theta}$ we get $c'_{1-\alpha}(\mathbf{R}(\boldsymbol{\theta})) \leq c'_{1-\alpha}(\mathbf{I}_r)$, where $\mathbf{I}_r$ is an identity matrix of rank r. This means that SCI based on the critical point of $Mt_r(\nu, \mathbf{I}_r)$ (or *MtI-SCI* for short), completely covers the exact

SCS and hence it is conservative. But from (5.10), MtI-SCI is less conservative than the Bonferroni-SCI. The two-sided MtI-SCI are the solutions of

$$Q_\ell(\gamma_\ell, c'_{1-\alpha}(\mathbf{I}_r)) \leq 0, \quad \ell = 1, \ldots, r.$$

For the one-sided case, Slepian inequality (see Hochberg and Tamhane, 1987) can be applied instead of Šidák's inequality. According to Slepian, $c_{1-\alpha}(\mathbf{R}(\boldsymbol{\theta}))$ is decreasing in all elements ρ_{ij} of $\mathbf{R}(\boldsymbol{\theta})$. This means that $c_{1-\alpha}(\mathbf{R}(\boldsymbol{\theta})) \leq c_{1-\alpha}(\mathbf{I}_r)$ for all $\boldsymbol{\theta}$ having only non-negative elements ρ_{ij} of $\mathbf{R}(\boldsymbol{\theta})$. Therefore, if such $\boldsymbol{\theta}$ belongs to the exact SCS, then it also belongs to the MtI-SCI. But, if for a point $\boldsymbol{\theta}$ some ρ_{ij}'s are negative, then the last statement does not hold true. In Section 5.5, the geometrical forms of the exact SCS and MtI-SCI will be demonstrated for the special case of many-to-one comparisons.

5.4.2 Projection Method

A method due to Scheffé (1970) is called the projection method and it uses quantiles of the F-distribution. Basically, it consists of projecting the simultaneous confidence sets onto the coordinate axes of the γ_ls. The two-sided SCI are the solutions of

$$Q_\ell(\gamma_\ell, [mF_{1-\alpha}(m, \nu)]^{\frac{1}{2}}) \leq 0, \quad \ell = 1, \ldots, r,$$

where m is the dimension of the space spanned by the vectors $\mathbf{c}_\ell - \gamma_\ell \mathbf{d}_\ell$, $\ell = 1, \ldots, r$. More details on this can be found in Scheffé (1970) and Zerbe *et al.* (1982). Note that the Scheffé-SCI is a simultaneous confidence interval not only for the r contrasts under study, but for all possible combinations of the m basic contrasts. Therefore, when $m = r$ (e.g., when all ratios have the same denominator), the conservativeness of the Scheffé method rapidly increases with r.

5.4.3 The Plug-in Approach

The idea of this competing approach is to derive approximate SCI by estimating $\mathbf{R}(\boldsymbol{\gamma})$ which depends on the unknown $\boldsymbol{\gamma}$. Under the ANOVA model in Section 5.2, the maximum likelihood estimator of $\boldsymbol{\mu}$ is $\overline{\mathbf{Y}}$. Hence, by the invariance property, the maximum likelihood estimators of the ratio parameters in (5.1) are $\widehat{\gamma}_\ell = \mathbf{c}_\ell' \overline{\mathbf{Y}} \,/ \mathbf{d}_\ell' \overline{\mathbf{Y}}$, $\ell = 1, \ldots, r$. Plugging these in $\mathbf{R}(\boldsymbol{\gamma})$, we obtain $\mathbf{R}(\widehat{\boldsymbol{\gamma}}) = [\widehat{\rho}_{ij}]$, where $\widehat{\rho}_{ij} = h(\widehat{\gamma}_i, \widehat{\gamma}_j), \quad 1 \leq i \neq j \leq r$. We call this method the 'plug-in' approach. Let $c'_{1-\alpha}(\mathbf{R}(\widehat{\boldsymbol{\gamma}}))$ denote the two-sided equicoordinate critical point associated with $\mathbf{R}(\widehat{\boldsymbol{\gamma}})$. The approximate SCI are obtained by solving

$$Q_\ell(\gamma_\ell, c'_{1-\alpha}(\mathbf{R}(\widehat{\boldsymbol{\gamma}})) \leq 0, \quad \ell = 1, \ldots, r.$$

The one-side SCI can be obtained similarly.

5.4.4 Resampling Techniques

In this section we describe two resampling methods. Let $y_{ij}^{(b)}$ denote the j^{th} observation under the i^{th} treatment for the b^{th} bootstrap sample, $i = 1, \ldots, k$; $j = 1, \ldots, n_i$; $b = 1, \ldots, N$. The bootstrap version of the test statistic $T_\ell = L_\ell \,/ S_{L_\ell}$ is given by

$$T_\ell^{*(b)} = \frac{(\widehat{\gamma}_\ell \mathbf{d}_\ell - \mathbf{c}_\ell)' \overline{\mathbf{y}}^{*(b)}}{s^{*(b)} [\widehat{\gamma}_\ell{}^2 \mathbf{d}_\ell' \mathbf{M} \mathbf{d}_\ell - 2\widehat{\gamma}_\ell \mathbf{c}_\ell' \mathbf{M} \mathbf{d}_\ell + \mathbf{c}_\ell' \mathbf{M} \mathbf{c}_\ell]^{\frac{1}{2}}},$$

where $\overline{\mathbf{y}}^{*(b)} = \left(\overline{y}_1^{*(b)}, \ldots, \overline{y}_k^{*(b)}\right)'$ and $s^{*(b)}$, respectively denote the vector of sample means and an estimate of the common standard deviation computed from the b^{th} bootstrap sample. The observations are mean-centred and pooled into one single dataset prior to resampling (Westfall and Young, 1993). The following procedures are considered for the estimation of the critical point(s).

a) T_{Max}. This is an adaptation of the method discussed by Westfall and Young (1993, p82) for ratios. For a two-sided SCI, an estimate of the critical point of interest, say $\widehat{c}_{1-\alpha}^*$, is obtained by computing the $(1-\alpha)^{\text{th}}$ quantile of the values

$$T_{\text{Max}}^{*(b)} = \text{Max}\left\{\left|T_1^{*(b)}\right|, \ldots, \left|T_r^{*(b)}\right|\right\}, \quad b = 1, \ldots, N.$$

Therefore, the two-sided SCI are the solutions of

$$Q_\ell(\gamma_\ell, \widehat{c}_{1-\alpha}^*) \leq 0, \quad \ell = 1, \ldots, r.$$

To construct one-sided SCI, we follow similar steps except that we drop the sign for the absolute value in the above expression for $T^{*(b)}_{\text{Max}}$.

We remark that unlike in the case of inference for the difference of means, the means do not vanish under the null hypothesis in the test statistics. The scheme in Westfall and Young (1993) works best for location shift problems.

b) Balanced SCI. This is a technique based on the idea of pre-pivoting and balance as described by Beran (1987, 1988). The purpose of balancing is to correct for uneven coverage probabilities for the individual parameters. The steps are as follows. Let H_ℓ be the left continuous cdf of the statistic T_ℓ, and let H be the cdf of $\text{Sup}\{H_\ell(T_\ell), \quad \ell = 1, \ldots, r\}$. Thus, H is a mapping from [0,1] to [0,1]. The bootstrap estimates of the critical points for a lower one-sided SCI are obtained as

$$\widehat{c}^{*}_{\ell,1-\alpha} = \widehat{H}^{-1}_{\ell}\left[\widehat{H}^{-1}(1-\alpha)\right],$$

where $\widehat{H}_\ell$ and $\widehat{H}$ are empirical estimates of H_ℓ and H, respectively. Note that in this case the critical points are estimated separately for each ratio parameter.

In summary, a two-sided approximate $(1-\alpha)100\%$ SCI for $(\gamma_1, \ldots, \gamma_r)'$ is given by

$$\left(\frac{-B_\ell - \sqrt{B_\ell^2 - 4A_\ell C_\ell}}{2A_\ell}, \frac{-B_\ell + \sqrt{B_\ell^2 - 4A_\ell C_\ell}}{2A_\ell} \right), \quad \ell = 1, 2, \ldots r$$

where

$$\begin{aligned} A_\ell &= (\mathbf{d}_\ell' \overline{\mathbf{Y}})^2 - q^2 S^2 \mathbf{d}_\ell' \mathbf{M} \mathbf{d}_\ell, \\ B_\ell &= -2 \left[(\mathbf{c}_\ell' \overline{\mathbf{Y}})(\mathbf{d}_\ell' \overline{\mathbf{Y}}) - q^2 S^2 \mathbf{c}_\ell' \mathbf{M} \mathbf{d}_\ell) \right], \\ C_\ell &= (\mathbf{c}_\ell' \overline{\mathbf{Y}})^2 - q^2 S^2 \mathbf{c}_\ell' \mathbf{M} \mathbf{c}_\ell, \end{aligned}$$

and

$$q = \begin{cases} t_{1-\frac{\alpha}{2r}}(\nu) & \text{if Bonferroni adjustment,} \\ c'_{1-\alpha}(\mathbf{I}_r) & \text{if Šidák,} \\ c'_{1-\alpha}(\mathbf{R}(\widehat{\boldsymbol{\gamma}})) & \text{if plug-in,} \\ \widehat{c}^*_{1-\alpha} & \text{if resampling } (T_{\text{Max}}). \end{cases}$$

Before providing some examples on the methods discussed in Sections 5.2 to 5.4, we give some general remarks.

(i) Applications to the general linear model

The methods discussed in the previous sections are equally applicable to the problems of constructing SCS and SCI for the ratios of regression coefficients in the general linear model. Let $\mathbf{X}$ be an $n \times p$ design matrix and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$ be a vector of unknown regression coefficients. The model is

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}_n(\mathbf{0}, \sigma^2 \mathbf{I}).$$

Interest lies in simultaneous confidence sets (intervals) for the ratios

$$\gamma_\ell = \frac{\sum_{i=1}^{p} c_{\ell i}\beta_i}{\sum_{i=1}^{p} d_{\ell i}\beta_i} = \frac{\mathbf{c}_\ell'\boldsymbol{\beta}}{\mathbf{d}_\ell'\boldsymbol{\beta}},$$

where $\mathbf{c}_\ell$ and $\mathbf{d}_\ell$ are known vectors of real constants of length p associated with the ℓ^{th} ratio, $\ell = 1, \ldots, r$. In order to get estimable linear combinations, the vectors $\mathbf{c}_\ell$ and $\mathbf{d}_\ell$ must be linear combinations of the rows of $\mathbf{X}$. Let S^2 be the unbiased estimator of σ^2 based on $\nu = n - \text{Rank}(\mathbf{X})$ degrees of freedom. Following the discussions in Section 5.2, the test statistic

$$T_\ell = \frac{(\gamma_\ell \mathbf{d}_\ell - \mathbf{c}_\ell)'\widehat{\boldsymbol{\beta}}}{S[\gamma_\ell^2 \mathbf{d}_\ell'\mathbf{M}\mathbf{d}_\ell - 2\gamma_\ell \mathbf{c}_\ell'\mathbf{M}\mathbf{d}_\ell + \mathbf{c}_\ell'\mathbf{M}\mathbf{c}_\ell]^{\frac{1}{2}}}$$

is distributed as $t(\nu)$, $\ell = 1, \ldots, r$, where $\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{Y}$, $\mathbf{M} = (\mathbf{X}'\mathbf{X})^{-}$ and $\mathbf{A}^{-}$ denotes a generalized inverse of $\mathbf{A}$. Jointly, $\mathbf{T} = (T_1, \ldots, T_r)' \sim Mt_r(\nu, \mathbf{R})$, where $\mathbf{R} = [\rho_{ij}]$ is as defined in (5.2). Therefore, SCS and SCI can be developed for $\boldsymbol{\gamma}$ as discussed in Sections 5.3 and 5.4.

(ii) Multiple assays

The multiple assays (parallel-line and slope-ratio) discussed by Jensen (1989) can be written in the form of the general linear model, and hence all of the methods discussed above are equally applicable to these problems.

(iii) Limit of the correlation matrix $\mathbf{R}(\boldsymbol{\gamma})$

The limit of the correlation matrix in (5.2) is

$$\begin{aligned}\lim_{\gamma_i,\gamma_j\to+\infty}\rho_{ij} &= \lim_{\gamma_i,\gamma_j\to-\infty}\rho_{ij} = \frac{\mathbf{d}_i'\mathbf{M}\mathbf{d}_j}{\sqrt{\mathbf{d}_i'\mathbf{M}\mathbf{d}_i}\sqrt{\mathbf{d}_j'\mathbf{M}\mathbf{d}_j}}\\ \lim_{\gamma_i\to-\infty,\gamma_j\to+\infty}\rho_{ij} &= \lim_{\gamma_i\to+\infty,\gamma_j\to-\infty}\rho_{ij} = -\frac{\mathbf{d}_i'\mathbf{M}\mathbf{d}_j}{\sqrt{\mathbf{d}_i'\mathbf{M}\mathbf{d}_i}\sqrt{\mathbf{d}_j'\mathbf{M}\mathbf{d}_j}}.\end{aligned}$$

A remarkable result is that $\mathbf{R}$ is free of $\boldsymbol{\gamma}$ asymptotically. Thus, it sounds that in some multiple ratio problems, the limit of the correlation matrix can be used to derive SCI. It is also interesting to note that the limit matrix depends only on the $\mathbf{d}_i$'s, the vectors which appear in the denominator of the ratios. Furthermore, if $\mathbf{d}_i = \mathbf{d}_j, 1 \le i < j \le r$; in other words, if all ratios have the same denominator, then the limit of the correlation in (5.2) will be 1 or -1 depending on whether γ_i and γ_j have the same or different signs. In all cases considered, the correlations do not depend on $\mathbf{M}$ if the design is balanced.

