ABSTRACT

Title of dissertation:	Statistical Methods in Bioequivalence Studies
	Meiyu Shen, Ph.D., 2015
Dissertation directed by:	Eric V. Slud, Ph.D., Department of Mathematics, University of Maryland Estelle Russek-Cohen, Ph.D., Food and Drug Administration

Bioequivalence studies are an essential part of the evaluation of generic drugs. The most common in-vivo bioequivalence (BE) study design is the two-period two-treatment crossover design. *AUC* (area under the concentration-time curve) and *Cmax* (maximum concentration) are obtained from the observed concentration-time profiles for each subject from each treatment under each sequence.

In the BE evaluation of pharmacokinetic crossover studies, the normality of the univariate response variable, e.g. log(*AUC*) or log(*Cmax*) is often assumed in the literature without much evidence. Therefore, we investigate the distributional assumption of the normality of response variables, log(*AUC*), log(*Cmax*), and log(*Tmax*) by simulating concentration-time profiles from the two-stage pharmacokinetic models for a wide range of pharmacokinetic parameters and measurement error structures. Our simulation shows that log(*AUC*) has heavy tails and log(*Cmax*) is skewed. We study the impact of the non-normality of response variable on the sample size and type I error rate.

Under the normality of the response variable, the most common approach to testing for bioequivalence is the two one-sided tests procedure. We develop the exact analytical formula for the probability of rejection in the two one-sided tests procedure

for crossover bioequivalence studies under general parameter settings. Our exact formulas for power and sample size are shown to improve in realistic parameter settings over the previous approximations.

We propose a new unblinded sample size re-estimation strategy. The new total sample size is calculated from our exact power function for the one stage using the estimated variance from the Stage 1 as the true variance. If the sample variance from Stage 1 is smaller than the initial variance from the historical data, then we stop at the end of Stage 1 and analyze Stage 1 data with the standard t- quantile. Otherwise, we collect data from additional subjects. We then analyze the combined data from both Stage 1 and Stage 2 with a new test statistic using the pooled variance of two stages. The exact critical values for the new test statistics are derived as the largest of u for which the following condition holds: the experimentwise type I error rate is exactly α.

STATISTICAL METHODS IN BIOEQUIVALENCE STUDIES

By

Meiyu Shen

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctoral in Philosophy
2015

Advisory Committee:
Eric V. Slud, Ph.D., Co-Advisor
Estelle Russek-Cohen, Ph.D. Co-Advisor
Paul Smith, Ph.D.
Xin He, Ph.D.
Frank Alt, Ph.D.

Dedication

I dedicate this to my husband, son, daughter, mother, and brother for their endless

support when I worked on my part-time graduate program. I also dedicate this to my

late father, in memoriam.

Acknowledgements

Table of Contents

List of Tables

List of Figures

Chapter 1 Overview

Following the 2001 Food and Drug Administration's Guidance for Industry [1]:

Statistical Approaches to Establishing Bioequivalence, the Center for Drug

Evaluation and Research (CDER) recommended that a standard in vivo

bioequivalence (BE) study design be based on the administration of either single or

multiple doses of the test (T) and the reference (R) products to healthy subjects on

separate occasions, with random assignment to the two possible sequences of drug

product administration. Hence the crossover design for in vivo BE study is the

primary design studied in regulatory trials. The reference product could be a marketed

innovator's product previously approved by Food and Drug Administration (FDA)

and the test product a potential generic substitute manufactured by a different

pharmaceutical company. The test and reference products could also be different

formulations, but manufactured by the same pharmaceutical company.

In a two-period two-treatment crossover design, a group of n_1 subjects (Sequence

1) receives the reference drug, and a profile of the drug concentration within blood

plasma over time is obtained for each subject. After a washout period to remove any

carryover effect, this group receives the test drug, and drug plasma concentration-

time profile for each subject is again obtained. A second group of n_2 subjects

(Sequence 2) receives the drugs in the reverse order. Therefore AUC (area under the

concentration-time curve), $Cmax$ (maximum concentration), and $Tmax$ (time to reach

the maximum concentration) are obtained from the correspondingly observed

concentration time profile for each subject under each treatment in each sequence. In

practice, AUC is estimated by a trapezoidal approximation method.

For the bioequivalence evaluation of pharmacokinetic crossover studies, univariate response variables such as log(AUC) and log($Cmax$) are often assumed to follow a normal distribution in literature and in practice [2, 3]. The investigation by Lacey et al. [4] showed the underlying distributions of AUC and $Cmax$ of four different Glaxo Wellcome compounds with the number of subjects varying from 29 to 69 are better approximated by log-normality rather than normality using p-values of the Shapiro-Wilk test [5]. Liu and Weng [6] briefly discussed the theoretical distributions of AUC and $Cmax$ under joint multivariate normal and multivariate log-normal assumptions of the observed plasma concentrations. The FDA Guidance published in 1992: Statistical procedures for bioequivalence studies using a standard two-treatment crossover design [7] and others [8, 9] provided the rationale for normality of log(AUC) and log($Cmax$) as follows. Assuming that elimination of the drug follows first order and only occurs from the central compartment, the following equation holds after an extravascular route of administration: $AUC_{0-\infty} = \dfrac{FD}{CL} = \dfrac{FD}{V_a Ke}$, where F is the fraction absorbed, D is the administered dose, FD is the amount of drug absorbed, and CL is the clearance of a given subject which is the product of the apparent volume of distribution (V_a) and the elimination rate constant (Ke). The use of AUC as a measure of the amount of drug absorbed thus involves a multiplicative term (CL) which might be regarded as a function of the subject. For this reason, Westlake [8, 9] contends that the subject effect is not additive if the data is analyzed on the original scale of measurement.

Logarithmic transformation of the *AUC* data brings the *CL* ($V_a K_e$) term into the equation in an additive fashion:

$$\ln AUC_{0-\infty} = \ln F + \ln D - \ln V_a - \ln Ke .$$

Clearly, the above argument is vague and lacks concrete scientific evidence. Thus, there is a need to have a systematic investigation of this assumption by simulations of individual pharmacokinetic profiles.

In Chapter 2, we investigate the assumption of normality of the response variables, log(*AUC*), log(*Cmax*), and log(*Tmax*) by simulating a large number of concentration time profiles from two-stage one–compartment pharmacokinetic models for a wide range of parameter choices and measurement error structures. Then the distributions of log(*AUC*), log(*Cmax*), and log(*Tmax*) can be investigated by examining the departure of the histograms of the standardized response variables from normality for large sample studies (e.g., 100,000 subjects) and examining the rejection rates of the Shapiro-Wilk normality test for the response variables at a 0.05 significance level for small sample size (e.g., 40 subjects) studies.

In Chapter 3, we present the linear mixed effect model for a two-period two-treatment cross-over bioequivalence study. Assuming the response variable (log(*AUC*) or log(*Cmax*)) is normally distributed, we give a brief introduction of test statistics for the two one-sided hypothesis tests most commonly used in bioequivalence studies and present the power function for the test based on two one-sided test statistics. We also take a quick look at the power approach. In practice, the power approach usually consists of testing the hypothesis of no difference at level

3

α (e.g., 0.05) and a lack of significance is often used to incorrectly infer equivalence. We derive the power function for the power approach. The two one-sided tests procedure is selected because it is recommended by FDA guidance and the two one-sided tests procedure is superior to the power approach as a test of the hypothesis H_0 below based on Shuirmann's comparison [10] of the two one-sided tests procedure, and the power approach for assessing the equivalence of average bioavailability in terms of rejection regions under the assumption that the response variable (*log(AUC)* or *log(Cmax)*) is normally distributed.

Denote the population mean bioavailability (*log(AUC)* or *log(Cmax)*) of the test product by μ_T and the population mean bioavailability of the reference product by μ_R. In order to conclude the bioequivalence of the test product and the reference product, we aim to reject the null hypothesis of the following null and alternative hypotheses [10]:

$$H_0 : \mu_T - \mu_R \le \theta_1 \text{ or } \mu_T - \mu_R \ge \theta_2$$
$$H_1 : \theta_1 < \mu_T - \mu_R < \theta_2 \qquad .$$

Here θ_1 and θ_2 are pre-specified constants, also called equivalence margins, and $\theta_1 < \theta_2$.

The null hypothesis, H_0, states that μ_T and μ_R are not equivalent. The alternative hypothesis, H_a, states that they are equivalent. H_a is the intersection of the two one-sided parameter regions, $\{\theta_1 < \mu_T - \mu_R\}$ and $\{\mu_T - \mu_R < \theta_2\}$. The statistical hypotheses H_0 and H_1 given above are referred to as the "interval hypotheses" in literature [11, 12].

In Chapter 4, we present the background, notation and standard assumptions in two-period two-sequence crossover designs for tests of bioequivalence and briefly derive the joint density function of test statistics and the exact formula for the power of the two one-sided tests procedure for testing bioequivalence based on a univariate normally distributed response variable. Our derivation for the exact power under general parameters only serves for completeness and for expository purposes since the explicit power formula under the equal variance for the test and reference products and balanced design was published before. However, by allowing for unequal variances, it might allow the user to assess the robustness of the power and sample size determinations when this assumption is violated. Modest differences in variability may not cause concern when the bioequivalence question is being addressed. We compare the simulated values with the numerical values and indicate how numerical integration easily provides accurate numerical values for power and sample size. Then, we compare the numerical results of the exact method with other methods including some approximate ones that have been proposed because of ease of calculation.

When planning a bioequivalence study, one needs to specify a true mean difference between test and reference for the response variable of interest, along with the variance of these responses for both test and reference formulations. This may come from a limited pilot study or by looking at data from the reference formulation to make an intelligent guess. Because these initial values may be off, providing an approach that allows one to update information in the study while still conserving

type I error can improve the chances of success. We henceforth refer to our new method as a two stage design for bioequivalence.

In Chapter 5, we propose a new unblinded sample size re-estimation strategy which re-estimates the new total sample size from the exact power function in Chapter 4 for the single stage with the replacement of the true variance by the estimated variance from the Stage 1. If the sample variance from the first stage is smaller than the initial variance from the historical data, then we stop at the end of the first stage, and we analyze the first stage data with the standard t quantile. Otherwise, we proceed collecting the data from additional subjects and then analyze the combined data from both Stage 1 and Stage 2. In order to analyze the combined data from Stage 1 and Stage 2, we propose new test statistics using the pooled variance of two stages. We search for exact critical values for the new test statistics subject to the following constraint: the maximum of the probability of rejecting the non-equivalent under the non-equivalent for the two-stage study plus the probability of rejecting the non-equivalent under the non-equivalent for the one-stage study in the whole range of variance is less than or equal to the nominal level α. With this exact critical value, the experiment-wise type I error rate for the sample size re-estimation procedure is not inflated.

Throughout Chapter 3 to Chapter 5, we assume the normality of the response variable. In Chapter 6, we consider the impact of the non-normality of the response variable on the sample size, power, and type I error rate for the single stage study planning and also on the expected sample size and type I error rate for the two-stage

study planning. In Chapter 7, we provide our final conclusions and recommendations for designing a new study using either a one stage design or a two stage design.

Chapter 2 Distributional assumptions for AUC, Cmax and Tmax

2.1 Introduction

In a typical pharmacokinetic bioequivalence study with a single dose administration, one of the drug products is a reference formulation and the other a test formulation. Each subject is administered both formulations in a randomized two-period crossover design [3]. A concentration-time profile is determined for each subject given each formulation. Each single concentration-time profile can be modeled by a pharmacokinetic compartmental model [13]. Many software programs exist for estimating the pharmacokinetic parameters such as the absorption rate, the volume of distribution, etc. [14]. Then, *AUC*, *Cmax*, and *Tmax* can be obtained from the fitted pharmacokinetic model. In spite of these elaborate pharmacokinetic models, the *AUC*, *Cmax*, and *Tmax* are obtained from the nonparametric method [1] for bioequivalence assessment.

In practice, the univariate response variables such as log(*AUC*) and log(*Cmax*) are often assumed to follow a normal distribution without much experimental data support. For instance, an investigation of observed pharmacokinetic studies in [4] was based on numbers of subjects from 29 to 69 and so the power of the Shapiro-Wilk test to detect departures from either distribution (lognormal or normal) may have been limited.

In this chapter, we investigate the normality assumption of log(*AUC*) or log(*Cmax*) using pharmacokinetic compartmental models typically used to describe

8

concentration profiles over time. In particular, if data is generated using the simplest pharmacokinetic models (namely one and two compartment models), will it ultimately lead to deciding which distribution of log(AUC), log($Cmax$), or log($Tmax$) is most plausible? There are many software packages, e.g., NONMEM [14] and many programs, e.g., SAS/IML [15] available for pharmacokinetic simulations. The SAS/IML program [15] provides an opportunity to use the statistical capabilities of the SAS package; NONMEM is a specialized-pharmacokinetic modeling software package. We write our own SAS program to simulate the plasma concentration profiles for streamlining the derived response variables and incorporating the desired variance-covariance structures for errors and pharmacokinetic parameters.

In this chapter, we will investigate the distributional assumption of the normality of response variables (log(AUC), log($Cmax$), and log($Tmax$)) by simulating a large number of concentration-time profiles from two-stage pharmacokinetic models for a wide range of variability in pharmacokinetic parameters, a wide range of correlations of pharmacokinetic parameter vector, and a wide range of measurement error structures.

In Stage 1, we simulate the mean plasma concentration-time profile of each subject from the one-compartment pharmacokinetic model using the pharmacokinetic parameters (absorption rate, elimination rate, bioavailability, and volume of distribution, etc.) of a particular drug whose values follow the log-normal distribution from subject to subject. In Stage 2, the plasma concentration-time profile of each subject is the result of the mean plasma concentration-time profile of each subject multiplied by the log-normally distributed residual errors. Then the distributions of

9

log(AUC), log($Cmax$), and log($Tmax$) from a simulated study are examined in three ways.

First, we compare the estimated density curve of the standardized response variables (log(AUC), log($Cmax$), and log($Tmax$)) for 400,000 simulated subjects with the standard normal density curve. We compare the percentiles of the sample distribution to those of the standard normal distribution for one case.

Second, we examine the normality departures of the histogram of the standardized response variable for 100,000 simulated subjects for several combinations of variability in pharmacokinetic parameters, correlations of pharmacokinetic parameters' vector, and measurement error structures so that the most severe normality departure can be spotted. From practical experience, correlation coefficients among pharmacokinetic parameters ranging from 0.1 to 0.5 seem reasonable. The coefficient of variation (CV) for measurement errors of 0.2 is small compared to 0.4 (a highly variable drug has CV>0.3). Coefficient of –log(0.5) is high since measurement errors are most commonly assumed to be independent for each subject in the pharmacokinetic modeling (see Chapter 3 in NONMEM users guide [14]).

Third, we compare the rejection rates of the Shapiro-Wilk normality test of these variables at 0.05 significance level for small sample size (e.g., 40 subjects) studies for some choice of parameters. Through the examination of a large sample size study, we can obtain a sampling distribution that is very close to the true distribution for the response variable and determine how the distribution of the response variable departs from normality. We obtain the proportion of trials which reject the null hypothesis of

normality from the rejection rate of the Shapiro-Wilk normality test for the response variable based on the small sample size study.. If the rejection rate is much larger than the nominal significance level, then the sampling distribution departs from normality. However, we would not know the nature of the departure. In addition, the normality testing for a small sample (e.g., 40 subjects) has very low power to reject the null hypothesis of normality. Hence the rejection rate is not a sensitive method. Sensitivity analyses investigate how the sampling distribution of the standardized $\log(AUC)$ (or the standardized $\log(Cmax)$) for a large number of simulated subjects deviates from normality if residual error is distributed as t (a heavy tail distribution), the mixture of two normal variables (two subgroups responding differently), or if the concentration-time profiles follow a two-compartment pharmacokinetic model with normal residual errors.

In Section 2.2, the two-stage one-compartment pharmacokinetic models are described in detail. In Section 2.3, we present the simulation scheme for the concentration profiles from a one–compartment pharmacokinetic model with first order absorption and first order elimination for many subjects who receive a single 1 mg oral dose of Ropinirole for treatment of Parkinson's disease. The simulation is motivated by the real example and the estimated means for parameters from the reference [16] are used as their true means in the simulation. Subsequently, the sampling distributions of $\log(AUC)$, $\log(Cmax)$, and $\log(Tmax)$ are obtained. In Section 2.4, we review univariate normality tests extensively and recommend that the Shapiro-Wilk test be used for sample sizes less than 2000. In Section 2.5, we closely examine the distributions of $\log(AUC)$, $\log(Cmax)$, and $\log(Tmax)$ for 400,000

simulated subjects for one case. We also examine the departure from normality of the histogram of the standardized response variable for 100,000 subjects simulated from one of several combinations of the variances and correlations of pharmacokinetic parameters' vector and measurement error structures. In Section 2.6.1, sensitivity analyses illustrate how the distribution of the standardized log(AUC) (or the standardized log($Cmax$)) for a large number of simulated subjects deviates from normality if e_{ij} is distributed as t with 5 to 20 degrees of freedom. In Section 2.6.2, sensitivity analyses study the validity of the normality assumptions of log(AUC), log($Cmax$), and log($Tmax$) if there is a subgroup with a slower absorption process. In Section 2.6.3, we present sensitivity analyses of the effect of different pharmacokinetic compartment models on the validity of the normality assumptions of log(AUC), log($Cmax$), and log($Tmax$) from two-compartment model under different combinations of the variations of the pharmacokinetic parameters and the variations of the measurement errors. In Section 2.7, distributions of log(AUC), log($Cmax$), and log($Tmax$) are examined for a real case with 39 subjects.

2.2 Pharmacokinetic models and assumed distribution

Assume that a typical person takes one tablet with dose D orally and the plasma concentration-time curve obtained after oral administration of one tablet can be described by a one-compartment model with the first-order absorption and elimination.

Let X be the true amount of drug in the body at time t after oral administration of one tablet with dose D. Let X_a be the true amount of drug at the absorption site at time t after oral administration of one tablet with dose D. For a drug that enters a body by

an apparent first-order absorption process, is eliminated by a first-order process, and distributes in the body according to a one-compartment model, the change in the amount of drug adheres to the following differential equations [13]:

$$\frac{dX}{dt} = kaX_a - keX .$$ (2.1)

$$\frac{dX_a}{dt} = -kaX_a .$$ (2.2)

Here ka is the apparent first-order absorption rate constant and ke is the apparent first-order elimination rate constant for the drug. Note that when $t=0$, $X=0$ and $X_a=FD$. Now we turn to solving the differential equations (2.1) and (2.2), we obtain the relationship between the amount of drug and time:

$$X = \frac{FDka}{(ka - ke)}\left(e^{-ka \cdot t} - e^{-ke \cdot t}\right).$$

Assuming the apparent volume of a typical person is V_a and bioavailability fraction is F, we obtain the relationship between the true concentration of drug (μ_{C_t}) and time (t):

$$\mu_{C_t} = \frac{FDka}{V_a(ka - ke)}\left(e^{-ka \cdot t} - e^{-ke \cdot t}\right).$$

Let C_{ij} denote the jth measurement of plasma concentration for $j=1,2,\ldots, n_i$; also for the ith subject with $i=1,2,\ldots,m$, taken at time t_{ij} after dosing. Thus a total of $N=\sum_{i=1}^{m} n_i$ plasma concentrations are obtained. Suppose that the relationship between the mean of C_{ij} and t_{ij} for a given subject i is a nonlinear function $f(t_{ij}, \boldsymbol{\beta_i})$, where $\boldsymbol{\beta_i}$ is a ($p \times 1$) vector of pharmacokinetic parameters for the ith subject which can vary from subject to subject, and t_{ij} is a nonrandom design constant. We further assume

that the form of f is common to all subjects, while $\boldsymbol{\beta}_i$ differs for each subject i. This

may be written as $E(C_{ij} \mid \boldsymbol{\beta}_i) = \mu_{C_{ij}} = f(t_{ij}, \boldsymbol{\beta}_i)$. Note that $f(t_{ij}, \boldsymbol{\beta}_i)$ is often assumed to be

a nonlinear function of t_{ij}, and $\boldsymbol{\beta}_i$ is assumed to be in the form of a summation of

exponential functions. It is common to represent the body as a system of

compartments and to assume that the rates of transfer between compartments follow

first-order or linear kinetics when we characterize the concentration of a drug in the

human body [13]. For example, $f(t_{ij}, \boldsymbol{\beta}_i) = \dfrac{F_i D k a_i}{V_{a,i}(ka_i - ke_i)} \left(\exp\!\left(-ka_i t_{ij}\right) - \exp\!\left(-ke_i t_{ij}\right) \right)$,

which is derived from the one-compartment linear pharmacokinetic model for plasma

concentration after a single oral dose, D, where $\boldsymbol{\beta}_i = (ka_i, ke_i, F_i/(1-F_i), V_{a,i})'$,

$ka_i > ke_i, 0 \leq F_i \leq 1$.