5.5 Many-to-One Comparisons

Many-to-one comparisons are often of interest in many areas of applications. As the name implies, the design consists of comparing many treatments to one treatment, often a control treatment. This comparison appears to be the simplest design on which to apply simultaneous inference for several ratios since all ratios have the same denominator.

Suppose that we have $k+1$ treatments (including the control). We assume that the responses of interest Y_{ij} are independent observations from

$$\mathcal{N}(\mu_i, \sigma^2), \quad i = 0, 1, \ldots, k;\ j = 1, \ldots, n_i,$$

where $i = 0$ refers to the control treatment. The aim is to construct confidence sets and SCI for the ratios of means $\gamma_\ell = \mu_\ell / \mu_0, \quad \ell = 1, \ldots, k$. As described in Chapter 4, this formulation can be utilized for ratio-based inference for non-inferiority (or superiority) trials.

Now the pivotal quantity

$$T_\ell = \frac{\gamma_\ell \overline{Y_0} - \overline{Y_\ell}}{S\sqrt{\frac{1}{n_\ell} + \frac{\gamma_\ell^2}{n_0}}}$$

has a t-distribution with $\nu = \sum_{i=0}^{k}(n_i - 1)$ degrees of freedom, $\ell = 1, \ldots, k$. Jointly, $\mathbf{T} = (T_1, \ldots, T_k)' \sim Mt_k(\nu, \mathbf{R})$. In many-to-one comparisons, the elements of the correlation matrix $\mathbf{R}$ simplify to

$$\rho_{ij} = \frac{\gamma_i \gamma_j}{\sqrt{\gamma_i^2 + \frac{n_0}{n_i}}\sqrt{\gamma_j^2 + \frac{n_0}{n_j}}}, \quad 1 \leq i \neq j \leq k.$$

Equivalently, the correlations can be written as $\rho_{ij} = \lambda_i \lambda_j$, where

$$\lambda_i = \frac{\text{sign}(\gamma_i)}{\sqrt{1 + \tau_i^{-2}}}, \tag{5.11}$$

$\tau_i = \gamma_i \sqrt{n_i / n_0}$ and $\text{sign}(\gamma_i)$ refers to the the sign of γ_i, with $\text{sign}(\gamma_i) = -1$ if $\gamma_i < 0$ and $\text{sign}(\gamma_i) = 1$ if $\gamma_i > 0$. The expression in (5.11) shows how the correlations are related to the ratios and the sample sizes. The correlations

tend to zero as τ_i's approach zero (i.e., as we increase the sample size for the control group or as the ratios tend to zero). More importantly, the index τ_i enables us to better compare the various methods of constructing SCI with respect to their coverage probabilities. T_ℓ is t-distributed with ν degrees of freedom and $c_{1-\alpha}(\mathbf{R})$ depends only on ν and the τ_i's as can be seen from (5.11). Therefore, for many-to-one comparisons, the coverage probability of the one-sided SCS (5.8) and analogously that of the two-sided SCS (5.9) depends only on ν and the τ_i's. Moreover, note that after little algebra T_ℓ can also be expressed as

$$T_\ell = \frac{\tau_\ell - \widehat{\tau}_\ell}{\widehat{CV}_{\overline{Y}_0}\sqrt{1+\tau_\ell^2}},$$

where $\widehat{\tau}_\ell = \overline{Y}_\ell\sqrt{n_\ell}\,/(\overline{Y}_0\sqrt{n_0}) = \widehat{CV}_{\overline{Y}_0}\Big/\widehat{CV}_{\overline{Y}_\ell}$ and $\widehat{CV}_{\overline{Y}_\ell}$ is an estimator of the coefficient of variation of $\overline{Y}_\ell$. Since the SCS for $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_r)'$ is defined analogously to (5.8) or (5.9), its shape depends only on ν, $\widehat{CV}_{\overline{Y}_0}$ and the $\widehat{\tau}_\ell$'s. From the SCS of $\boldsymbol{\tau}$, we get the SCS of $\boldsymbol{\gamma}$ by the linear transformations $\gamma_\ell = \tau_\ell\sqrt{n_0\,/n_\ell}$.

In the following sections, a wide variety of examples in many-to-one comparisons are used to illustrate the methodologies described in Sections 5.2 to 5.5. Since all ratios have the same denominator, the corresponding g values defined in (5.5) are identical.

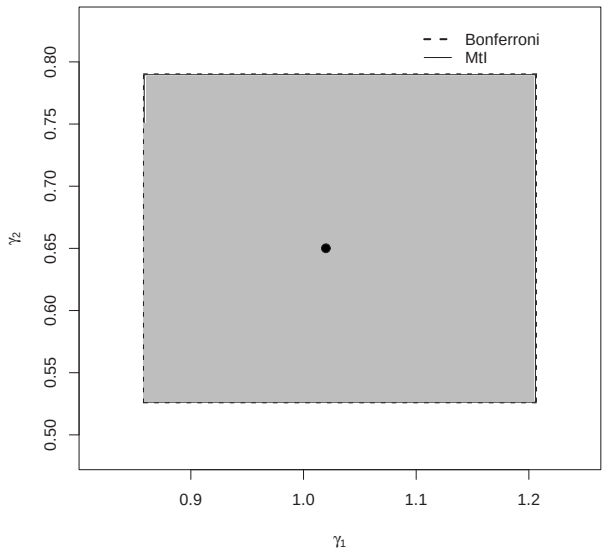

Figure 5.2: *Two-sided* 95% *SCS and Bonferroni SCI for* $\boldsymbol{\gamma} = (\gamma_1, \gamma_2)'$ *with* $\widehat{\tau}_1 = 0.85$, $\widehat{\tau}_2 = 0.65$ *and* $\widehat{CV}_{\overline{Y}_0} = 0.05$ *(Body weight data).*

Example 1 *(Two-sided SCI and small g value,* $k = 2$*).* Recall the body weight data described in Section 2.2. Let μ_0, μ_1 and μ_2 denote the true average weight gain for the control, Thyroxin, and Thiouracil treatment, respectively. The interest is to simultaneously compare Thyroxin and Thiouracil with the control based on the ratios of means $\gamma_1 = \mu_1/\mu_0$ and $\gamma_2 = \mu_2/\mu_0$. The summary statistics are $\overline{y}_0 = 106.6$, $\overline{y}_1 = 108.7$, $\overline{y}_2 = 69.3$, $s^2 = 240.66$ (an estimate of σ^2 under the assumption of variance homogeneity) and $\widehat{CV}_{\overline{Y}_0} = 0.05$ (an estimate of the coefficient of variation for the mean of the control group). The two-sided 95% SCI for γ_1 and γ_2 are constructed

Table 5.1: Two-sided 95% SCI (Body weight gain data)

Ratio	Bonferroni	*MtI*	Plug-in	T_{Max}
γ_1	(0.858, 1.207)	(0.859, 1.206)	(0.860, 1.205)	(0.860, 1.205)
γ_2	(0.526, 0.790)	(0.526, 0.790)	(0.527, 0.789)	(0.528, 0.789)

by using the various methods described in Section 5.4. The results are summarized in Table 5.1. In this example, the confidence limits for all methods are very close to each other. Moreover, the exact confidence set (the shaded region) and the Bonferroni SCI are almost identical, as seen from Figure 5.2. The value of g associated with the Bonferroni SCI is 0.009, which is quite small. This is computed by using $t_{1-\frac{\alpha}{2r}}(\nu)$ in (5.5). In the figure, the point estimate of $\boldsymbol{\gamma} = (\gamma_1, \gamma_2)'$ is shown by a heavy dot which corresponds to $\widehat{\gamma}_1 = 1.02$ and $\widehat{\gamma}_2 = 0.65$, and the estimates of τ_1 and τ_2 are 0.85 and 0.65, respectively. From the upper confidence limits for γ_2, we infer that the weight gain induced by Thiouracil is by at least 20 per cent less than that of the control. The plug-in confidence interval for γ_1 varies from 0.860 to 1.205, and therefore we cannot exclude that Thyroxin and the control have similar effects (Table 5.1).

Example 2 *(Two-sided SCI and large g value, $k = 2$).* In order to illustrate the influence of g, we show in Figure 5.3 the two-sided SCS and SCI based on

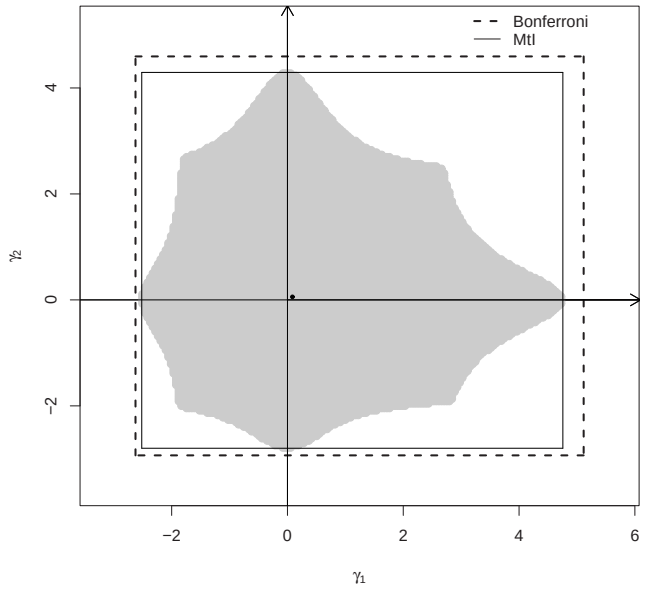

Figure 5.3: *Two-sided* 95% *SCS, Bonferroni and MtI-SCI for* $\boldsymbol{\gamma} = (\gamma_1, \gamma_2)'$ *with* $n_0 = n_1 = n_2 = 10$, $\widehat{\tau}_1 = 0.086$, $\widehat{\tau}_2 = 0.0571$, $\widehat{CV}_{\overline{Y}_0} = 0.41$ *(Example 2).*

a balanced hypothetical dataset with relatively large coefficient of variation for the mean of the control group ($\widehat{CV}_{\overline{Y}_0} = 0.41$). In this example, the differences between the exact simultaneous confidence sets, the Bonferroni SCI and the MtI-SCI are visible. The g values associated with the Bonferroni and MtI-SCI are computed to be 0.931 and 0.923, respectively. The exact SCS is circumscribed by the MtI-SCI rectangle as is expected from Šidák inequality. Note also the irregularity in the shape of the SCS and the absence of any symmetry.

Table 5.2: Two-sided 95% SCI for ratios to the control (leaf chroma value)

Ratio	Bonferroni	MtI	Plug-in	T_{Max}
γ_1	(0.895, 1.216)	(0.896, 1.216)	(0.899, 1.211)	(0.917, 1.187)
γ_2	(0.948, 1.278)	(0.949, 1.278)	(0.952, 1.273)	(0.970, 1.248)
γ_3	(1.296, 1.692)	(1.296, 1.691)	(1.300, 1.685)	(1.322, 1.655)

Table 5.3: Two-sided 95% SCI for ratios to the control (chlorophyll content)

Ratio	Bonferroni	MtI	Plug-in	T_{Max}
γ_1	(0.805, 0.939)	(0.805, 0.939)	(0.806, 0.939)	(0.854, 0.886)
γ_2	(0.649, 0.773)	(0.649, 0.773)	(0.650, 0.773)	(0.614, 0.725)
γ_3	(0.405, 0.517)	(0.406, 0.517)	(0.406, 0.516)	(0.447, 0.475)

Example 3 *(Two-sided SCI and small g values, $k = 3$).* Recall the data on leaf chroma and leaf chlorophyll content described in Section 2.1. The two-sided 95% SCI for the ratios to control are shown in Tables 5.2 and 5.3. The details of the raw data used to obtain the results for the resampling method (under T_{Max}) can be found in Mutui (2005). From the lower confidence limits for γ_3 (Table 5.2), we see that the average chroma value at the 2 $\mu l/l$ level of Ethylene is by at least 30% greater than that of the control. Since the confidence intervals for γ_1 and γ_2 do cover 1, there is no statistically significant increase in the chroma values at Ethylene levels 0.5 and 1. On

the other hand, for the leaf chlorophyll content, the simultaneous confidence intervals fall to the left of 1. This implies that the three levels of Ethylene induced a reduction in the percentage of chlorophyll content as compared to the control. More specifically, Table 5.3 indicate a reduction of about 5%, 25% and 50% for the ethylene levels 0.5, 1 and 2, respectively.

Example 4 *(One-sided SCI and large g value, $k = 2$).* To illustrate one-sided SCS and SCI, we again shall use hypothetical data. Figure 5.4 consists of the lower 95% exact confidence set and the Bonferroni SCI for two ratios. Both the iterative and point-wise testing described in Section 5.3 produce the same confidence set. The g value for the Bonferroni SCI is 0.269. In the figure, it appears as if the exact SCS coincides with the Bonferroni SCI in the second and fourth quadrants. However, there are some discrepancies as can be seen from Figure 5.5, which is a zoom of Figure 5.4 in the neighbourhood of γ_1-intercept. Note that γ_1 and γ_2 have different signs in the second and fourth quadrants. This implies that $\rho_{12} = \lambda_1\lambda_2 < 0$, and hence $c_{1-\alpha}(\mathbf{R}(\boldsymbol{\gamma})) \geq c_{1-\alpha}(\mathbf{I}_2)$ in these quadrants. Therefore, if $\boldsymbol{\gamma}$ belongs to the MtI-SCI, then it also belongs to the exact confidence set. In other words, the exact confidence set completely covers the MtI-SCI in these regions. Another point of interest is the corner points of the confidence sets. For the one-sided confidence set in Figure 5.4, the point is denoted by C. This point

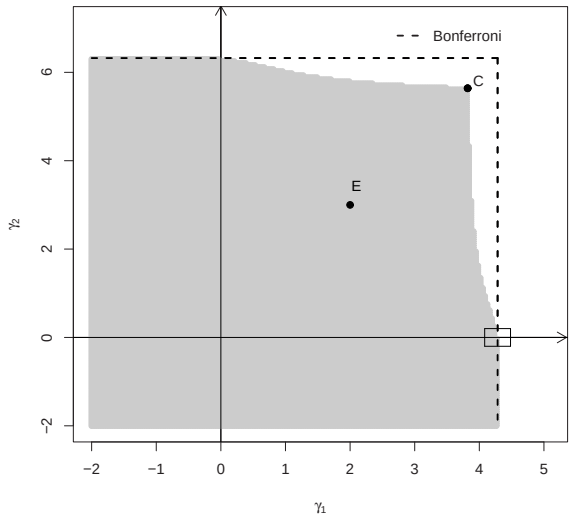

Figure 5.4: *Lower* 95% *SCS and Bonferroni SCI for* $\boldsymbol{\gamma} = (\gamma_1, \gamma_2)'$ *with* $n_0 = n_1 = n_2 = 10$, $\widehat{\tau}_1 = 2$, $\widehat{\tau}_2 = 3$, $\widehat{CV}_{\overline{Y}_0} = 0.25$ *(Example 4).*

can easily be obtained by iterating both components of $\boldsymbol{\gamma}$ simultaneously in the iterative approach described in Section 5.3. In fact, in order to effectively choose the grid points in the iterative approach, one has to locate the corner point in advance. This point can serve as a measure of deviation of the exact confidence set from the Bonferroni SCI.

Example 5 *(One-sided SCI and large g value,* $k = 3$*).* Consider another hypothetical data set with relatively large coefficient of variation for the mean of the control group. The summary statistics are $\overline{y}_0 = 0.82$, $\overline{y}_1 = 2.19$,

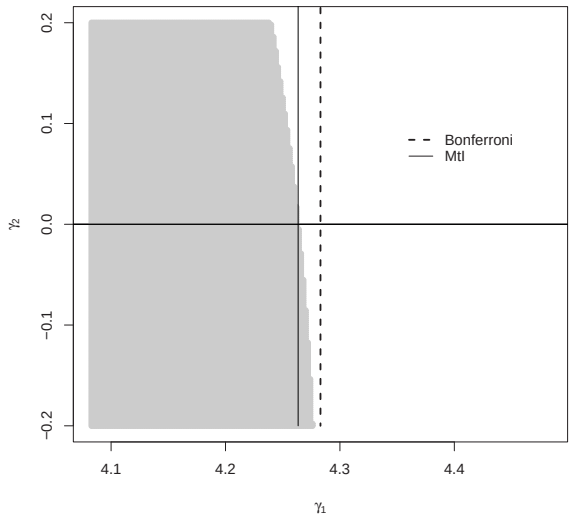

Figure 5.5: *Zoom of part of Fig. 3 (Example 4).*

$\overline{y}_2 = 2.75$, $\overline{y}_3 = 3.75$, $s = 0.80$. The sample sizes are taken to be $n_0 = 10$, $n_1 = 11$, $n_2 = 13$ and $n_3 = 13$. For this data, we construct a one-sided 95% SCS and SCI. The 95% exact one-sided SCS along with the Bonferroni SCI is shown in Figure 5.6. By definition, the lower confidence limits extend to $-\infty$. In Figure 5.6, only part of the SCS in the first octant (where γ_1, γ_2 and γ_3 are all positive) is shown for illustration purposes. In this example, $\widehat{\tau}_1 = 2.80$, $\widehat{\tau}_2 = 3.83$, $\widehat{\tau}_3 = 5.22$, and they are all greater than one. Consequently, we see a marked discrepancy between the exact SCS and the Bonferroni SCI. The estimate of $CV_{\overline{Y}_0}$ is 0.307. For this dataset, the Bonferroni SCI is clearly too conservative. Suppose that the ratio parameters are a

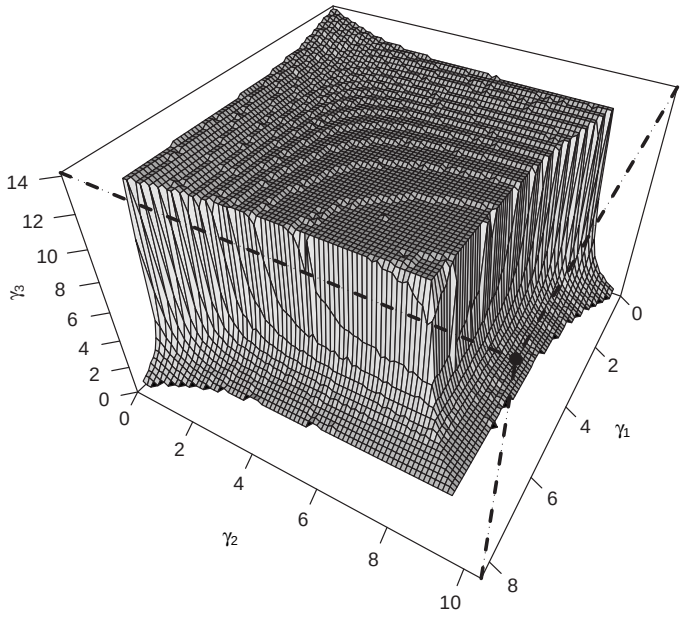

Figure 5.6: *Lower 95% SCS and Bonferroni SCI (Example 5).*

priori known to be positive, thus the elements of the correlation matrix are positive. Therefore, as mentioned in Section 5.4.1, the MtI method yields SCI which is slightly less conservative than Bonferroni. In this example, the exact SCS lie well inside the SCI based on MtI method. Hence, it should be preferred to Bonferroni SCI. On the other hand, the upper simultaneous confidence limits associated with the plug-in and the resampling are smaller than the MtI limits, hence the plug-in and the resampling SCI cut out part of

Table 5.4: Upper bounds of the one-sided 95% simultaneous confidence intervals for γ_1, γ_2 and γ_3

Ratio	Bonferroni	Slepian	Plug-in	T_{Max}
γ_1	8.348	8.275	6.471	6.337
γ_2	10.428	10.337	8.077	7.910
γ_3	14.173	14.049	10.976	10.747

the exact SCS. Table 5.4 summarizes the estimates of the upper confidence limits by the various methods described earlier. In this case, the estimates of the limits are quite different. As one would expect, the Bonferroni SCI are the most conservative.

In summary, for the many-to-one comparisons considered in this section, it is observed that the shape of the SCS mainly depends on the coefficient of variation of the mean of the control group. If $\widehat{CV}_{\overline{Y}_0}$ is small and hence if g is very close to zero, the shape of the exact SCS is very close to that of Bonferroni and MtI-SCI. On the other hand, if $\widehat{CV}_{\overline{Y}_0}$ is large (or if g is close to 1), then there is a substantial difference between the SCS and the SCI. However, note that there is no connection between shapes of SCS and SCI and the coverage probabilities, as will be discussed after the following simulation study.

5.6 Simulation Study

In this section, a simulation study is carried out to investigate the behavior of the different methods with respect to coverage probabilities and estimates of the critical points. To simplify the matter, we present the simulation results for the case of two ratio parameters for the many-to-one comparisons described in the previous section. Three values for the coefficient of variation (10%, 20% and 50%, regarded as small, intermediate and large) for the comparison group and two sample sizes (10 and 20) in a balanced design are selected. The number of simulation runs and the number of bootstrap samples (N) are each set to 10^4. The observations are generated from a normal distribution and lower 90% SCI are constructed in each case.

In Tables 5.5 and 5.6, estimates of the coverage probabilities for the individual ratios and for all ratios simultaneously, are displayed for the various scenarios. For all random samples, the g values are observed to be less than 1 as desired. The tabulated g^* values are the population quantities $g^* = t^2 CV^2_{\overline{Y}_0}$. In Table 5.5 and 5.6, they are provided for the Bonferroni SCI. As can be seen from Table 5.5, estimates of the coverage probabilities are the largest for Bonferroni SCI (both for the individual and all ratios simultaneously) followed by MtI when the τ_i's are greater than one. The results under MtI are based on the critical points of $Mt_2(\nu, I_2)$. For the plug-in and

Table 5.5: Estimates of the coverage probabilities ($r = 2,\ \tau_1 = 2, \tau_2 = 3,\ 1 - \alpha = 0.90$)

$CV_0(\%)$	n	g^*	Ratio	Bonferroni	MtI	Plug-in	T_{Max}	Balanced
10	10	0.003	γ_1	0.947	0.945	0.927	0.927	0.927
			γ_2	0.947	0.945	0.927	0.927	0.926
			$\boldsymbol{\gamma}$	0.924	0.921	0.897	0.897	0.897
10	20	0.001	γ_1	0.947	0.946	0.928	0.928	0.928
			γ_2	0.950	0.949	0.932	0.932	0.932
			$\boldsymbol{\gamma}$	0.926	0.924	0.903	0.902	0.902
20	10	0.012	γ_1	0.951	0.950	0.932	0.932	0.932
			γ_2	0.953	0.951	0.933	0.933	0.933
			$\boldsymbol{\gamma}$	0.932	0.930	0.904	0.905	0.905
20	20	0.006	γ_1	0.948	0.946	0.928	0.928	0.928
			γ_2	0.948	0.947	0.929	0.929	0.929
			$\boldsymbol{\gamma}$	0.927	0.925	0.899	0.899	0.899
50	10	0.073	γ_1	0.947	0.945	0.928	0.928	0.928
			γ_2	0.949	0.948	0.931	0.931	0.931
			$\boldsymbol{\gamma}$	0.926	0.924	0.901	0.901	0.901
50	20	0.035	γ_1	0.949	0.947	0.930	0.930	0.930
			γ_2	0.950	0.949	0.931	0.930	0.930
			$\boldsymbol{\gamma}$	0.929	0.926	0.902	0.902	0.902

Table 5.6: Estimates of the coverage probabilities ($r = 2,\ \tau_1 = 0.8, \tau_2 = 0.4, 1 - \alpha = 0.90$)

$CV_0(\%)$	n	g^*	Ratio	Bonferroni	MtI	Plug-in	T_{Max}	Balanced
10	10	0.003	γ_1	0.947	0.946	0.944	0.944	0.943
			γ_2	0.948	0.947	0.945	0.945	0.945
			$\boldsymbol{\gamma}$	0.901	0.899	0.896	0.895	0.894
10	20	0.001	γ_1	0.947	0.946	0.944	0.943	0.944
			γ_2	0.951	0.949	0.947	0.948	0.947
			$\boldsymbol{\gamma}$	0.904	0.900	0.897	0.898	0.898
20	10	0.012	γ_1	0.948	0.947	0.946	0.946	0.946
			γ_2	0.952	0.952	0.950	0.950	0.950
			$\boldsymbol{\gamma}$	0.907	0.905	0.903	0.903	0.903
20	20	0.006	γ_1	0.948	0.947	0.946	0.945	0.946
			γ_2	0.948	0.946	0.944	0.944	0.944
			$\boldsymbol{\gamma}$	0.903	0.900	0.897	0.897	0.897
50	10	0.073	γ_1	0.951	0.950	0.949	0.949	0.948
			γ_2	0.949	0.948	0.948	0.948	0.947
			$\boldsymbol{\gamma}$	0.907	0.905	0.904	0.904	0.903
50	20	0.035	γ_1	0.951	0.949	0.948	0.948	0.948
			γ_2	0.949	0.948	0.946	0.946	0.947
			$\boldsymbol{\gamma}$	0.906	0.903	0.901	0.901	0.901

the two resampling approaches, the estimates of the simultaneous coverage probabilities are pretty close to the nominal level of 0.90. The estimates of the coverage probabilities for the individual ratios for the balanced resampling approach and for the T_{Max} are very similar. Thus, there is no important difference between the two resampling methods for the particular situation considered in the simulation. The simulation results for the two-sided SCI are very similar to the one-sided case (results not shown here). For three ratio parameters, the behavior of the coverage probabilities is observed to be similar to the case of two ratios (results not shown here). That is, the plug-in and the resampling methods have coverage probabilities close to the nominal level, and the Bonferroni and the MtI methods are conservative if the τ_i's are greater than one.