Now we can define the following two-stage models:

Stage 1 (between subject variability)

Variation among subjects is accounted for through the subject-specific regression

parameters ($\boldsymbol{\beta}_i$). Parameters may differ due to unexplained variation from the natural

biological or physical variability among subjects or the run-to-run variation in assay

procedures.

In general, subjects in pharmacokinetic bioequivalence studies are chosen from a

relatively homogeneous population of healthy volunteers. Thus, variation among

pharmacokinetic parameters across subjects is often attributable mainly to random

variation among subjects rather than to differences in individual demographic and

physiological characteristics that would be more pronounced in a heterogeneous

patient population, e.g., body weight, genetics, and disease status.

In bioequivalence studies, it is appropriate to assume that inter-subject variation is due to unexplained noise:

$$log(\boldsymbol{\beta_i}) = log(\boldsymbol{\gamma}) + \mathbf{b_i}. \tag{2.3}$$

In Model (2.3), $log(\boldsymbol{\beta_i})$ is the vector of logarithms of the components of the vector $\boldsymbol{\beta_i}$, $\boldsymbol{\gamma}$ is a positive vector of population pharmacokinetic parameters, $log(\boldsymbol{\gamma})$ is the vector of logarithms of the components of the vector $\boldsymbol{\gamma}$, and the error vector $\mathbf{b_i}$ is the normal random component of inter-subject variation, which might be taken to have mean vector zero and covariance matrix $\boldsymbol{\Sigma}$. In practice, pharmacologists often assume $log(\boldsymbol{\beta_i})$ is distributed as a normal random variable [14, 16-18]. This assumption is based on physiological and biological reasons such as positively skewed $\boldsymbol{\beta_i}$ [17]. On the contrary, this assumption has not been validated. However, in one type of bimodal population, a small percentage of the population has a slower absorption process. This two-subpopulation case corresponds to a two-component mixture of the multivariate-normal $log(\boldsymbol{\beta_i})$ distribution where two components differ only by $log(ka)$.

Stage 2 (Within subject variability)

Assume that for Subject i, the jth concentration follows the model

$$y_{ij} = log(C_{ij}) = g(t_{ij}, \boldsymbol{\beta_i}) + e_{ij} . \tag{2.4}$$

Here e_{ij} is a normal random measurement error with $E(e_{ij} | \boldsymbol{\beta_i}) = 0$ and $Var(e_{ij} | \boldsymbol{\beta_i}) = \sigma^2$, given the ith subject, and $f(t_{ij}, \boldsymbol{\beta_i}) = \exp\left(g(t_{ij}, \boldsymbol{\beta_i}) + \sigma^2 / 2\right)$.

Let $\mathbf{y_i} = [y_{i1},...,y_{in_i}]'$ be the log-transformed concentrations of the ith subject and furthermore let $\mathbf{e_i} = [e_{i1},...,e_{in_i}]'$ be the errors of the ith subject. Let $\mathbf{g_i}(\mathbf{\beta_i})$ be the vector of concentration functions of the ith subject.

$$\mathbf{g_i}(\mathbf{\beta_i}) = \left[g(t_{i1},\mathbf{\beta_i}) \quad g(t_{i2},\mathbf{\beta_i}) \quad ... \quad g(t_{in_i},\mathbf{\beta_i}) \right]. \tag{2.5}$$

We can summarize the data for the ith subject as $\mathbf{y_i} = \mathbf{g_i}(\mathbf{\beta_i}) + \mathbf{e_i}$, where we assume $E(\mathbf{e_i}|\mathbf{\beta_i}) = 0$, $Var(\mathbf{e_i}|\mathbf{\beta_i}) = \mathbf{R_i}$, and $\mathbf{R_i}$ is the variance-covariance matrix of log-transformed data within the ith subject.

Let $e_{ij} = X_i(t_{ij})$, where $\{X_i(t), t \geq 0\}$ is all Ornstein-Uhlenbeck process [19] defined by the following stochastic differential equation:

$dX_i(t) = -\xi X_i(t)dt + \sigma dW(t)$, $X_i(0) \sim N(0,\sigma^2)$, and $t \geq 0$. Here $\xi > 0$ and $\sigma > 0$ are unknown parameters, and $W(t)$ is the standard unit Wiener process. The solution to the preceding differential equation is $X_i(t) = \exp(-\xi t)X_i(0) + \sigma \int_0^t \exp(-\xi(t-s))dW_s$.

Thus we have $e_{i1} = u_{i1}$, $e_{i,j+1} = \exp(-\xi(t_{j+1} - t_j))e_{ij} + u_{i,j+1}$, $j = 1,2,..,m$, where $u_{i1},...,u_{im}$ are independent distributed normal variables with $E(u_{ij}|\mathbf{\beta_i}) = 0$, $Var(u_{i1}|\mathbf{\beta_i}) = \sigma^2$, and $Var(u_{ij}|\mathbf{\beta_i}) = \sigma^2(1 - \exp(-2(t_{j+1} - t_j)\xi))$. Hence $Cov(e_{ij}, e_{ik}|\mathbf{\beta_i}) = \sigma^2 \exp(-2(t_k - t_j)\xi), k > j$. If CV denotes the coefficient of variation of the untransformed concentration data, then the variance of the log-transformed concentration is computed as $\sigma^2 = \log(1+CV^2)$.

We assume that $\mathbf{b_i}$ is independent of e_{ij}. A supposition of normality of $\mathbf{b_i}$ and e_{ij} in Models (2.3) to (2.5) cannot be justified on a physiological or pharmacological basis,

and hence some sensitivity analyses of bioavailability parameters ($\log(AUC)$ and $\log(Cmax)$) to these distributions are essential.

2.3 Simulation scheme

Motivated by a real world example, we will simulate the concentration profiles from Models (2.3) to (2.5). Assuming the estimate means are the true value, this is a one–compartment pharmacokinetic model with the first order absorption and first order elimination for many subjects to whom are administered a single 1 mg oral dose of Ropinirole for treatment of Parkinson's disease. This drug is a novel non-ergoline dopamine D2 receptor agonist, for which Kaye and Nicholls [20] summarized clinical pharmacokinetics. We obtain the estimated means of untransformed pharmacokinetic parameters (ka, ke, F, and V_a) from [20]. Here, ka is the absorption rate in hr^{-1}; ke is the elimination rate in hr^{-1}; F is the bioavailability fraction, $0 \leq F \leq 1$; and V_a is the apparent volume in liters (L). In the reference [20], F is reported to be approximately 0.5, V_a at steady state is approximately 7.2 L/kg after oral administration, *Tmax* approximately lies in the range from 0.5 to 4 hours after dosing, and the elimination half-life is approximately 6 hours after dosing. The average of ke is about 0.12 *hr^{-1}* obtained by $ke = \dfrac{0.693}{t_{1/2e}}$ (the well-known approximate relationship [13] between elimination half-life ($t_{1/2e}$) and ke). The average of ka is 1.5 *hr^{-1}* obtained by $T\max = \dfrac{2.3026}{ka - ke} \ln\left(\dfrac{ka}{ke}\right)$ (the approximate relationship [13], among *Tmax*, absorption rate, ka, and ke) when we assume *Tmax* to be 4.21 hours. According to Kaye and Nicholls [20], patients with Parkinson's disease usually have body weights

in the range from 65 to 75 kg. The average weight of a patient is assumed 70 Kg, and so the average of V_a is 525 Liters.

The following detailed steps delineate how to simulate the plasma concentration profiles:

1) The coefficients of variation for person-level untransformed pharmacokinetic parameters are assumed after consideration of the estimated value in [20] as:

$$\boldsymbol{\eta} = E\left[ka, \quad ke, \quad \frac{F}{1-F}, \quad V_a \right]' = [1.5, \quad 0.12, \quad 1, \quad 525]'$$

and $\mathbf{cv} = \left[cv\_ka, \quad cv\_ke, \quad cv\_\frac{F}{1-F}, \quad cv\_V_a \right]'.$

2) Assume that the log-transformed vector of pharmacokinetic parameters for the ith subject follow a multivariate normal distribution and write this as log of vector

entries: $\log\left[ka_i, \quad ke_i, \quad \left(\frac{F}{1-F}\right)_i, \quad V_{a,i} \right]' \sim N(\boldsymbol{\lambda}, \boldsymbol{\Lambda})$. For a log-normal ($\lambda, \sigma^2$) variable,

the coefficient of variation squared cv_{pk}^2 (variance divided by the square of mean) is

$e^{\sigma^2} - 1$ and $\sigma = \sqrt{1 + cv_{pk}^2}$. The correlation matrix of these log-transformed

pharmacokinetic parameters is assumed to be $\mathbf{R}$. We further assume each parameter on this scale is equally correlated. Here we need to convert the marginal mean ($\eta[j], \forall j \in \{1,2,3,4\}$) and coefficient of variation ($cv_{pk}[j], \forall j \in \{1,2,3,4\}$) for each untransformed pharmacokinetic parameter obtained from the reference into the marginal mean ($\lambda[j], \forall j \in \{1,2,3,4\}$) of each log-transformed pharmacokinetic parameter and variance matrix ($\boldsymbol{\Lambda}$) of log-transformed pharmacokinetic parameters by the following formulas.