Graphical comparisons of the estimates of the critical points for two of the cases in Table 5.5 (with the smallest and the largest g^* value) are shown in Figures 5.7 and 5.8. From both figures, we see that the estimates of the critical points are symmetrically distributed about the actual value which is computed by substituting $\boldsymbol{\gamma} = (2, 3)'$ in $\mathbf{R}$. For relatively small $CV_{\overline{Y}_0}$ (2.24%, $CV_{Y_0} = 10\%$ and $n_0 = 20$), there are small variabilities in the estimates of the critical point based on the plug-in and the two resampling approaches as can be seen from Figure 5.7. In the figure, balanced C1 and balanced C2 refer to estimates of the critical points associated with the first and the second ra-

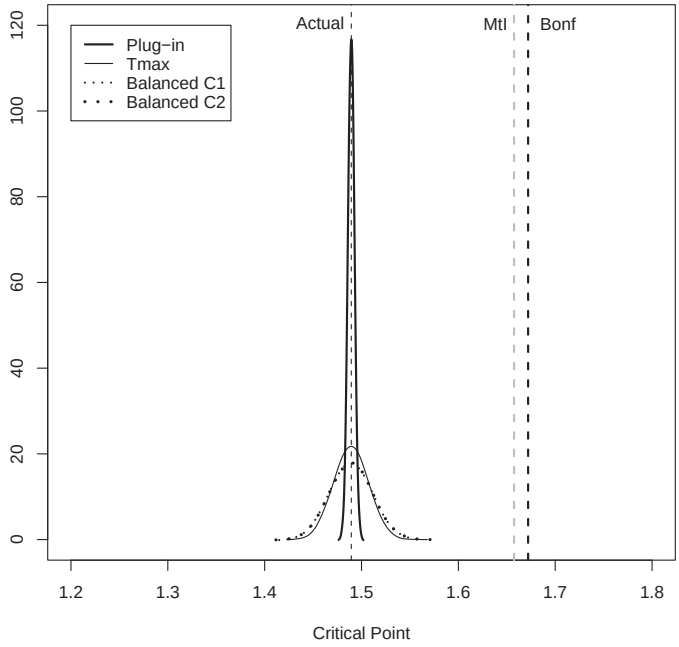

Figure 5.7: *Density estimate of the estimates of the critical point* ($\tau_1 = 2$, $\tau_2 = 3$, $CV_0 = 0.1$, $n_0 = n_1 = n_2 = 20$, $1 - \alpha = 0.90$*).*

tio based on the balanced resampling approach. The conservativeness of the Bonferroni and MtI methods relative to the actual and the estimated critical points is also clear from the figure. In contrast, when $CV_{\overline{Y}_0}$ is relatively large (15.8%), the variability in the estimated critical points increases as can be seen from Figure 5.8. However, estimates of the critical points based on the plug-in appear to be more stable than that of the resampling approaches.

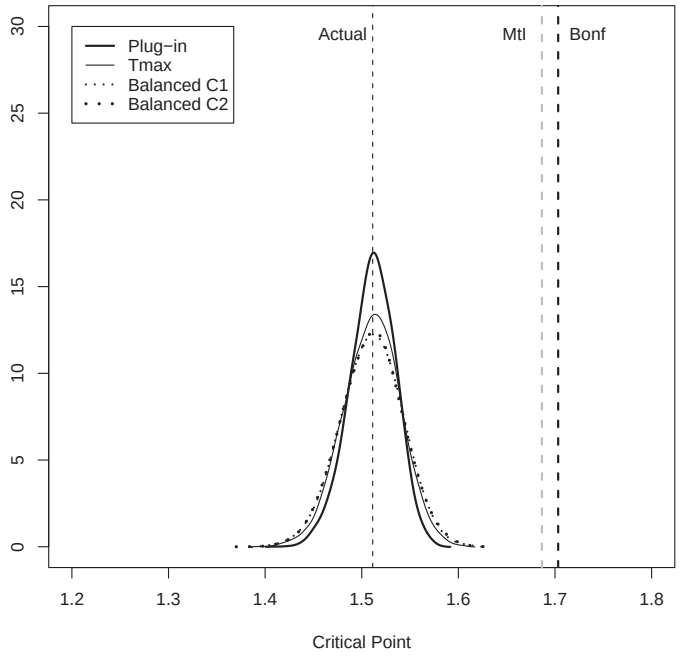

Figure 5.8: *Density estimate of the estimates of the critical point ($\tau_1 = 2$, $\tau_2 = 3$, $CV_0 = 0.5$, $n_0 = n_1 = n_2 = 10$, $1 - \alpha = 0.90$).*

5.7 Discussion

It is trustworthy to note the following points regrading shapes of SCS and conservativeness of SCI based on probability inequalities. The shape of exact SCS can appear to be rectangular and almost identical with the Bonferroni and MtI SCI, but yet the SCI can be conservative. This happens when the true ratio is very close to the border of the SCS. On the other hand, the shapes of SCS can be quite different from the SCI box, but the coverage probabilities

of Bonferroni and MtI methods can be as good as that of the exact SCS (or the same as the nominal confidence level). In other words, the shapes are different but the coverage probabilities of the exact SCS, Bonferroni and MtI-SCI methods are almost identical. This happens when the vector of ratio parameters is not close to the boundary of the SCS. Therefore, the discrepancy between the shapes of SCS and the SCI boxes is not a good indicator of conservatism.

Chapter 6

Conclusions and Further Research

6.1 Summary and Conclusions

In this research, simultaneous inference for ratios is considered with a special focus on sample size determinations in non-inferiority trials and simultaneous estimation of several ratios like, for example, in multiple assays. In the first part, a numerical study revealed that the sample sizes required for ratio-based inferences is smaller than that of the difference-based inferences when the relative non-inferiority margin is less than one and when large response values indicate better treatment effects. The designs of non-inferiority trials (or superiority trials) based on the ratio view are also illustrated using two data

examples. From the examples and the numerical study, we can conclude that the ratio view has an advantage in selected multiple tests for non-inferiority (or superiority). If small response values indicate better treatment effects, one may still take advantages of the ratio view by making inference for the ratio of control mean to that of the test treatments (instead of the test treatments to the control).

In Chapter 5, methods for constructing simultaneous confidence sets and confidence intervals for multiple ratios are discussed. The main difficulty with problems involving more than one ratio parameter is that the joint distribution of the relevant statistics (multivariate t-distribution) depends on the vector of unknown ratios. This means that the equi-coordinate critical points depend on the ratios. Consequently, there is no direct means of deriving the confidence sets and the SCI. Two methods of determining the exact confidence sets and several methods of constructing SCI for the ratios are discussed. Simulation studies based on the many-to-one comparisons are carried out to assess the performance of the two proposed methods of constructing SCI, namely, the plug-in and resampling methods. It is found that when the ratio of the coefficient of variation of the mean of the control group to that of the other treatments are greater than one (i.e., $\tau_i > 1$), estimates of the simultaneous coverage probabilities of the plug-in and the resampling methods are substantially closer to the nominal level than Bonferroni and

MtI methods. Therefore, we infer that for datasets with $\widehat{\tau}_i > 1$, the Bonferroni and MtI methods are too conservative. However, if all τ_i's are less than one, the coverage probabilities for all methods are very similar, and hence all methods are equally effective. It is also observed that the estimates of the critical points based on the plug-in method are more stable than estimates based on the resampling methods. In summary, the plug-in method behaved very well for any configuration of the τ_i's. The resampling methods have also good coverage probabilities, but, in practice, this approach can be computationally expensive. Both parametric and nonparametric bootstrap methods produced similar results.

6.2 Further Researches

The methods discussed in this thesis can be extended in several directions for further researches.

Linear mixed model- The first extension of this research can be simultaneous confidence intervals estimation in linear mixed models. These models are becoming increasing important in agricultural and medical researches. For example, Young *et al.* (1997) discuss simultaneous confidence estimation based on the Scheffé (1970) approach. However, this approach can be too conservative. Therefore, it remains to search for other methods (based on

multivariate t) of constructing SCI for ratios of the levels of a fixed factor in the presence of random factor(s).

Sequential analysis- Sequential approaches can alleviate difficulties with the coverage probabilities and expected diameter of Fieller confidence intervals in non-sequential settings. Gleser and Hwang (1987) showed the impossibility of constructing confidence intervals (for ratios) which have both positive confidence and finite expected length. This problem is also discussed by Hwang and Liu (1990, 1992). In particular, Hwang and Liu (1990) proposed a fully sequential procedure of constructing simultaneous confidence sequences which overcomes the problems with non-sequential and finite stage sequential procedures. They also compared with the Scheffé type confidence intervals. Therefore, in situations where it gives sense to implement a fully sequential procedure, it might be of interest to further investigate these methods with a possible improvement.

Variance heterogeneity- Variance homoscedasticity is a very common assumption in the general linear model. However, this can be an unrealistic assumption for some datasets. In inference for ratio of means of two normal distributions with unknown and heterogeneous variances, Mendoza and Gutiérrez-Pena (1999) discuss Bayesian methods of constructing confidence intervals for the ratio while Lee and Lin (2004) use generalized confidence intervals. In multiple comparisons, Tamhane and Logan (2004) utilized Welch-

Satterthwaite approximations. This problem is not adequately addressed in the context of simultaneous estimation of ratios. We conjecture that similar approximations (with a plug-in estimate for the number of degrees of freedom and/or the correlation matrix) can be used to develop SCI for ratios from heteroscedastic data.

Bayesian analysis- The Bayesian approach is also an intuitively attracting area of research. As pointed out in Section 5.2, Fieller solutions can result in unbounded (simultaneous) confidence intervals. See also Appendix C for several examples on unbounded two-sided confidence sets in the case of two ratios. This is not much a problem in Bayesian statistics. For example, Gelman *et al.* (2004) analyzed a bioassay data in which the two-sided Fieller confidence interval for $LD50$ results in $(-\infty, \infty)$. Whereas, the Bayesian interval with independent non-informative priors is given by (-0.277, 0.125). As described by Buonaccorsi and Gatsonis (1988), an obvious advantage of a Bayesian approach is that it enables one to express the posterior information on the parameter of interest (ratio) using finite length probability regions. Even when we have finite length Fieller intervals, another important feature of the Bayesian approach is that the prior distributions can be tuned to have desirable credible intervals for the ratios (e.g., not to include negative values in the intervals). Bayesian approach for ratios is also addressed by Mendoza (1990) for slope-ratio bioassay, Mendoza and Gutiérrez-Pena (1999) for the

ratio of the means of two normal populations, Ghosh *et al.* (2003) for ratios of regression coefficients in linear models, Heitjan *et al.* (1999) for cost-effectiveness ratios analysis, and others.

The idea is now to extend the existing Bayesian confidence intervals (credibility intervals) for a single ratio to that of constructing joint credibility intervals for multiple ratios. There are already methods for constructing joint credibility intervals for inferences based on the difference of means (Westfall *et al.*, 1999; Liu and Hayter, 2001). We leave a detailed discussion of Bayesian simultaneous confidence intervals for multiple ratios for a future research, but here we give two general steps on how these intervals can be obtained. First, generate samples from the posterior distributions of the ratios of interest. This can be done, for example, using WinBUGS (see Spiegelhalter *et al.*, 2003). Then, use the *BayesIntervals* Macro in SAS (Westfall *et al.*, 1999, p. 358) to get the simultaneous confidence intervals.

References

Bennett, B.M. (1961). Confidence limits for multivariate ratios. *Journal of the Royal Statistical Society, Series B* **23**, 08-112.

Beran, R. (1987). Prepivoting to reduce level error of confidence sets. *Biometrika* **74**, 457-468.

Beran, R. (1988). Balanced simultaneous confidence sets. *Journal of the American Statistical Association* **83**, 679-686.

Biesheuvel, E. (2002). Many-to-one comparisons in stratified designs. Unpublished Ph.D. Thesis, Faculty of Horticulture, University of Hannover, Germany, pp. 77-92.

Blackwelder, W.C. (1982). 'Proving the null hypothesis' in clinical trials. *Controlled Clinical Trials* **3**, 345-353.

Bretz, F., Genz, A. and Hothorn, L.A. (2001). On the numerical availability of multiple comparison procedures. *Biometrical Journal* **43**, 645-656.

Bretz, F., Hothorn, L.A. and Hsu, J. (2003). Identifying effective and/or safe doses by stepwise confidence intervals for ratios. *Statistics in Medicine* **22**, 847-858.

Briggs, A.H., Mooney, C.Z. and Wonderling, D.E. (1999). Constructing confidence intervals for cost-effectiveness ratios: An evaluation of parametric

and non-parametric techniques using Monte Carlo simulation. *Statistics in Medicine* **18**, 3245-3262.

Buonaccorsi, J.P. and Iyer, H.K. (1984). A comparison of confidence regions and designs in estimaton of a ratio. *Communications in Statistics-Theory and Methods* **13**, 723-741.

Buonaccorsi, J.P. (1985). Ratios of linear combinations in the general linear model. *Communications in Statistics-Theory and Methods* **14**, 635-650.

Buonaccorsi, J. P. and Gatsonis, A.G. (1988). Bayesian inference for ratios of coefficients in a linear model. *Biometrics* **44**, 87-101.

Chikuse, Y. (1981). Simultaneous pseudo confidence regions for ratios of the discriminant coefficients. *Communications in Statistics-Theory and Methods* **10**, 531-547.

Cox, D.R. (1967). Fieller's theorem and a generalization. *Biometrika* **54**, 567-572.

D'Agostino, R.B.Sr., Massaro, J.M. and Sullivan, L.M. (2003). Non-inferiority trials: design concepts and issues - the encounters of academic consultants in statistics. *Statistics in Medicine* **22**, 169-186.

Dilba, G., Bretz, F., Guiard, V. and Hothorn, L.A. (2004). Simultaneous confidence intervals for ratios with applications to the comparison of several treatments with a control. *Methods of Information in Medicine*

43, 465-469.

Dilba, G., Bretz, F. and Guiard, V. Simultaneous confidence sets and confidence intervals for multiple ratios. *Journal of Statistical Planning and Inference* (in press).

Dilba, G., Bretz, F., Hothorn, L.A. and Guiard, V. (2005). Power and sample size computations in simultaneous tests for non-inferiority based on relative margins. *Statistics in Medicine* **24**, 000-000.

Dunnett, C.W. (1955). A multiple comparison procedure for comparing several treatments with a control. *Journal of the American Statistical Association* **50**, 1096-1121.

Faraggi, D., Izikson, P. and Reiser, B. (2003). Confidence intervals for the 50 per cent response dose. *Statistics in Medicine* **22**, 1977-1988.

Fieller, E.C. (1954). Some problems in interval estimation. *Journal of the Royal Statistical Society, Series B* **16**, 175-185.

Fox, D.R. (1991). Interval estimation in caliberation models: An application of Fieller's theorem. *Communications in Statistics-Theory and Methods* **20**, 711-720.

Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D.B. (2004). *Bayesian Data Analysis*, 2nd ed, Chapman & Hall/CRC, pp. 88-93.

Genz, A. and Bretz, F. (1999). Numerical computation of multivariate t-probabilities with application to power calculation of multiple contrasts. *Journal of Statistical Computation and Simulation* **63**, 361-378.

Genz, A. and Bretz, F. (2002). Methods for the computation of multivariate t-probabilities. *Journal of Computational and Graphical Statistics* **11**, 950-971.

Ghosh, M., Yin, M. and Kim, Y. (2003). Objective Bayesian inference for ratios of regression coefficients in linear models. *Statistica Sinica* **13**, 409-422.

Gleser, L.Y. and Hwang, J.T. (1987). The nonexistence of $100(1-\alpha)\%$ confidence sets of finite expected diameter in errors-in-variables and related models. *Annals of Statistics* **15**, 1351-1362.

Gold, M.R., Siegel, J.E., Russell, L.B. and Weinstein, M.C. (1996). Cost-effectiveness in health and medicine. New York: Oxford University Press.

Guiard, V. (2002). Studie zum einseitigen simultanen Konfidenzintervall im Fieller-Problem. Technical report. Teaching Unit Bioinformatics, University of Hannover.

Hauschke, D. and Kieser, M. (2001). Multiple testing to establish non-inferiority of k treatments with a reference based on the ratio of two means. *Drug Information Journal* **35**, 1247-1251.

Hauschke, D., Kieser, M. and Hothorn, L.A. (1999a). Proof of safety in toxicology based on the ratio of two means for normally distributed data. *Biometrical Journal* **41**, 295-304.

Hauschke, D., Kieser, M., Diletti, E. and Burke, M. (1999b). Sample size determination for proving equivalence based on the ratio of two means for normally distributed data. *Statistics in Medicine* **18**, 93-105.

Heitjan, D.F., Moskowitz, A.J. and Whang, W. (1999). Bayesian estimation of cost-effectiveness ratios from clinical trials. *Health Economics* **8**, 191-201.

Hochberg, J. and Tamhane, A. (1987). *Multiple comparison procedures*. Wiley, New York.

Horn, M. and Vollandt, R. (1998). Sample sizes for comparisons of k treatments with a control based on different definitions of power. *Biometrical Journal* **40**, 589-612.

Hothorn, L.A. and Bauss, F. (2004). Biostatistical design and analyses of long-term animal studies simulating human postmenopausal osteoporosis. *Drug Information Journal* **38**, 47-56.

Hothorn, T., Bretz, F. and Genz, A. (2001). On Multivariate t and Gauss Probabilities. *R News* **1**/2, 27-29.

Hsu, J.C. (1996). *Multiple Comparisons*. Chapman & Hall, London.

Hsu, J.C., Hwang, J.T.G., Liu, H.K. and Ruberg, S.J. (1994). Confidence intervals associated with tests for bioequivalence. *Biometrika* **81**, 103-114.

Hwang, I.K. and Morikawa, T. (1999). Design issues in noninferiority/equivalence trials. *Drug Information Journal* **33**, 1205-1218.

Hwang, J.T. and Liu, H. (1990). Sequential confidence regions in inverse regression problems. *Annals of Statistics* **18**, 1389-1399.

Hwang, J.T. and Lui, H.K. (1992). Existence and nonexistence theorems of finite diameter sequential confidence regions for errors-in-variables models. *Statistics and Probability Letters* **13**, 45-55.

Jensen, D.R. (1989). Joint confidence sets in multiple dilution assays. *Biometrical Journal* **31**, 841-853.

Kanis, J.A., Oden, A., Johnell, O., Caulin, F., Bone, H., Alexandre, J.M., Abadie, E. and Lekkerkerker, F. (2002). Uncertain future of trials in osteoporosis. *Osteoporosis International* **13**, 443-449.

Kendall, M.G., 1999. *The Advanced Theory of Statistics*, Vol. 2A, pp. 126-127.

Knapp, H.H., Schrott, H., Ma, P., Knopp, R., Chin, B., Gaziano, J.M., Donovan, J.M., Burke, S.K. and Davidson, M.H. (2001). Efficacy and safety of combination simvastatin and colesevelam in patients with pri-

mary hypercholesterolemia. *The American Journal of Medicine* **110**, 352-360.

Koschat, M.A. (1987). A characterization of the Fieller solution. *Annals of Statistics* **15**, 462-468.

Kotz, S. and Nadarajah, S. (2004). *Multivariate t Distributions and Their Applications*. UK, Cambridge University Press.

Kshirsagar, A.M. (1961). Some extensions of the multivariate t-distribution and the multivariate generalization of the distribution of the regression coefficient. *Proceedings of the Cambridge Philosophical Society* **57**, 80-85.

Laska, E.M., Meisner, M. and Siegel, C. (1997). Statistical inference for cost-effectiveness ratios. *Health Economics* **6**, 229-242.

Laster, L.L. and Johnson, M.F. (2003). Non-inferiority trials: the 'at least as good as' criterion. *Statistics in Medicine* **22**, 187-200.

Lee, J.C. and Lin, S. (2004). Generalized confidence intervals for the ratio of means of two normal populations. *Journal of Statistical Planning and Inference* **123**, 49-60.

Lee, J.J. (1998). Calibration. In *Encyclopedia of Biostatistics*. John Wiley & Sons Ltd, pp. 461-466.

Liu, W. and Hayter, A.J. (2001). Selecting and sharpening inferences in

simultaneous inference with a Bayesian approach. *Communications in Statistics-Theory and Methods* **30**, 135-145.

Malley, J.D. (1982). Simultaneous confidence intervals for ratios of normal means. *Journal of the American Statistical Association* **77**, 170-176.

Mendoza, M. (1990). A Bayesian analysis of slope ratio bioassay. *Biometrics* **46**, 1059-1069.

Mendoza, M. and Gutiérrez-Pena, E. (1999). Bayesian inference for the ratio of the means of two normal populations with unequal variances. *Biometrical Journal* **41**, 133-147.

Miller, R.G. (1977). Developments in Multiple Comparisons 1966-1976. *Journal of the American Statistical Association* **72**, 779-788.

Miller, R.G. (1981). *Simultaneous Statistical Inference.* 2nd ed., Springer, New York.

Mutui, T.M. (2005). Physiological and molecular effects of Thidiazuron and ethylene on leaf yellowing and rooting of Pelargonium of (*Pelargonium zonale* hybrids) cuttings. Ph.D. Thesis in Floriculture, Institute of Floriculture, Tree Nursery and Plant Breeding, Faculty of Natural Sciences, University of Hannover, Germany.

Mutui, T.M., Mibus, H., Serek, M. (2005). Effects of thidiazuron, ethylene, abscisic acid and dark storage on leaf yellowing and rooting of *Pelargo-*

nium cuttings. *Journal of Horticultural Science & Biotechnology* **80**, 000-000.

Pigeot, I., Schäfer, J., Röhmel, J. and Hauschke, D. (2003). Assessing non-inferiority of a new treatment in a three-arm clinical trial including a placebo. *Statistics in Medicine* **22**, 883-899.

Polsky, D., Glick, H.A., Willke, R. and Schulman, K. (1997). Confidence intervals for cost-effectiveness ratios: A comparison of methods. *Health Economics* **6**, 243-252.

Rashid, M.M. (2003). Rank-based tests for non-inferiority and equivalence hypotheses in multi-centre clinical trials using mixed models. *Statistics in Medicine* **22**, 291-311.

Röhmel, J. (1998). Therapeutic equivalence investigations: statistical considerations. *Statistics in Medicine* **17**, 1703-1714.

Roy, S.N. (1953). On heurestic method of test construction and its use in multivariate analysis. *The Annals of Mathematical Statistics* **24**, 220-238.

Scheffé, H. (1970). Multiple testing versus multiple estimation, improper confidence sets, estimation of directions and ratios. *The Annals of Mathematical Statistics* **41**, 1-29.

Sen, P.K. (1998). Biological Assay, Overview. In *Encyclopedia of biostatis-*

tics. John Wiley & Sons Ltd, pp. 372-384.

Šidák, Z. (1967). Rectangular confidence regions for the means of multivariate normal distributions. *Journal of the American Statistical Association* **62**, 626-633.

Spiegelhalter, D., Thomas, A., Best, N. and Lunn, D. (2003). *WinBUGS User Manual*, version 1.4, http://www.mrc-bsu.cam.ac.uk/bugs.

Steffens, F.E. (1971). On confidence sets for the ratio of two normal means. *South African Statistical Journal* **5**, 105-113.

Tamhane, A.C. and Logan, B.R. (2004). Finding the maximum safe dose level for heteroscedastic data. *Journal of Biopharmaceutical Statistics* **14**, 843-856.

Tamhane, A.C., Shi, K. and Strassburger, K. Power and sample size determination for a stepwise test procedure for finding the maximum safe dose. *Journal of Statistical Planning and Inference* (to appear).

Tong, Y.L. (1990). *The Multivariate Normal Distribution.* Springer-Verlag: New York; 191-3.

Westfall, P.H., Tobias, R.D., Rom, D., Wolfinger, R.D. and Hockberg, Y. (1999). *Multiple comparisons and multiple tests using the SAS system.* Cary, NC, SAS Institute Inc.

Westfall, P.H. and Young, S.S. (1993). *Resampling-Based Multiple Testing.* Wiley, New York.

Young, D.A., Zerbe, G.O. and Hay, W.Jr. (1997). Fieller's Theorem, Scheffé's simultaneous confidence intervals, and ratios of parameters of linear and nonlinear mixed-effect models. *Biometrics* **53**, 835-847.

Zerbe, G.O., Laska E., Meisner, M. and Kushner, H.B. (1982). On multivariate confidence regions and simultaneous confidence limits for ratios. *Communications in Statistics-Theory and Methods* **11**, 2401-2425.

Appendix A

R Code for Sample Size Determination

Power and sample size computations are done in R, an open source statistical software package available at www.r-project.org. For the computation of multivariate t probabilities and the equi-coordinate percentage points c, the *pmvt* function from the *mvtnorm* package (a package which computes multivariate normal probabilities and critical points) is used, which is also available under the URL above. See Hothorn *et al.* (2001) for further details. The code below is for the ratio-based inference when large response values indicate better treatment effects, but it can also be easily modified for inferences based on differences by redefining the correlation matrix and the non-centrality parameters. In the program, *n.ratio* is a function that computes the smallest sample size per treatment (balanced design) given the number of comparisons r, the value of $m1$ (often there is no prior information about m, therefore, $m1 = 1$), the relative non-inferiority margin (*psi*), the minimal power (*min.power*), the coefficient of variation of the control group ($CV0$), the clinically irrelevant percentage of the control to be detected (*theta.star*), the familywise type I error rate (*alpha*), and a starting value for the sample size (*n.start*).

```
 library (mvtnorm)
 n.ratio <- function (r, m1, psi, min.power, CV0,theta.star,
                      alpha,n.start) {
 rho <- (psi^2)/sqrt((psi^2 + 1)*(psi^2 +1))
 RHO <- matrix(rep(rho,r*r), nr=r)
diag(RHO) <- rep(1,r) # correlation matrix (balanced design)


 n <- n.start
 power <- 0
 eps <- 0.00001
 while(power < min.power) {
   nu <- (r+1)*(n-1)
   probq <- function(q){
   pmvt(lower=rep(-Inf,r),upper=rep(q,r),nu,corr=RHO,
        delta=rep(0,r),abseps=eps)-(1-alpha)}
   cp <- uniroot(probq, lower=0, upper=4)$root #computes c

   theta.vec <- rep(theta.star,m1)
   deltaR <- (theta.vec - psi)/(CV0*sqrt(1/n + (psi^2)/n)) #ncp
   RHO.LFC <- matrix(rep(rho,m1*m1), nr=m1)
   diag(RHO.LFC) <- rep(1,m1)
   power <- 1-pmvt(lower=rep(-Inf,m1),upper=rep(cp,m1),nu,
```

```
                corr=RHO.LFC,delta=deltaR,abseps=eps)
  n <- n + 1
  }
cbind(c(sample.size=round(n-1,0),power=round(power,4))) }
```

For example, for the pharmacological study in Sections 2.3 and 4.9, we have $r = 3$, $m_1 = 1$, $\psi = 0.70$, $1 - \beta = 0.80$, $CV_0 = 0.5$, $\theta^* = 0.95$ and $\alpha = 0.05$. Set the starting value for the sample size to 2 and run the following command.

```
n.ratio(r=3, m1=1, psi=0.7, min.power=0.8, CV0=0.5,
        theta.star=0.95,alpha=0.05, n.start=2)
```

For sample size computations based on the complete power, we change only the fifth line from the end of the *n.ratio* function to the following command (replacing $m1$ by $m2$ and *min.power* by *com.power*).

```
power <- pmvt(lower=rep(cp,m2),upper=rep(Inf,m2),nu,
              corr=RHO,delta=deltaR,abseps=eps)
```

It should also be remarked that sample size calculation based on the normal approximation (Section 4.6) is very fast and can be directly calculated in many standard statistical softwares.