$\lambda[j] = \log(\eta[j]) - 0.5 * \log(1 + (cv_{pk}[j])^2), \forall j \in \{1,2,3,4\}$, and

$$\mathbf{M} = (diag\{\sqrt{\log(1 + (cv_{pk}[j]^2))}\})_{4*4}.$$

Let $\mathbf{R} = \begin{pmatrix} 1 & \rho_0 & \rho_0 & \rho_0 \\ \rho_0 & 1 & \rho_0 & \rho_0 \\ \rho_0 & \rho_0 & 1 & \rho_0 \\ \rho_0 & \rho_0 & \rho_0 & 1 \end{pmatrix}$. Therefore, $\mathbf{\Lambda} = \mathbf{M} \cdot \mathbf{R} \cdot \mathbf{M}$.

3) Independently generate m subjects' random vectors of log-transformed

pharmacokinetic parameters, $\log(ka_i, ke_i, F_i/(1-F_i), V_{a,i})'$, from the distribution in

Step 2. Then convert this to $(ka_i, ke_i, F_i, V_{a,i})'$, for i=1, 2,..., m.

4) For a given individual i, simulate concentration profile at time points

$\mathbf{t} = (1/4, 1/2, 1, 1.5, 2, 3, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 28, 32, 36)'$, measured in hours.

Let n_i be the number of sampling times. The choice of the sampling schedule is

typical for these types of studies and follows the general rule: more frequent

samplings earlier on after dosing (e.g., sampling every 15 minutes for the first few

samples and sampling every half hour for the next few samples) and less frequent

samplings later on (e.g., sampling every 2 or 3 hours after half-life). This flexible

sampling schedule allows more information for the rapidly changing period prior to

the half-life.

Let t_{ij} be the jth sampling time point after dosing for Subject i and let C_{ij} be the

concentration at t_{ij}, j=1,2,...,n_i. Here $\mu_{C_{ij}} = \dfrac{F_i Dka_i}{V_{a,i}(ka_i - ke_i)}\left(e^{-ka_i \cdot t_{ij}} - e^{-ke_i \cdot t_{ij}}\right)$.

$\mu_{C_{ij}} = e^{g(t_{ij}, \mathbf{\beta_i}) + \sigma_{ij}^2/2}$, $\log(C_{ij}) \sim N(g(t_{ij}, \mathbf{\beta_i}), \sigma_{ij}^2)$.

5) Obtain AUC_i by evaluating the following summation: $\sum_{j=1}^{n_i-1} 0.5\left(C_{ij} + C_{i,j+1}\right)\left(t_{i,j+1} - t_{ij}\right)$

for Subject i, since analytical integration from this complicated nonlinear and stochastic model is not tractable, this calculation is done in practice.

6) Obtain $Cmax_i = \underset{j}{Max}\left(C_{ij}\right)$ from all observed values for Subject i.

7) Obtain $Tmax_i$ (time to reach $Cmax_i$) from all observed values for Subject i.

8) Perform the Shapiro-Wilk W test (discussed in Section 2.4) for the goodness of fit of normal distribution of $\log(AUC_i)$, $\log(Cmax_i)$, and $\log(Tmax_i)$, $i=1,2,\ldots,m$. Small sample sizes such as $m=40$ will be investigated and a large sample size of 400,000 will also be investigated for the true distribution arising from random-effects pharmacokinetic models.

9) Repeat Steps 2 to 8 for $Snum=10,000$ times when $m=40$.

10) Calculate the rejection rate of the goodness of fit test of log-normality at nominal significance level 0.05 by $\dfrac{\sum_{l=1}^{Snum} I\left(pvalue_l \leq 0.05\right)}{Snum}$. The 0.05 significance level is chosen based on the usual type I error rate of 5% used by the regulatory and industrial statisticians for each small bioequivalence study.

2.4 Univariate normality test

The goodness-of-fit problem is to decide whether or not the random sample (x_i, $i=1,..,n$) follows one of a parameterized family $F(x,\boldsymbol{\theta})$ of distributions, where $\boldsymbol{\theta}$ may be either be specified in advance or estimated. The W statistic proposed by Shapiro and Wilk [5] provides an omnibus measure of non-normality when $n<2,000$ and the

power of concluding the non-normality which slightly deviates from normality is very

high when $n>2,000$ [21]. The test statistic W [5] was defined for a random sample

$x_1,...,x_n$ as: $W = \left(\sum_{i=1}^{n} a_i x_{(i)}\right)^2 \bigg/ \sum_{i=1}^{n}(x_i - \bar{x})^2$, where $x_{(i)}$ is the ith order statistic (the

i^{th} value from the bottom of the list of sample variables in increasing sorted order),

$\bar{x}$ is the sample mean, and a_i is a constant determined from the expectation and

variance of the ith order statistic of a sample of n standard normal observations,

$i=1,2,…, n$. W must be greater than zero and less than or equal to one. Small values of

W lead to the rejection of the null hypothesis of normality. The distribution of W is

highly skewed. When the sample size is greater than three, the coefficients to

compute the linear combination of the order statistics in Proc Univariate, a SAS

procedure [22], are approximated with an approximate normalizing transformation

suitable for computer implementation [23]. According to this method, variable Z_n

defined by $Z_n = \left(-\log(\gamma_0 - \log(1-W_n)) - \mu_0\right)/\sigma_0$ for $4 \leq n \leq 11$ or

$Z_n = \left(\log(1-W_n) - \mu_0\right)/\sigma_0$ for $12 \leq n \leq 2000$ are treated as precisely normally

distributed, where $\gamma_0, \mu_0,$ and σ_0 are functions of n from simulation results, and W_n is

W-statistic value for sample size n. Royston [23] extended the Shapiro-Wilk W test up

to sample size 2000. Large values of Z_n indicate departure from normality. When the

sample size is larger than 2000, other goodness of fit tests such as the chi-squared test

or tests based on the empirical distribution function such as the original or modified

Kolmogorov-Smirnov test, Cramer-von Mises or Anderson-Darling tests are used.

   The main purpose of the chi-squared test is to provide a quantitative test of the

discrepancies between the observed and the expected frequencies. The null

hypothesis is: the observations are randomly drawn from a specified theoretical distribution. The chi-squared test can be used with either continuous or discrete distribution. Under the null, the chi-squared test follows the chi-squared distribution in large samples with the degrees of freedom equal to the number of cells minus 1 minus the dimension of the estimated parameter $\boldsymbol{\theta}$ [24, 25].

Another class of goodness-of-fit statistics is empirical distribution function (EDF) statistics because they are based on a comparison of $F(x)$ with the empirical distribution function $F_n(x)$. Kolmogorov-Smirnov, Cramer-Von Mises, and Anderson-Darling test are examples of EDF statistics tests.

If $F(x)$ is continuous and completely specified, EDF statistics is more powerful than the chi-squared test. The Kolmogorov-Smirnov statistic is the largest vertical distance between the fully specified cumulative distribution function ($F(x)$) and the EDF ($F_n(x)$), which is a step function that takes a step of height 1/n at each observation. This test only applies to continuous distributions and tends to be more sensitive near the center of the distribution than at the tails. Anderson-Darling test [26] uses the quadratic class EDF, which is based on the weighted and integrated squared difference ($F_n(x) - F(x))^2$. This test gives more weight to the tails than the Kolmogorov-Smirnov test does by setting a function weighting the square difference. Similarly, Cramer-von Mises test [27] is also based on the quadratic class of EDF statistics, but the weight function is set to 1. The Cramer-von Mises test and the Anderson-Darling test better takes into account the variation in the whole data; while the Kolmogorov-Smirnov test is more sensitive to the outliers in the sample.

If $F(x)$ is not completely specified, the mean and/or the variance have to be estimated from the data. In such cases, the critical threshold for any of the three EDF statistics is no longer valid. However, some work [28-30] has made it possible to use EDF statistics for two very important practical situations in which the distribution tested is normal or exponential, with parameters to be estimated. For a large sample size (n>2000), the Kolmogorov-Smirnov test, Cramer-von Mises test, and Anderson-Darling test are usable for normal and exponential distributions when the mean and variance are estimated from the data after the percent points are obtained from simulations [30]. Once the EDF test statistics (Kolmogorov-Smirnov test, Cramer-von Mises test and Anderson-Darling test) are computed, then the associated p-values are calculated. SAS procedure: PROC UNIVARIATE [22] uses internal tables of probability levels that are similar to those given by D'Agostino and Stephens [31].

For continuous, asymmetric and long-tailed distributions (e.g., $\chi^2(1)$, $\chi^2(2)$, $\chi^2(4)$, and $\chi^2(10)$) and continuous, asymmetric, and short-tailed distributions (e.g., $Beta(2,1)$ and $Beta(3,2)$), the W statistic is most sensitive and the Kolmogorov-Smirnov test and Cramer-von Mises test show, in general, relatively poor sensitivity for these alternative [21]. For continuous, symmetric, and long-tailed distributions (e.g., $Cauch(1)$, $Cauch(2)$, and $Cauch(4)$) and continuous, symmetric, and short-tailed distributions (e.g., $Beta(1,1)$ and $Beta(2,2)$), the W statistic outperforms the Kolmogorov-Smirnov test and Cramer-von Mises test [27]. Hence the Anderson-Darling test implemented by SAS PROC UNIVARIATE is chosen in this chapter for assessing the log-normality of the response variable when n>2000 and SAS PROC

UNIVARIATE's *W* test extended by Royston is chosen for assessing the log-normality of the response variable when *n*<2000.

2.5 Distributions of log(AUC), log(Cmax), and log(Tmax)

2.5.1 One case study

To illustrate the simulated plasma concentration-profiles and distributions of the response variables, we will simulate plasma concentration-time profiles using the simulation scheme in Section 2.3 for one particular set of parameters, that is , a one-compartment model with *CV*=0.2, ξ =-log(0.5), $\rho_0 = 0.5, cv_{pk} = (0.3, 0.3, 0.3, 0.3)'$, and

$$\left(ka_i, ke_i, F_i/(1-F_i), V_{a,i}\right)' = (1.5, 0.12, 1, 525)'.$$ Figure 2.1 shows concentration-time profiles for a sample of 20 subjects simulated under the models (2.3)-(2.5).