Appendix B

Appendix B1: An R function for constructing simultaneous confidence intervals

```
##  Description: The function sci.ratio constructs SCI for ratios to
##                the control treatment using the Bonferroni adjustment,
##                Sidak (Slepian) and the plug-in approaches.
##  Inputs: User dataframe with two columns (treatment and response)
##            control.T - the control group
##            alternative - two.sided or one.sided
##            conf.level - confidence level

sci.ratio <- function(User.DataFr, control.T='',alternative = 'two.sided',
                      conf.level = 0.95) {
library(mvtnorm)
if (is.numeric(User.DataFr[,1])){
    Response <- User.DataFr[,1]
    Treatment<- User.DataFr[,2]}
else {
    Response <- User.DataFr[,2]
    Treatment<- User.DataFr[,1]}
```

```
Cont.Flag <- Treatment==control.T

Cont.Posi <- order(unique(Treatment))[unique(Treatment)==control.T]

if (control.T==’’) stop(’The control group is not specified.’)

if (sum(unique(Treatment)==control.T)==0)

    stop(’Incorrect name of the control group’)


n.Control  <- length(Response[Cont.Flag])

ybar.Contr <- mean(Response[Cont.Flag])

n.Treat <- tapply(Response[!Cont.Flag],Treatment[!Cont.Flag],

                  length)[-Cont.Posi]

k <- length(n.Treat)  # Number of comparisons with a control

d.f <- n.Control + sum(n.Treat) - (k+1)

ybar.Treat <- tapply(Response[!Cont.Flag],Treatment[!Cont.Flag],

                     mean)[-Cont.Posi]

var.vec <- tapply(Response,Treatment,var)    # k+1 groups

n.vec <- tapply(Response,Treatment,length)

s <- sqrt(sum(var.vec*(n.vec - 1)/d.f)) # estimate of sigma


gammaC.vec <- ybar.Treat/ybar.Contr


Quad.root <- function(ratioV, gValue, nC, nT){
```

```
        Discrimi <- gValue*(ratioV^2 + (1-gValue)*nC/nT)
        if ((gValue < 1)&(Discrimi >= 0)) {
           Limit.s <- (ratioV + plus.minus*sqrt(Discrimi))/(1-gValue)}
        else  Limit.s <- 'NSC'
        return(Limit.s)}
cat("                                        ","\n")
if (alternative=='two.sided'){
    side <- 2
    plus.minus <- c(-1,1)
    #
    # BONFERRONI
    #
    cpBon <- qt(1- (1-conf.level)/(side*k), d.f, lower.tail = TRUE)
    gBon  <- (cpBon^2)*(s^2)/(n.Control*(ybar.Contr^2))
    #
    #  MtI
    #
    qmt <- function(q) {pmvt(rep(-q,k),rep(q,k),d.f,corr=diag(k),
           delta=rep(0,k))-conf.level}
    cpMtI <- uniroot(qmt, lower=0, upper=4)$root
    gMtI  <- (cpMtI^2)*(s^2)/(n.Control*(ybar.Contr^2))
    #
```

```
    #  Plug-in
    #
    Corr.Mat <- matrix(rep(NA,k*k),nr=k)
    for(i in 1:k) {
        for(j in 1:k) {
            Corr.Mat[i,j] <- (gammaC.vec[i]*gammaC.vec[j])/
            sqrt((gammaC.vec[i]^2 + n.Control/n.Treat[i])*(gammaC.vec[j]^2
                  + n.Control/n.Treat[j]))
        }
    }
    diag(Corr.Mat) <- rep(1,k)

    qmt0 <- function(q) {pmvt(rep(-q,k),rep(q,k),d.f,corr=Corr.Mat,
                        delta=rep(0,k))-conf.level}
    Cplug <- uniroot(qmt0, lower=0, upper=4)$root
    gPlug  <- (Cplug^2)*s^2/(n.Control*(ybar.Contr^2))
    } # End of two-sided CI

if ((alternative=='less')|(alternative=='greater')){
    side <- 1
    if (alternative=='less') plus.minus <- 1
    else plus.minus <- -1
```

```
#
# BONFERRONI
#
cpBon <- qt(1- (1-conf.level)/(side*k), d.f, lower.tail = TRUE)
gBon  <- (cpBon^2)*(s^2)/(n.Control*(ybar.Contr^2))
#
#  MtI
#
qmt <- function(q) {pmvt(rep(-Inf,k),rep(q,k),d.f,corr=diag(k),
       delta=rep(0,k))-conf.level}
cpMtI <- uniroot(qmt, lower=0, upper=4)$root
gMtI  <- (cpMtI^2)*(s^2)/(n.Control*(ybar.Contr^2))
#
#  Plug-in
#
Corr.Mat <- matrix(rep(NA,k*k),nr=k)
for(i in 1:k) {
    for(j in 1:k) {
        Corr.Mat[i,j] <- (gammaC.vec[i]*gammaC.vec[j])/
        sqrt((gammaC.vec[i]^2 + n.Control/n.Treat[i])*(gammaC.vec[j]^2
              + n.Control/n.Treat[j]))
    }
```

```
    }
    diag(Corr.Mat) <- rep(1,k)

    qmt0 <- function(q) {pmvt(rep(-Inf,k),rep(q,k),d.f,corr=Corr.Mat,
                        delta=rep(0,k))-conf.level}
    Cplug <- uniroot(qmt0, lower=0, upper=4)$root
    gPlug  <- (Cplug^2)*s^2/(n.Control*(ybar.Contr^2))
    } # End of one-sided CI

BonCL <- MtICL <- PlugCL <- matrix(rep(NA,side*k),nr=k)
for(j in 1:k) {
    BonCL[j,]  <- Quad.root(gammaC.vec[j], gBon,  n.Control, n.Treat[j])
    MtICL[j,]  <- Quad.root(gammaC.vec[j], gMtI,  n.Control, n.Treat[j])
    PlugCL[j,] <- Quad.root(gammaC.vec[j], gPlug, n.Control, n.Treat[j])
    }

sci.table <- round(data.frame(gammaC.vec,PlugCL,MtICL,BonCL),3)
cat("                                        ","\n")
cat("                                        ","\n")
if (alternative=='two.sided'){
names(sci.table) <- c('Point.Estimate','Plug.Lower','Plug.Upper',
'Sidak.Lower','Sidak.Upper','Bon.Lower','Bon.Upper')
```

```
cat("Two-sided",conf.level*100, "%",
    "simultaneous confidence intervals for ratios to control:","\n")}
if (alternative=='less'){
names(sci.table) <- c('Point.Estimate','Plug.Upper',
      'Slepian.Upper','Bon.Upper')
cat("Upper",conf.level*100, "%",
    "simultaneous confidence limits for ratios to control:","\n")}
else if (alternative=='greater'){
names(sci.table) <- c('Point.Estimate','Plug.Lower',
     'Slepian.Lower','Bon.Lower')
cat("Lower",conf.level*100, "%",
    "simultaneous confidence limits for ratios to control:","\n")}
cat("                                        ","\n")
print(sci.table)
if (sum(sci.table[,2]=='NSC')>0){
cat("                                        ","\n")
cat("   NSC = Mean of the control group is not significantly
              different from zero. ","\n")}
cat("                                        ","\n")
}   # END OF sci.ratio
```

```
##############
##  EXAMPLE  #
##############
  sample0 <- c( 107,91,115,90,133,95,112,115,117,91)   #  Control
  sample1 <- c(119,88,84,133,87,118,132)               #  Thyroxin
  sample2 <- c(61,68,89,80,69,52,80,63,63,68)          #  Thiouracil
Y <- c(sample0,sample1,sample2)
Treat <- factor(rep(c('Control','Thyroxin','Thiouracil'),c(10,7,10)))
Data <- data.frame(Treat,Y)

sci.ratio(Data,control.T='Control')

Two-sided 95 % simultaneous confidence intervals for ratios to control:

           Point.Estimate Plug.Lower Plug.Upper Sidak.Lower Sidak.Upper
Thiouracil           0.65      0.527      0.789       0.526       0.790
Thyroxin             1.02      0.860      1.205       0.859       1.206
           Bon.Lower Bon.Upper
Thiouracil     0.526     0.790
Thyroxin       0.858     1.207
```

Appendix B2: Example of an R code for constructing one-sided simultaneous confidence set ($k = 3$)

```
library(mvtnorm)
CLevel <- 0.95
gridsize <- 20

n0<-10; n1<- 11; n2<- 13; n3<-13
xbar0 <- 0.82; xbar1 <- 2.19; xbar2 <- 2.75; xbar3<- 3.75; s<- 0.8

gamma1C<-xbar1/xbar0; gamma2C<-xbar2/xbar0; gamma3C<-xbar3/xbar0

cpBon <- qt(1- (1-CLevel)/3,df,lower.tail = TRUE) #Bon. adjustment
gBon  <- (cpBon^2)*(s^2)/(n0*xbar0*xbar0)

Qroot.U <- function(ratioV, gValue, nC, nT){
  (ratioV + sqrt(gValue*(ratioV^2 + (1-gValue)*nC/nT)))/(1-gValue)}
BonSUCL1 <- Qroot.U(gamma1C, gBon, n0, n1)
BonSUCL2 <- Qroot.U(gamma2C, gBon, n0, n2)
BonSUCL3 <- Qroot.U(gamma3C, gBon, n0, n3)
```

```
BonSUCL1; BonSUCL2; BonSUCL3


thetha1s <- seq(0, BonSUCL1,length=gridsize)

thetha2s <- seq(0, BonSUCL2,length=gridsize)

thetha3s <- seq(0, BonSUCL3,length=gridsize)

thethaH0 <- matrix(NA,nr=gridsize^3,nc=3)

##

## ONE-SIDED CONFIDENCE SET

##

L <- 1

for (i in 1:gridsize) {

  for (j in 1:gridsize) {

    for (l in 1:gridsize) {

      rho012 <- (thetha1s[i]*thetha2s[j])/

                sqrt((thetha1s[i]^2 +n0/n1)*(thetha2s[j]^2 +n0/n2))

      rho013 <- (thetha1s[i]*thetha3s[l])/

                sqrt((thetha1s[i]^2 +n0/n1)*(thetha3s[l]^2 +n0/n3))

      rho023 <- (thetha2s[j]*thetha3s[l])/

                sqrt((thetha2s[j]^2 +n0/n2)*(thetha3s[l]^2 +n0/n3))

      Rho0 <-matrix(c(1,rho012,rho013,rho012,1,rho023,rho013,

                      rho023,1),nrow=3,byrow=T)
```

```
      qmt <- function(q) {pmvt(c(-Inf,-Inf,-Inf),c(q,q,q),df,
                          corr=Rho0,delta=c(0,0,0))-CLevel}
      cp <- uniroot(qmt, lower=0, upper=4)$root

      T1  <- (thetha1s[i]*xbar0-xbar1)/
               (s*sqrt((1/n1 + (thetha1s[i]^2)/n0)))
      T2  <- (thetha2s[j]*xbar0-xbar2)/
              (s*sqrt((1/n2 + (thetha2s[j]^2)/n0)))
      T3  <- (thetha3s[l]*xbar0-xbar3)/
               (s*sqrt((1/n3 + (thetha3s[l]^2)/n0)))
      if ((T1 <= cp)&(T2 <= cp)&(T3 <= cp))
           thethaH0[L,] <- c(thetha1s[i],thetha2s[j],thetha3s[l])
      L <- L + 1
      }
  }
}
thethaH02 <- thethaH0
thethaH02[is.na(thethaH02)] <- 0  # This works for gammas >= 0

group <- c(1:gridsize)
```

```
BoundZ <- matrix(NA,nr=gridsize,nc=gridsize)
for (i in 1:gridsize) {
    for (j in 1:gridsize) {
        BoundZ[i,j] <- max(thethaH02[group,3])
        group <- group + gridsize
    }
 }
BoundZ[BoundZ==0] <- NA

x <- thetha1s; y <- thetha2s; z <- BoundZ

persp(x, y, z, theta = 120, phi = 40, expand=0.7, col="grey91",
     ltheta = 120, shade = 0.75, ticktype = "detailed",ntick=6,
     xlab = '', ylab = '',xlim=c(0,BonSUCL1),ylim=c(0,BonSUCL2),
     zlim=c(0,BonSUCL3), zlab = '', box=TRUE) ->res

##   ADDING POINTS & LINES TO 3D PLOT
trans3d <- function(x,y,z, pmat) {
           tr <- cbind(x,y,z,1) %*% pmat
           list(x = tr[,1]/tr[,4], y= tr[,2]/tr[,4])
           }
```

```
text(trans3d(4, 12, z= 0, pm = res),expression(gamma[1]))

text(trans3d(10, 5, z= 0, pm = res),expression(gamma[2]))

mtext(side=2,expression(gamma[3]))


##  Superimposing Bonferroni SCI

##

lines (trans3d(x, y=BonSUCL2, z= BonSUCL3, pm=res), lty=2,lwd=3)

lines (trans3d(BonSUCL1, y, z=BonSUCL3, pm=res), lty=2,lwd=3)

lines (trans3d(BonSUCL1, BonSUCL2,seq(0,BonSUCL3,length=20),

        pm = res), lty=2,lwd=3)

points(trans3d(BonSUCL1,BonSUCL2,BonSUCL3,pm=res),pch=16,cex=1.4)
```