Throughout this chapter, the log(*AUC*), log(*Cmax*), and log(*Tmax*) are each standardized through centering by its sample mean and scaling by its sample standard deviation. The density curve of the standardized log(*AUC*) or log(*Cmax*) is estimated by the normal kernel density estimation method with bandwidth which is 0.9 times the minimum of the standard deviation and the interquartile range divided by 1.34 times the sample size to the negative one-fifth power [32].

Figure 2.2 shows that the estimated density curve of the standardized log(*AUC*) for 400,000 simulated subjects is very close to the standard normal density. According to Table 2.1, the estimated quantiles for the standardized log(*AUC*) are generally not far away from the standard normal percentiles, but for example the 1% and 2% quantiles for standardized log(*AUC*) differ noticeably from the normal percentiles, showing that the tails of log(*AUC*) are a bit asymmetric even when the errors and random

effects in the pharmacokinetic model are normally distributed. The differences are as high as 0.2 in the tails, becoming closer (within about 0.05) in the center of the distribution.

Figure 2.3 shows that the estimated density curve of the standardized log(*Cmax*) is skewed to the left compared to the standard normal density. Table 2.1 shows that the estimated quantiles for the standardized log(*Cmax*) are consistently larger than those for the standard normal. The differences are as high as 0.3 in the tail.

Figure 2.4 shows that the cumulative distribution curve of the standardized log(*Tmax*) with a fixed number of discrete points significantly deviates from the standard normal cumulative distribution curve. Table 2.1 shows that the 1% tail quantile of standardized log(*Tmax*) is 17% larger than the 1% standard normal quantile; and the 99% tail quantile of standardized log(T*max*) is 12% larger than the 99% standard normal quantile. In practice, *Tmax* is assessed subjectively and is not assessed by equivalence criteria.

Figure 2.1 An example of concentration-time profiles from 20 subjects simulated from one-compartment model with CV=0.2, ξ =-log(0.5), $\rho_0 = 0.5, cv_{pk} = (0.3, 0.3, 0.3, 0.3)'$, and

$$\left(k_{a,i}, k_{e,i}, F_i/(1-F_i), V_{a,i}\right)' = (1.5, 0.12, 1, 525)'.$$
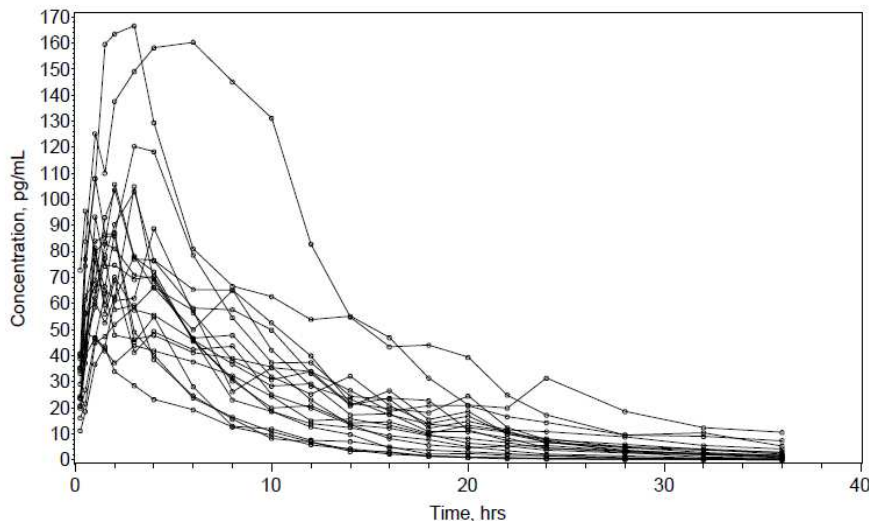
Figure 2.2 Comparison of the standard normal density and the estimated density curve of log(*AUC*) from 400,000 subjects simulated from one-compartment model with CV=0.2, ξ =-log(0.5), $\rho_0 = 0.5, cv_{pk} = (0.3,0.3,0.3,0.3)'$ and $\left(k_{a,i}, k_{e,i}, F_i/(1-F_i), V_{a,i}\right)' = (1.5,0.12,1,525)'$.
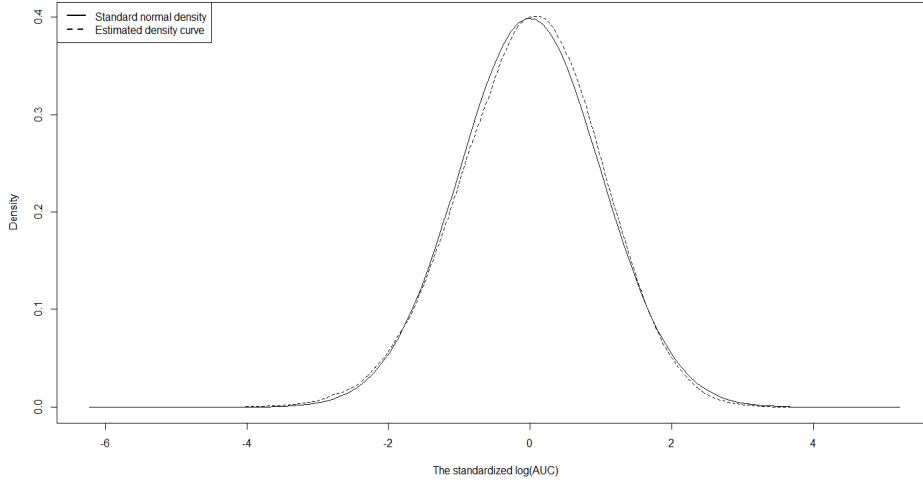
Figure 2.3 Comparison of the standard normal density and the estimated density curve of log(*Cmax*) from 400,000 subjects simulated from one-compartment model with CV=0.2, ξ =-log(0.5), $\rho_0 = 0.5, cv_{pk} = (0.3,0.3,0.3,0.3)'$ and $\left(k_{a,i}, k_{e,i}, F_i/(1-F_i), V_{a,i}\right)' = (1.5,0.12,1,525)'$.
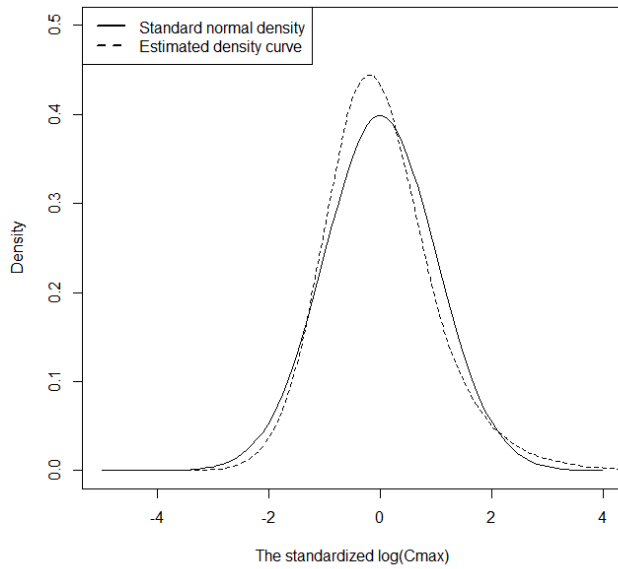
Figure 2.4 Comparison of standard normal distribution function and cumulative distribution function of log($Tmax$) from 400,000 subjects simulated from a one-compartment model with CV=0.2, ξ=-log(0.5), $\rho_0 = 0.5, cv_{pk} = (0.3, 0.3, 0.3, 0.3)'$ and

$$\left(k_{a,i}, k_{e,i}, F_i/(1-F_i), V_{a,i}\right)' = (1.5, 0.12, 1, 525)'$$
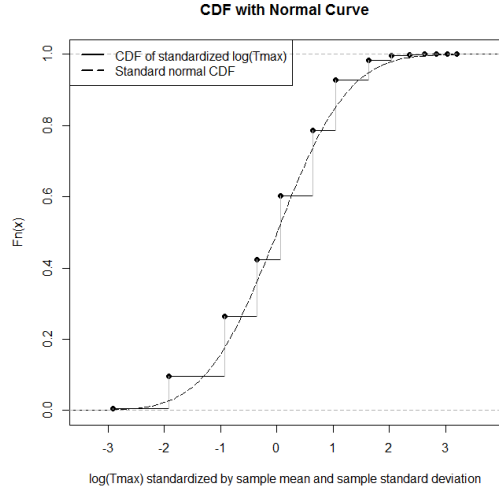
Table 2.1 Comparison of observed extreme quantiles for standardized log(AUC), log($Cmax$), and log($Tmax$) with standard normal extreme quantiles for 400,000 subjects simulated from one-compartment model with CV=0.2 , ξ=-log(0.5), $\rho_0 = 0.5, cv_{pk} = (0.3, 0.3, 0.3, 0.3)'$, and

$$\left(k_{a,i}, k_{e,i}, F_i/(1-F_i), V_{a,i}\right)' = (1.5, 0.12, 1, 525)'$$

Probability	Estimated quantile			Standard normal quantile
	Standardized log(AUC)	Standardized log($Cmax$)	Standardized log($Tmax$)	
0.10%	-3.338	-2.694	-2.910	-3.090
0.50%	-2.749	-2.258	-2.910	-2.576
1%	-2.451	-2.043	-1.920	-2.326
2%	-2.142	-1.816	-1.920	-2.054
5%	-1.692	-1.475	-1.920	-1.645
95%	1.595	1.771	1.628	1.645
97.50%	1.882	2.264	1.628	1.960
99%	2.204	2.908	2.039	2.326
99.50%	2.423	3.359	2.039	2.576
99.90%	2.883	4.335	2.357	3.090

2.5.2 Examination of distributions of response variables for 8 scenarios

We compare the distributions of the standardized log(AUC), log($Cmax$), and

log($Tmax$) for 100,000 simulated subjects to the standard normal density. We also

obtain the rejection rate of the normality testing for 2500 trials, each of which has 40 subjects. Note that testing normality for a small sample (e.g., 40 subjects) has very low power to reject the null hypothesis of normality.

From Table 2.2, it can be seen that 1) the median of the standardized log(AUC) is off 0 by less than 0.1 for all cases; 2) the 75th percentile is about 0.15 larger than the 75th standard normal quantile for Cases 1 to 4 and about 0.15 smaller for Cases 5 to 8, and 3) the 25th percentile is 0.07 smaller than the 25th standard normal quantile for Cases 1 to 4 and about 0.08 larger for Cases 5 to 8. Figure 2.5 shows that the histogram (300 breaks) of the standardized log(AUC) has heavy tails compared to the density of $N(0,1)$ for all 8 cases.

The rejection rates of the Shapiro-Wilk normality test of log(AUC) at 0.05 significance level are about 5% for Cases 3, 5, 6, 7, and 8. The rejection rates of the Shapiro-Wilk normality test of log(AUC) at 0.05 significance level are approximately 6.5% for Cases 1, 2, and 6.

From Table 2.2, it can be seen that 1) the mean of the standardized log($Cmax$) is less than 0 by 0.15 for Cases 1, 3, 5, and 7; 2) the 75th percentile is about 0.1 larger than the 75th standard normal quantile for Cases 1 and 3; the same holds for Cases 5 and 7, and about 0.3 smaller for Cases 2, 4, 6, and 8; and 3) the 25th percentile is 0.2 to 0.3 larger than the 25th standard normal quantile for Cases 1, 3, 5, and 7 and about 0.2 smaller for Cases 2, 4, 6, and 8. Figure 2.6 shows that the histogram (300 breaks) of the standardized log($Cmax$) is skewed to the right of $N(0,1)$ if $\xi = 0$ or to the left if $\xi = -\log(0.5)$ for 8 cases. All rejection rates of the Shapiro-Wilk normality test of

log(*Cmax*) at 0.05 significance level for Cases 1, 3, 5, 6, 7, and 8 are less than 7.5%

and for Cases 2 and 4 are greater than 23%.

Figure 2.7 shows that the CDF of the standardized log(*Tmax*) has the fixed number

of mass points and is above the standard normal CDF for all cases. The rejection rates

of the Shapiro-Wilk normality test of log(*Tmax*) at 0.05 significance level for almost

all cases are larger than 50%.

Table 2.2 Description of cases with simulated 100,000 subjects from Models 2.1 to 2.5 where PK
parameters with $\boldsymbol{\eta} = E\left[ka,\ \ ke,\ \ F/(1-F),\ \ V_a\right]' = \left[1.5,\ \ 0.12,\ \ 1,\ \ 525\right]'$, $\mathbf{cv} = cv_{pk}[1,1,1,1]$, and
$\mathbf{e_i} = [e_{i1},...,e_{in_i}]'$.

| Case | $\log\boldsymbol{\eta}$ $\sim N(\boldsymbol{\lambda},\boldsymbol{\Lambda})$ | | Residual error e_{ij} | | Quantiles for $N(0,1)$: $(25^{th}, 50^{th}, 75^{th})=(-0.674,0,0.674)$ | | | | | |
| | | | | | Quantiles of the standardized log(*AUC*) | | | Quantiles of the standardized log(*Cmax*) | | |
	ρ_0	cv_{pk}	ξ	CV	25th	50th	75th	25th	50th	75th
1	0.5	0.3	0	0.4	-0.749	0.022	0.780	-0.413	0.152	0.737
2	0.5	0.3	0.5	0.2	-0.699	0.097	0.856	-0.910	-0.244	0.486
3	0.1	0.3	0	0.4	-0.766	0.009	0.773	-0.490	0.148	0.804
4	0.1	0.3	0.5	0.2	-0.715	0.075	0.852	-0.973	-0.253	0.539
5	0.5	0.2	0	0.4	-0.590	-0.052	0.478	-0.337	0.136	0.625
6	0.5	0.2	0.5	0.2	-0.551	0.017	0.576	-0.865	-0.294	0.382
7	0.1	0.2	0	0.4	-0.597	-0.050	0.483	-0.378	0.134	0.673
8	0.1	0.2	0.5	0.2	-0.560	0.008	0.574	-0.894	-0.294	0.413

Figure 2.5 Histograms (breaks=80) of the standardized log(*AUC*) for cases in Table 2.2 compared to *N(*0,1)

Histogram with Normal Curve

Case 5

Case 5

$\rho_0 = 0.5$

cv_pk $= 0.2$

$\xi = 0$

CV $= 0.4$

N(0,1)

Density

The standardized log(AUC)

Histogram with Normal Curve

Case 6

Case 6

$\rho_0 = 0.5$

cv_pk $= 0.2$

$\xi = -\log(0.5)$

CV $= 0.2$

N(0,1)

Density

The standardized log(AUC)

Histogram with Normal Curve

Case 7

Case 7

$\rho_0 = 0.1$

cv_pk $= 0.2$

$\xi = 0$

CV $= 0.4$

N(0,1)

Density

The standardized log(AUC)

Histogram with Normal Curve

Case 8

Case 8

$\rho_0 = 0.1$

cv_pk $= 0.2$

$\xi = -\log(0.5)$

CV $= 0.2$

N(0,1)

Density

The standardized log(AUC)

31
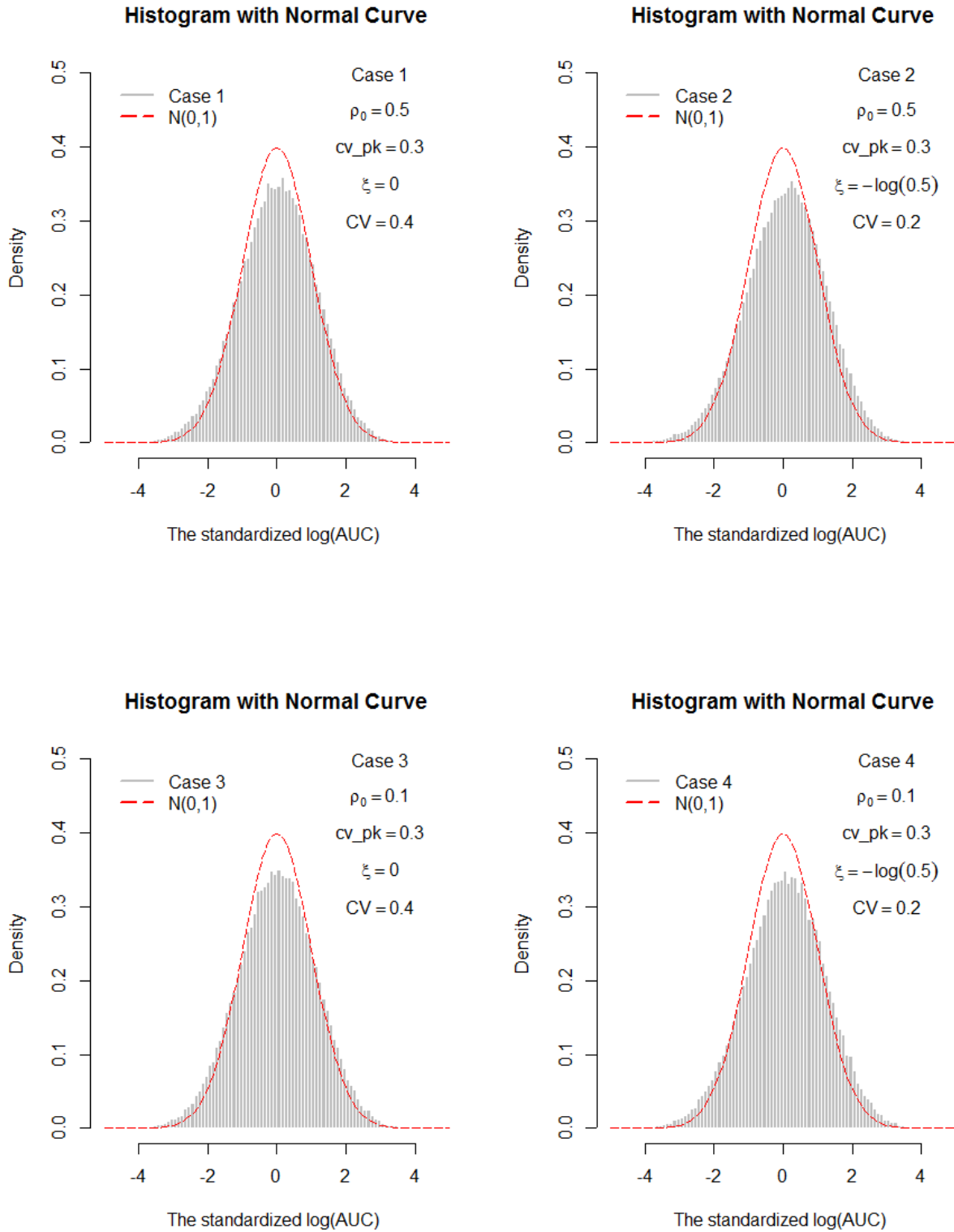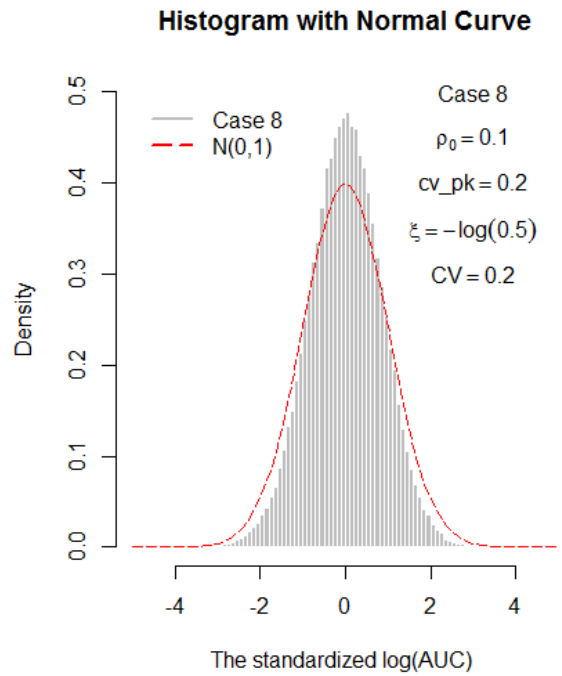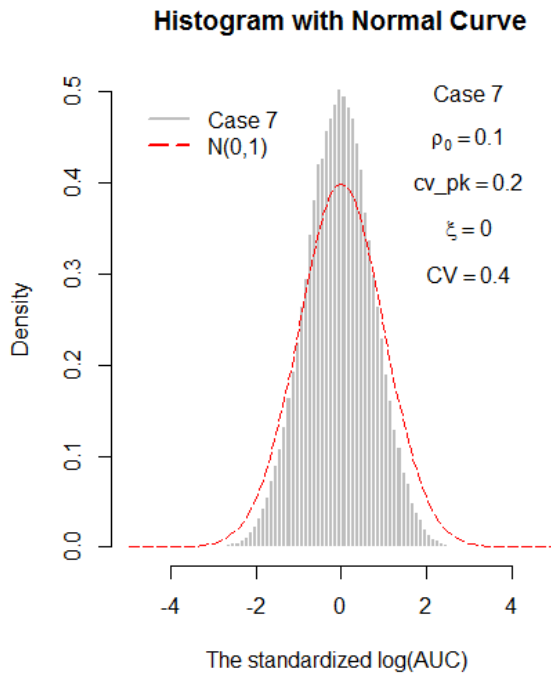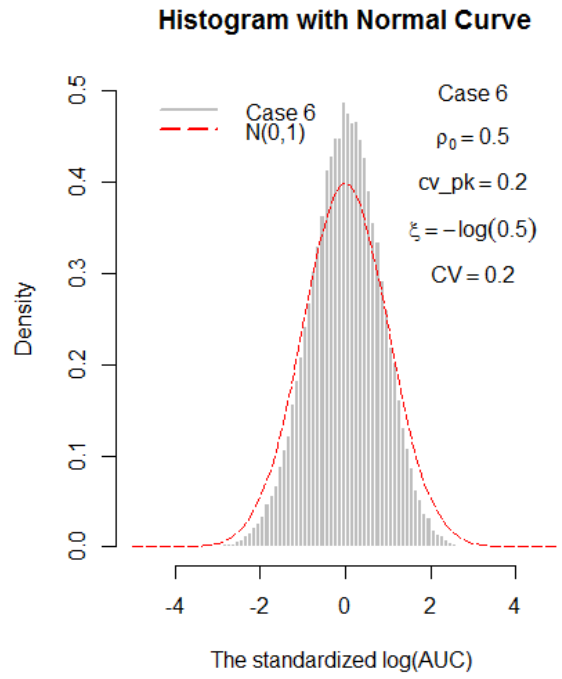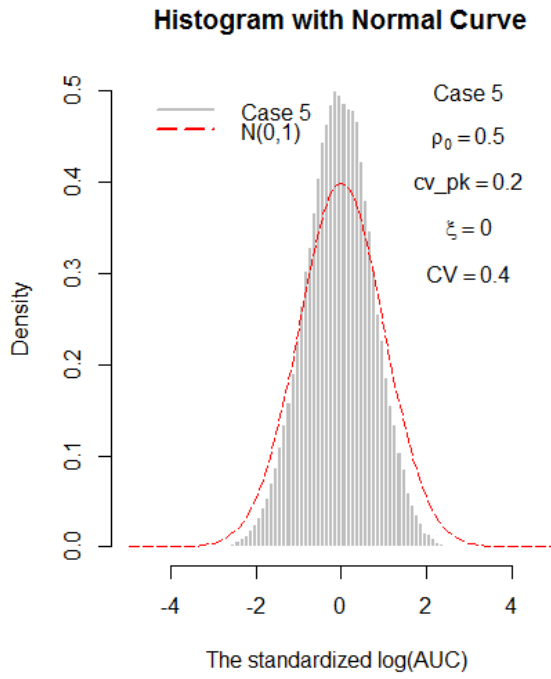
Figure 2.6 Histograms (breaks=80) of the standardized log(*Cmax*) for cases in Table 2.2 compared to *N(*0,1)