Appendix B3: Simulation program for comparing the coverage probabilities of methods for constructing SCI (Section 5.6)

```
#  Inputs:
#
#     N = Number of iterations
#     B = Number of bootstrap samples
#   CV0 = Coefficient of variation of the control group
#   mu0 = mean of the control group
#   mu1 = mean of treatment 1
#   mu2 = mean of treatment 2
#   mu3 = mean of treatment 3

SIM <- function(N,B,CLevel,CV0,mu0,mu1,mu2,mu3,n0,n1,n2,n3) {
  print(paste("                                        "), quote=FALSE)
  print(paste("  ==============================        "), quote=FALSE)
  print(paste("                                        "), quote=FALSE)
  print(paste("  SIMULATION PARAMETERS: "), quote=FALSE)
  print(paste("                                        "), quote=FALSE)
  print(paste("  N = ",N,  " B = ",B), quote=FALSE)
  print(paste("  Confidence Level = ", CLevel*100, "%"), quote=FALSE)
  print(paste("  mu0 = ",mu0, " mu1 = ",mu1, "  mu2 = ",mu2,
```

```
              "  mu3 = ",mu3, " CV0 = ", CV0*100, "%"), quote=FALSE)
  print(paste("   n0 = ",n0,  " n1 = ",n1,
              "  n2 = ",n2,"  n3 = ",n3), quote=FALSE)
  print(paste("                                        "), quote=FALSE)

sigma  <- mu0*CV0
df     <- n0+n1+n2+n3 - 4

Gamma1 <- mu1/mu0; Gamma2 <- mu2/mu0;  Gamma3 <- mu3/mu0

# Elements of the true correlation matrix
rt12    <- (Gamma1*Gamma2)/
                     sqrt((Gamma1^2 +n0/n1)*(Gamma2^2 +n0/n2))
rt13    <- (Gamma1*Gamma3)/
                     sqrt((Gamma1^2 +n0/n1)*(Gamma3^2 +n0/n3))
rt23    <- (Gamma2*Gamma3)/
                     sqrt((Gamma2^2 +n0/n2)*(Gamma3^2 +n0/n3))
Rhot<-matrix(c(1,rt12,rt13,rt12,1,rt23,rt13,rt23,1),nrow=3,byrow=T)
library(mvtnorm)
qmt <- function(q) {pmvt(c(-Inf,-Inf,-Inf),c(q,q,q),df,corr=Rhot,
                   delta=c(0,0,0))-CLevel}
Ct <- uniroot(qmt, lower=0, upper=4)$root
```

```
Ratio<- matrix(NA,nrow=N,ncol=3)

Rp <- Cplug <- Corrboot <- rep(NA,N)

Cpb1 <- CTMax <- cnp <- dpn1 <- dpn2 <- dpn3 <- rep(NA,N)


## Vectors for computing overall coverage probabilities

##

CovPrBon<-CovPrMtI<-CovPrC<-CovPrTMa<-CovPrBab <- rep(NA,N)


## Vectors for computing individual coverage probabilities

##

L1BonCov<-L2BonCov<-L3BonCov<-L1MtICov<-L2MtICov<-L3MtICov<-rep(NA,N)

L1CCov<-L2CCov<-L3CCov<-L1TMaCov<-L2TMaCov<-L3TMaCov<-rep(NA,N)

L1babCov<-L2babCov<-L3babCov<-rep(NA,N)


step <- round(N/10,0)

stage <- step

# Critical points for Bonferroni & MtI (Slepian) SCIs:

 cpBon <- qt(1- (1-CLevel)/3, df, lower.tail = TRUE)


 qmt <- function(q) {pmvt(c(-Inf,-Inf,-Inf),c(q,q,q),df,

          corr=diag(3), delta=c(0,0,0))-CLevel}

 cpMtI <- uniroot(qmt, lower=0, upper=4)$root
```

```
## QUADRATIC ROOT  (Upper confidence limit)
##
 Qroot.U <- function(ratioV, gValue, nC, nT){
  (ratioV + sqrt(gValue*(ratioV^2 + (1-gValue)*nC/nT)))/(1-gValue)}

##
## Loop i
for (i in 1:N) {
 sample0 <- rnorm(n0,mu0,sigma)
 sample1 <- rnorm(n1,mu1,sigma)
 sample2 <- rnorm(n2,mu2,sigma)
 sample3 <- rnorm(n3,mu3,sigma)

 xbar0 <- mean(sample0);  xbar1 <- mean(sample1)
 xbar2 <- mean(sample2);  xbar3 <- mean(sample3)

 v0 <- var(sample0);  v1 <- var(sample1)
 v2 <- var(sample2);  v3 <- var(sample3)

 s2 <- ((n0-1)*v0 + (n1-1)*v1 + (n2-1)*v2
      + (n3-1)*v3)/(n0+n1+n2+n3-4)
```

```
Ratio[i,1] <- xbar1/xbar0

Ratio[i,2] <- xbar2/xbar0

Ratio[i,3] <- xbar3/xbar0

#

#   BONFERRONI

#

gBon  <- (cpBon^2)*(s2)/(n0*xbar0*xbar0)


L1Bon <- Qroot.U(Ratio[i,1], gBon, n0, n1)

L2Bon <- Qroot.U(Ratio[i,2], gBon, n0, n2)

L3Bon <- Qroot.U(Ratio[i,3], gBon, n0, n3)


L1BonCov[i] <- Gamma1 <= L1Bon

L2BonCov[i] <- Gamma2 <= L2Bon

L3BonCov[i] <- Gamma3 <= L3Bon

CovPrBon[i] <- (Gamma1 <= L1Bon)&(Gamma2 <= L2Bon)&(Gamma3 <= L3Bon)

#

#   MtI  (one-sided SCI based on Slepian inequality)

#

gMtI  <- (cpMtI^2)*(s2)/(n0*xbar0*xbar0)
```

```
L1MtI <- Qroot.U(Ratio[i,1], gMtI, n0, n1)

L2MtI <- Qroot.U(Ratio[i,2], gMtI, n0, n2)

L3MtI <- Qroot.U(Ratio[i,3], gMtI, n0, n3)


L1MtICov[i] <- Gamma1 <= L1MtI

L2MtICov[i] <- Gamma2 <= L2MtI

L3MtICov[i] <- Gamma3 <= L3MtI

CovPrMtI[i] <- (Gamma1 <= L1MtI)&(Gamma2 <= L2MtI)&(Gamma3 <= L3MtI)

#

#   PLUG-IN

#

rp12 <- (Ratio[i,1]*Ratio[i,2])/

                  sqrt((Ratio[i,1]^2 +n0/n1)*(Ratio[i,2]^2 +n0/n2))

rp13 <- (Ratio[i,1]*Ratio[i,3])/

                  sqrt((Ratio[i,1]^2 +n0/n1)*(Ratio[i,3]^2 +n0/n3))

rp23 <- (Ratio[i,2]*Ratio[i,3])/

                  sqrt((Ratio[i,2]^2 +n0/n2)*(Ratio[i,3]^2 +n0/n3))

Rp <-matrix(c(1,rp12,rp13,rp12,1,rp23,rp13,rp23,1),nrow=3,byrow=T)

qmt0 <- function(q)

        {pmvt(c(-Inf,-Inf,-Inf),c(q,q,q),df,corr=Rp,

                delta=c(0,0,0))-CLevel}

Cplug[i] <- uniroot(qmt0, lower=0, upper=4)$root
```

```
  #
  # Sim_CI:
gplug  <- (Cplug[i]^2)*s2/(n0*xbar0*xbar0)

L1plug <- Qroot.U(Ratio[i,1], gplug, n0, n1)
L2plug <- Qroot.U(Ratio[i,2], gplug, n0, n2)
L3plug <- Qroot.U(Ratio[i,3], gplug, n0, n3)

L1CCov[i] <- Gamma1 <= L1plug
L2CCov[i] <- Gamma2 <= L2plug
L3CCov[i] <- Gamma3 <= L3plug
CovPrC[i] <- (Gamma1 <= L1plug)&(Gamma2 <= L2plug)&(Gamma3 <= L3plug)

#################################
#  Resampling methods           #
#################################

TMat<-matrix(NA,nrow=B,ncol=3)
TMax <- rep(NA,B)
E0 <- sample0 - mean(sample0)
E1 <- sample1 - mean(sample1)
E2 <- sample2 - mean(sample2)
```

```
E3 <- sample3 - mean(sample3)


#PoolErr <- c(E0,E1,E2,E3)   #  For non-parametric resampling


s <- sqrt(s2)


for (b in 1:B) {
        #
        # Non-parametric
        #
 #bE0 <- sample(PoolErr,n0,replace =TRUE)
 #bE1 <- sample(PoolErr,n1,replace =TRUE)
 #bE2 <- sample(PoolErr,n2,replace =TRUE)
 #bE3 <- sample(PoolErr,n3,replace =TRUE)
       #
       # Parametric
       #
 bE0 <- rnorm(n0,xbar0,s)
 bE1 <- rnorm(n1,xbar1,s)
 bE2 <- rnorm(n2,xbar2,s)
 bE3 <- rnorm(n3,xbar3,s)
```

```
bEbar0 <- mean(bE0)

bEbar1 <- mean(bE1)

bEbar2 <- mean(bE2)

bEbar3 <- mean(bE3)


bv0 <- var(bE0)

bv1 <- var(bE1)

bv2 <- var(bE2)

bv3 <- var(bE3)


bs <- sqrt(((n0-1)*bv0 + (n1-1)*bv1 + (n2-1)*bv2

            + (n3-1)*bv3)/(n0+n1+n2+n3-4))

TMat[b,1] <- (Ratio[i,1]*bEbar0-bEbar1 )/

                (bs*sqrt((1/n1 + (Ratio[i,1]^2)/n0)))

TMat[b,2] <- (Ratio[i,2]*bEbar0-bEbar2 )/

                (bs*sqrt((1/n2 + (Ratio[i,2]^2)/n0)))

TMat[b,3] <- (Ratio[i,3]*bEbar0-bEbar3 )/

                (bs*sqrt((1/n3 + (Ratio[i,3]^2)/n0)))

TMax[b] <- max(TMat[b,1],TMat[b,2],TMat[b,3])

}

#

#   Tmax,     max{T1,T2,T3}
```

```
#
TMax <- sort(TMax)
CTMax[i] <- TMax[ceiling(CLevel*B)]
gTMax  <- (CTMax[i]^2)*s2/(n0*xbar0*xbar0)
L1TMax <- Qroot.U(Ratio[i,1], gTMax, n0, n1)
L2TMax <- Qroot.U(Ratio[i,2], gTMax, n0, n2)
L3TMax <- Qroot.U(Ratio[i,3], gTMax, n0, n3)

L1TMaCov[i] <- Gamma1 <= L1TMax
L2TMaCov[i] <- Gamma2 <= L2TMax
L3TMaCov[i] <- Gamma3 <= L3TMax
CovPrTMa[i]<-(Gamma1 <= L1TMax)&(Gamma2 <= L2TMax)&(Gamma3 <= L3TMax)
#
#  BALANCED RESAMPLING
#
Rn1b <- TMat[,1]          # test statistics, bootstrap version
Rn2b <- TMat[,2]
Rn3b <- TMat[,3]

HpRb1 <- (rank(Rn1b)-1)/B # left-cont. cdfs of the boot. test stat.
HpRb2 <- (rank(Rn2b)-1)/B
HpRb3 <- (rank(Rn3b)-1)/B
```

```
  Snj <- rep(NA,B)


# superimum of the three cdfs at the data
# points in the bootstrap sample
  for (j in 1:B) {Snj[j] <- max(HpRb1[j],HpRb2[j], HpRb3[j])}


   cnp[i] <- quantile(Snj, CLevel + 1/B)


   dpn1[i] <- quantile(Rn1b, cnp[i])
   dpn2[i] <- quantile(Rn2b, cnp[i])
   dpn3[i] <- quantile(Rn3b, cnp[i])


   gbab1  <- (dpn1[i]^2)*s2/(n0*xbar0*xbar0)
   gbab2  <- (dpn2[i]^2)*s2/(n0*xbar0*xbar0)
   gbab3  <- (dpn3[i]^2)*s2/(n0*xbar0*xbar0)


   L1bab <- Qroot.U(Ratio[i,1], gbab1, n0, n1)
   L2bab <- Qroot.U(Ratio[i,2], gbab2, n0, n2)
   L3bab <- Qroot.U(Ratio[i,3], gbab3, n0, n3)


   L1babCov[i] <-  Gamma1 <= L1bab
```

```
   L2babCov[i] <-  Gamma2 <= L2bab

   L3babCov[i] <-  Gamma3 <= L3bab

   CovPrBab[i] <- (Gamma1 <= L1bab)&(Gamma2 <= L2bab)&(Gamma3 <= L3bab)


   # Indicator of the simulation stage:

   #

   if (i==stage) {print(paste(round(100*(i/N),0), "%  completed:",

                  date()), quote=FALSE)

       stage <- stage + step }

 } ##

   ## End of loop i

   ##


Covstats <- data.frame(L1BonCov,L2BonCov,L3BonCov,CovPrBon,

                       L1MtICov,L2MtICov,L3MtICov,CovPrMtI,

                       L1CCov, L2CCov,L3CCov,CovPrC,

                       L1TMaCov,L2TMaCov,L3TMaCov,CovPrTMa,

                       L1babCov,L2babCov,L3babCov,CovPrBab)


CovprMAT <- apply(Covstats+0,2, sum)  # changes T & F to 1 & 0

CovprMAT <- matrix(CovprMAT,nr=4,byrow=FALSE)

CovprMAT
```

```
EstCover <-data.frame(CovprMAT/N)  # Estimate of coverage prob.

names(EstCover) <- c('Bonf','MtI','Plug-in','Tmax','Balanced')

row.names(EstCover) <- c('Gamma1','Gamma2','Gamma3','Gamma')

print(paste(" "), quote=FALSE); print(paste(" "), quote=FALSE)

print(paste("  Estimates of the coverage probabilities: "), quote=FALSE)

print(paste(" "), quote=FALSE)

print(EstCover)

print(paste(" "), quote=FALSE)

   ################################

 } # End of function SIM          ############################

   ################################
```

```
SIM(N=1000,B=1000,CLevel=0.90,CV0=0.20,mu0=1,mu1=2,mu2=3,mu3=4,

            n0=10,n1=10,n2=10,n3=10)
```

```
Output:

[1]

[1]   ==============================
```

```
[1]

[1]   SIMULATION PARAMETERS:

[1]

[1]   N =  1000  B =  1000

[1]   Confidence Level =  90 %

[1]   mu0 =  1  mu1 =  2   mu2 =  3   mu3 =  4  CV0 =  20 %

[1]    n0 =  10  n1 =  10   n2 =  10   n3 =  10

[1]

[1] 10 %  completed: Fri May 20 00:53:15 2005

[1] 20 %  completed: Fri May 20 00:54:25 2005

[1] 30 %  completed: Fri May 20 00:55:35 2005

[1] 40 %  completed: Fri May 20 00:56:45 2005

[1] 50 %  completed: Fri May 20 00:57:56 2005

[1] 60 %  completed: Fri May 20 00:59:07 2005

[1] 70 %  completed: Fri May 20 01:00:16 2005

[1] 80 %  completed: Fri May 20 01:01:26 2005

[1] 90 %  completed: Fri May 20 01:02:37 2005

[1] 100 %  completed: Fri May 20 01:03:48 2005

[1]

[1]

[1]   Estimates of the coverage probabilities:

[1]
```

```
        Bonf   MtI Plug-in  Tmax Balanced

Gamma1 0.968 0.966   0.947 0.948    0.948

Gamma2 0.973 0.970   0.949 0.950    0.947

Gamma3 0.968 0.965   0.946 0.944    0.944

Gamma  0.947 0.943   0.911 0.910    0.906

[1]
```