Histogram with Normal Curve

Case 5
$\rho_0 = 0.5$
cv_pk = 0.2
$\xi = 0$
CV = 0.4

— Case 5
-- N(0,1)

Density

The standardized log(Cmax)

Histogram with Normal Curve

Case 6
$\rho_0 = 0.5$
cv_pk = 0.2
$\xi = -\log(0.5)$
CV = 0.2

— Case 6
-- N(0,1)

Density

The standardized log(Cmax)

Histogram with Normal Curve

Case 7
$\rho_0 = 0.1$
cv_pk = 0.2
$\xi = 0$
CV = 0.4

— Case 7
-- N(0,1)

Density

The standardized log(Cmax)

Histogram with Normal Curve

Case 8
$\rho_0 = 0.1$
cv_pk = 0.2
$\xi = -\log(0.5)$
CV = 0.2

— Case 8
-- N(0,1)

Density

The standardized log(Cmax)
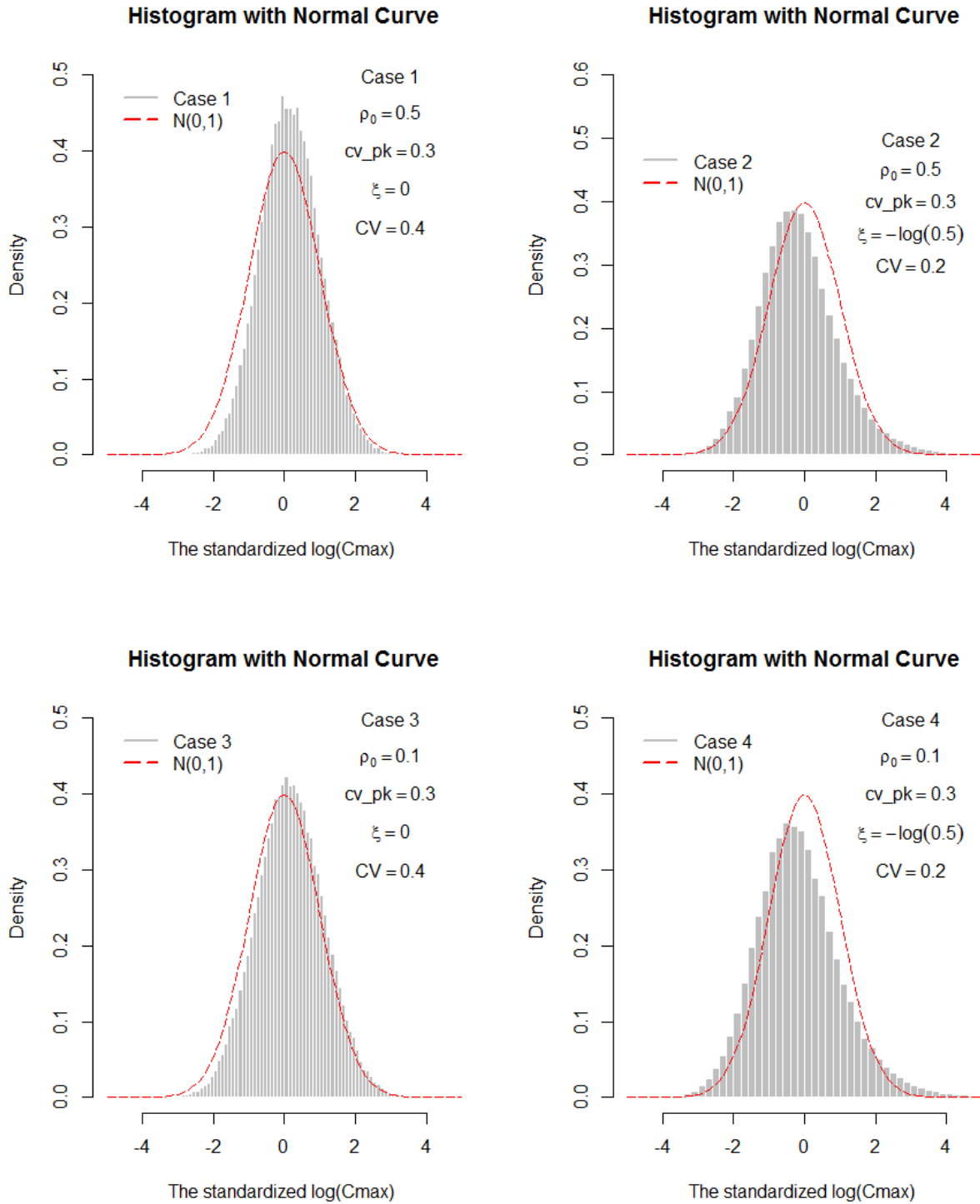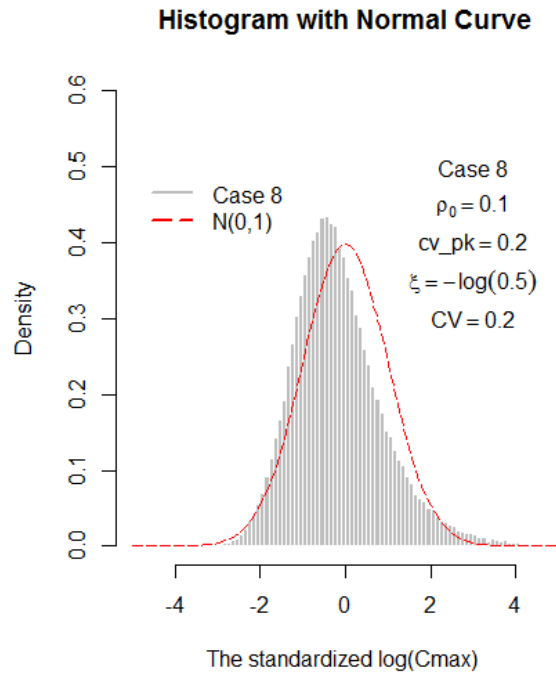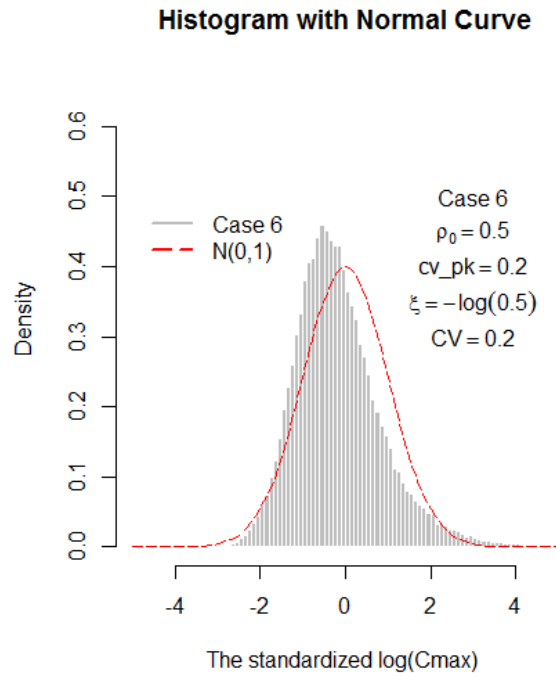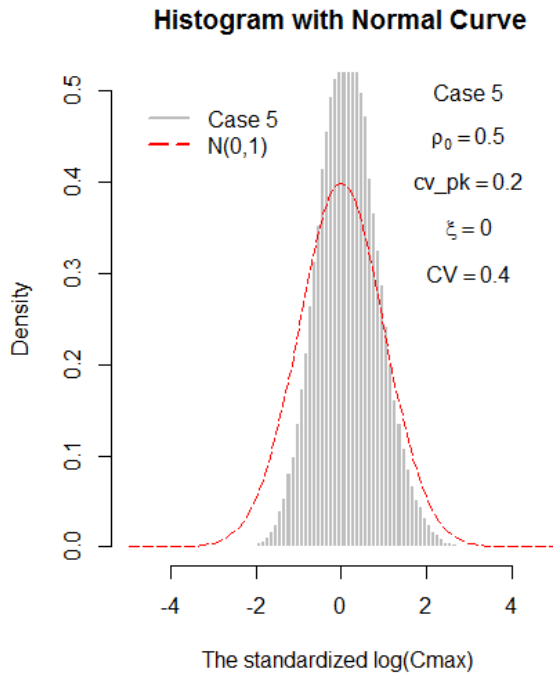
Figure 2.7 CDF of the standardized log(*Cmax*) for cases in Table 2.2 compared to CDF of *N(*0,1)