Appendix C

Examples on unbounded two-sided simultaneous confidence sets. Consider the case of two ratios ($r = 2$): $\gamma_1 = \mu_1/\mu_0$ and $\gamma_2 = \mu_2/\mu_0$.

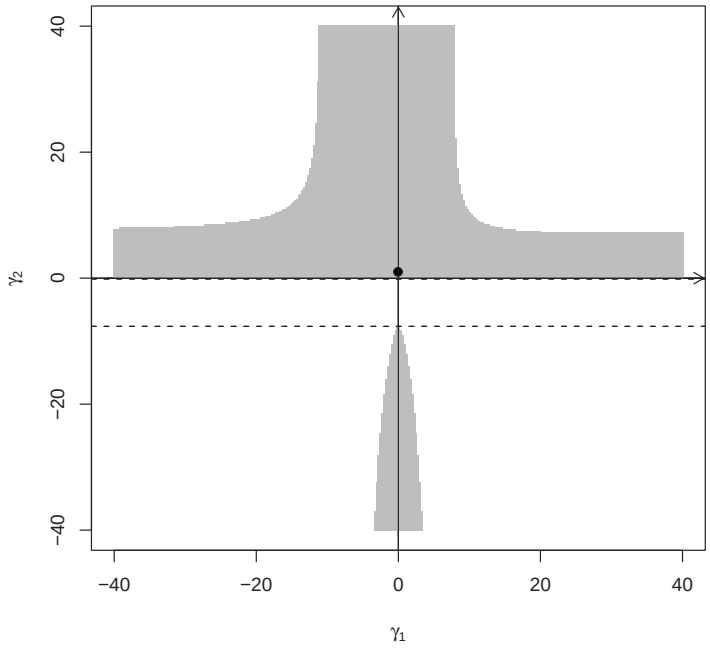

Figure 6.1: *Unbounded SCS for* $\boldsymbol{\gamma} = (\gamma_1, \gamma_2)'$ *with the MtI-SCI given by* $\gamma_1 \in (-\infty, \infty)$ *and* $\gamma_2 \in (-\infty, -7.65) \bigcup (-0.131, \infty)$. *The MtI-SCI limits for* γ_2 *are shown by dotted lines. Data:* $\overline{y}_0 = 0.4$, $\overline{y}_1 = -0.02$, $\overline{y}_2 = 0.4$, $s = 0.6$, $n_0 = n_1 = n_2 = 10$.

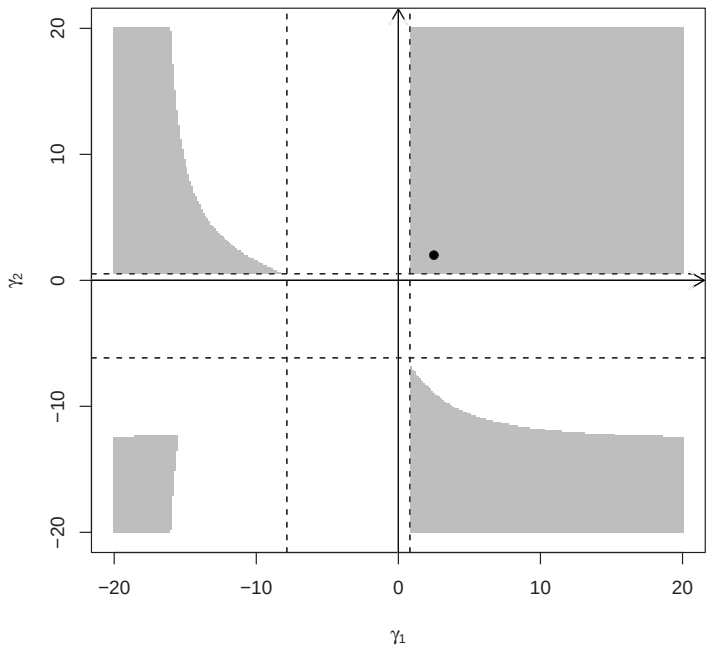

Figure 6.2: *Unbounded SCS for* $\boldsymbol{\gamma} = (\gamma_1, \gamma_2)'$ *with the MtI-SCI given by* $\gamma_1 \in (-\infty, -7.848) \bigcup (0.814, \infty)$ *and* $\gamma_2 \in (-\infty, -6.151) \bigcup (0.524, \infty)$. *Both confidence intervals are complements of finite length intervals. The limits of MtI-SCI are shown by dotted lines. Data:* $\overline{y}_0 = 0.4$, $\overline{y}_1 = 1$, $\overline{y}_2 = 0.8$, $s = 0.7$, $n_0 = n_1 = n_2 = 10$.

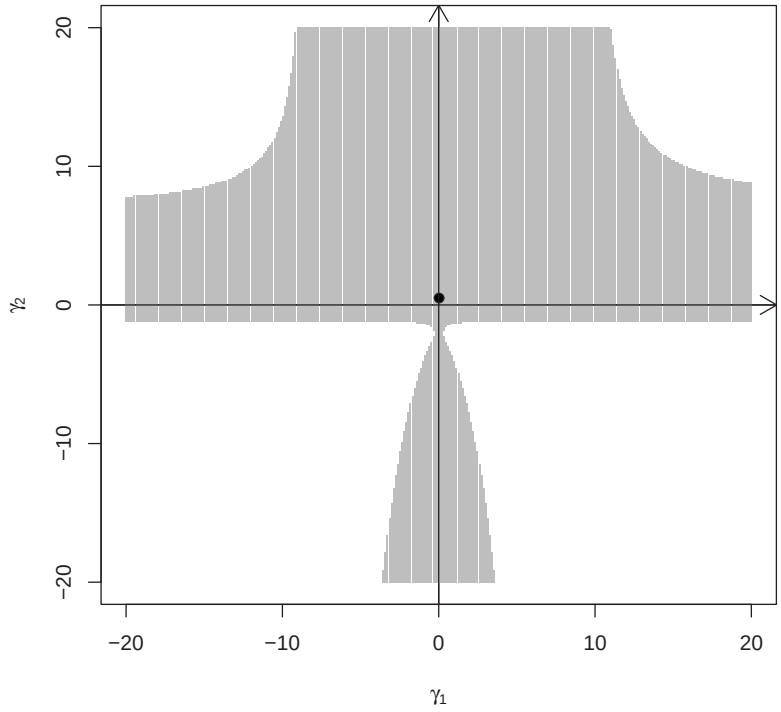

Figure 6.3: *Unbounded SCS for* $\boldsymbol{\gamma} = (\gamma_1, \gamma_2)'$ *with the MtI-SCI given by* $\gamma_1 \in (-\infty, \infty)$ *and* $\gamma_2 \in (-\infty, \infty)$*. The limits of both* γ_1 *and* γ_2 *extend to infinity in all directions. Data:* $\overline{y}_0 = 0.4$*,* $\overline{y}_1 = 0.01$*,* $\overline{y}_2 = 0.2$*,* $s = 0.6$*,* $n_0 = n_1 = n_2 = 10$*.*

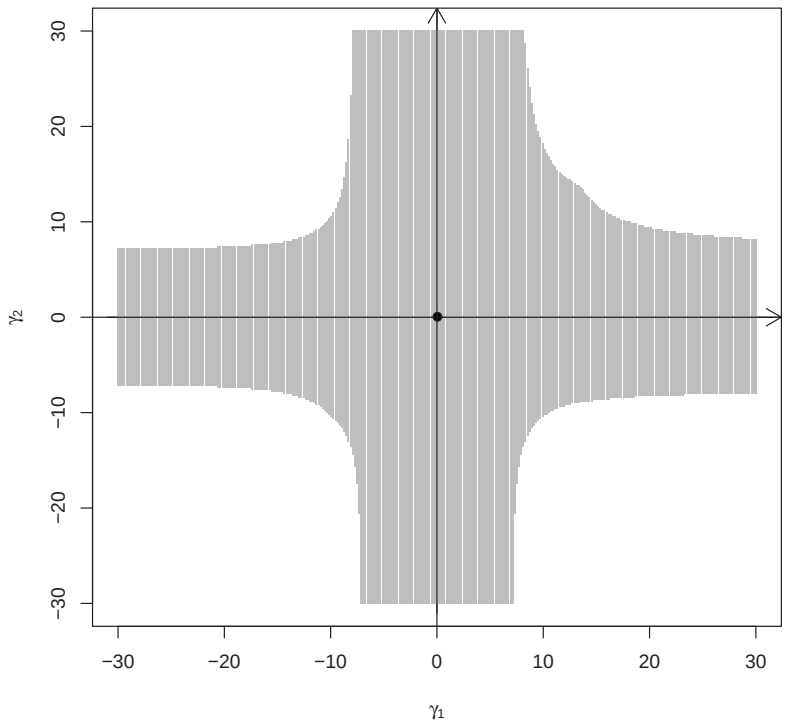

Figure 6.4: *Unbounded SCS for* $\boldsymbol{\gamma} = (\gamma_1, \gamma_2)'$ *with the MtI-SCI given by* $\gamma_1 \in (-\infty, \infty)$ *and* $\gamma_2 \in (-\infty, \infty)$. *The limits of both* γ_1 *and* γ_2 *extend to infinity in all directions. Data:* $\overline{y}_0 = 0.4$, $\overline{y}_1 = 0.02$, $\overline{y}_2 = 0.02$, $s = 0.6$, $n_0 = n_1 = n_2 = 10$.

Acknowledgements

I would like to express my gratitude to my supervisors Prof. Ludwig A. Hothorn and Dr. Frank Bretz for their guidance and encouragements at all stages of this research. I also had a great deal of crucial discussions with Dr. Volker Guiard, thanks indeed.

Many people helped me in order to accomplish this task. I thank from my heart all my colleagues at Lehrgebiet Bioinformatik for their contributions and day-to-day support. In this regard, I would like to thank Mrs. Hannelore Visser and Mr. Clemens Buczilowski for their help. I thank Mrs. Cornelia Frömke for all important discussions we had, and for assisting me in so many practical ways. I thank Mr. Frank Schaarschmidt for taking his precious time to translate the summary into German. I also thank Dr. Lemma Ebssa (Institute of Plant Diseases and Plant Protection) for all lively discussions we had.

I am always indebted to the unfailing love, prayer and encouragement of my parents, three sisters, and brothers throughout my life. My special thanks goes to Nathan and my wife. Bikiltu, thank you for your patience and support on those hectic days. I also thank Prof. Bichaka Fayissa for many words of encouragement and for showing interest in this research. God bless you all!

Gemechis

July 2005

Curriculum Vitae

Personal Data

Name Gemechis Dilba Djira

Address Bioinformatics Unit, Herrenhaeuser Strasse 2, 30419 Hannover

Tel.: 0511 762 5564, Fax: 0511 762 4966

Email: dilba@bioinf.uni-hannover.de

Date and place of birth March 6th, 1968, Aira, Wollega, Ethiopia

Sex Male

Nationality Ethiopian

Marital status Married, one child

Education

9/1974 - 6/1980 *Primary School*, Tschallia and Walgo Aira

9/1980 - 3/1986 *Secondary School*, Lalo Aira

9/1986 - 6/1990 *BSc in Statistics*, Addis Ababa University (AAU), Ethiopia

9/1991 - 6/1994 *MSc in Statistics*, AAU, Ethiopia

9/1995 - 1/2000 *Attended extension diploma courses in Theology*, Mekane Yesus Theological Seminary, Ethiopia

10/2000 - 7/2001 *MSc in Applied Statistics*, Limburgs Universitair Centrum (LUC), Belgium

10/2001 - 9/2002 *MSc in Biostatistics*, LUC, Belgium

10/2002 - 7/2005 *PhD student of biostatistics*, University of Hannover, Germany

Work Experience

Lecturer at Department of Statistics, Addis Ababa University, Ethiopia (9/1990 - 9/2000).