CDF with Normal CDF Curve

Case 1
— Case 1
-- N(0,1)

Case 1
$\rho_0 = 0.5$
$cv\_pk = 0.3$
$\xi = 0$
$CV = 0.4$

The standardized log(Tmax)

CDF with Normal CDF Curve

— Case 2
-- N(0,1)

Case 2
$\rho_0 = 0.5$
$cv\_pk = 0.3$
$\xi = -\log(0.5)$
$CV = 0.2$

The standardized log(Tmax)

CDF with Normal CDF Curve

— Case 3
-- N(0,1)

Case 3
$\rho_0 = 0.1$
$cv\_pk = 0.3$
$\xi = 0$
$CV = 0.4$

The standardized log(Tmax)

CDF with Normal CDF Curve

— Case 4
-- N(0,1)

Case 4
$\rho_0 = 0.1$
$cv\_pk = 0.3$
$\xi = -\log(0.5)$
$CV = 0.2$

The standardized log(Tmax)

CDF with Normal CDF Curve

— Case 5
-- N(0,1)

Case 5
$\rho_0 = 0.5$
$cv\_pk = 0.2$
$\xi = 0$
$CV = 0.4$

The standardized log(Tmax)

CDF with Normal CDF Curve

— Case 6
-- N(0,1)

Case 6
$\rho_0 = 0.5$
$cv\_pk = 0.2$
$\xi = -\log(0.5)$
$CV = 0.2$

The standardized log(Tmax)

CDF with Normal CDF Curve

— Case 7
-- N(0,1)

Case 7
$\rho_0 = 0.1$
$cv\_pk = 0.2$
$\xi = 0$
$CV = 0.4$

The standardized log(Tmax)

CDF with Normal CDF Curve

— Case 8
-- N(0,1)

Case 8
$\rho_0 = 0.1$
$cv\_pk = 0.2$
$\xi = -\log(0.5)$
$CV = 0.2$

The standardized log(Tmax)

34

2.6 Sensitivity analyses

In Section 2.5, the examination of distributions of 9 cases with large samples shows some deviations of log(AUC) from normality. The 25th and 75th percentiles of log(AUC) are slightly different from the 25th and 75th percentiles for the standard normal distribution, respectively. The 50th percentile of log(AUC) is close to 0. It also shows that log($Cmax$) distributes skewed to the right or the left. The 25th, 50th, and 75th percentiles of log($Cmax$) are all different from the 25th, 50th, and 75th percentiles for the standard normal distribution, respectively. But the rejection rate of the Shapiro-Wilk normality test of log($Cmax$) is slightly higher than 5% percent, but less than 7.5% for all cases. Clearly, the Shapiro-Wilk normality test of log(AUC) with small samples for all cases has very low power to reject the null hypothesis of normality.

Since simulations in Section 2.5 are based on one-compartment pharmacokinetic models with normal measurement errors, we will demonstrate whether or not log(AUC) and log($Cmax$) can be reasonably assumed as random normal variables by simulations if data come from the one-compartment pharmacokinetic models with symmetrical measurement errors distributed as t and with a bimodal population corresponding to one subgroup having a slower absorption process. We will also demonstrate whether or not log(AUC) and log($Cmax$) can be reasonably assumed to be normally distributed by simulations if data come from two-compartment pharmacokinetic models with normal measurement errors.

2.6.1 Symmetrical measurement errors distributed as *t*

We investigate how the distribution of the standardized log(AUC) (or the standardized log($Cmax$)) for large samples of simulated subjects deviates from normality if e_{ij} is distributed as a $t(v)$ with $v = $ 5, 10, 15, or 20. We compare the histogram of the empirical standardized log(AUC) (or the standardized log($Cmax$)) with the standard normal density curve. For the standardized log($Tmax$), we compare the cumulative distribution function (CDF) of the standardized log($Tmax$)) with the normal CDF.

For Cases 1 to 8 in Table 2.3, we assume that $e_{i1},...,e_{im}$ are independent and identically distributed *t*-variables for each *i* by letting $\xi = 0$, and we also vary the v values from 5 to 20 by 5 and CV values from 0.2 to 0.4 by 0.2 but fix the pharmacokinetic parameters' CVs and correlation at $cv_{pk} = (0.3,0.3,0.3,0.3)'$ and $\rho_0 = 0.5$, respectively. For Cases 9 to 12, $e_{i1},...,e_{im}$ are the correlated *t*-variables for each *i* by letting $\xi = -\log(0.5)$.

Figure 2.8 shows that the histogram of the standardized log(AUC) for 100,000 simulated subjects is very close to the standard normal density curve for all cases except Cases 1 and 9 (v=5 and CV=0.4) in which the histograms are slightly skewed to the right of the standard normal density curve. Table 2.3 shows that the 25th, 50th, and 75th percentiles of log(AUC) are all very close to the 25th, 50th, and 75th percentiles for the standard normal distribution, respectively. The rejection rate of the Shapiro-Wilk normality test at 0.05 significance level for log(AUC) is about 5% if $v \geq 10$ and greater than 5% if v=5.

Figure 2.9 shows that the histogram of the standardized log($Cmax$) is skewed to the right compared to the standard normal density when $\nu \geq 5$ and CV= 0.4, and the histogram of the standardized log($Cmax$) has a sharper peak and skew to the left compared to the standard normal density when $\nu \geq 5$ and CV= 0.2. The 25th, 50th, and 75th percentiles of log($Cmax$) are all different from the 25th, 50th, and 75th percentiles for the standard normal distribution, respectively. The rejection rate of the Shapiro-Wilk normality test at 0.05 significance level for log($Cmax$) is about 5% when $\nu \geq 10$ except Case 3 (8.1%) and above 10% when $\nu = 5$.

Figure 2.10 shows that the CDF of the standardized log($Tmax$) has the fixed number of mass points and is above the standard normal CDF for all cases.

In conclusion, the normality assumption of log(AUC) seems reasonable if the independent and identical measurement errors are assumed to be $t(10)$, moderately heavy tails. The normality assumption of log($Cmax$) is skewed even if the independent and identical measurement errors are assumed to follow t distribution. Normality of log($Tmax$) cannot be assumed in general.

Table 2.3 Description of cases with 100,000 simulated subjects from Models 2.1 to 2.5 where PK parameters with $\mathbf{\eta} = E\left[ka, \quad ke, \quad F/(1-F), \quad V_a\right]' = \left[1.5, \quad 0.12, \quad 1, \quad 525\right]'$, and $\mathbf{e_i} = \left[e_{i1},...,e_{in_i}\right]'$, with $cv_{pk} = (0.3, 0.3, 0.3, 0.3)'$ and $\rho_0 = 0.5$.

| Case | Error $e_{ij} \sim t(\nu)$ | | | Quantiles for $N(0,1)$: $(25^{th}, 50^{th}, 75^{th}) = (-0.674, 0, 0.674)$ | | | | | |
| | | | | Standardized log(AUC) | | | Standardized log($Cmax$) | | |
	ξ	ν	CV	25th	50th	75th	25th	50th	75th
1		5	0.4	-0.626	0.081	0.774	-0.366	0.353	1.160
2		5	0.2	-0.651	0.025	0.677	-0.768	-0.226	0.343
3		10	0.4	-0.669	0.020	0.694	-0.460	0.197	0.892
4	0	10	0.2	-0.668	-0.001	0.656	-0.815	-0.298	0.232
5		15	0.4	-0.675	0.013	0.681	-0.491	0.153	0.823
6		15	0.2	-0.672	0.001	0.653	-0.829	-0.317	0.211
7		20	0.4	-0.691	0.003	0.669	-0.506	0.133	0.793
8		20	0.2	-0.670	-0.005	0.648	-0.833	-0.323	0.191
9		5	0.4	-0.645	0.070	0.763	-0.514	0.123	0.837
10	-log(0.5)	5	0.2	-0.622	0.040	0.682	-0.698	-0.171	0.435
11		10	0.4	-0.669	0.015	0.676	-0.565	0.020	0.666
12		10	0.2	-0.630	0.018	0.648	-0.716	-0.212	0.352

Figure 2.8 Histograms of the standardized $\log(AUC)$ for cases in Table 2.3 compared to $N(0,1)$
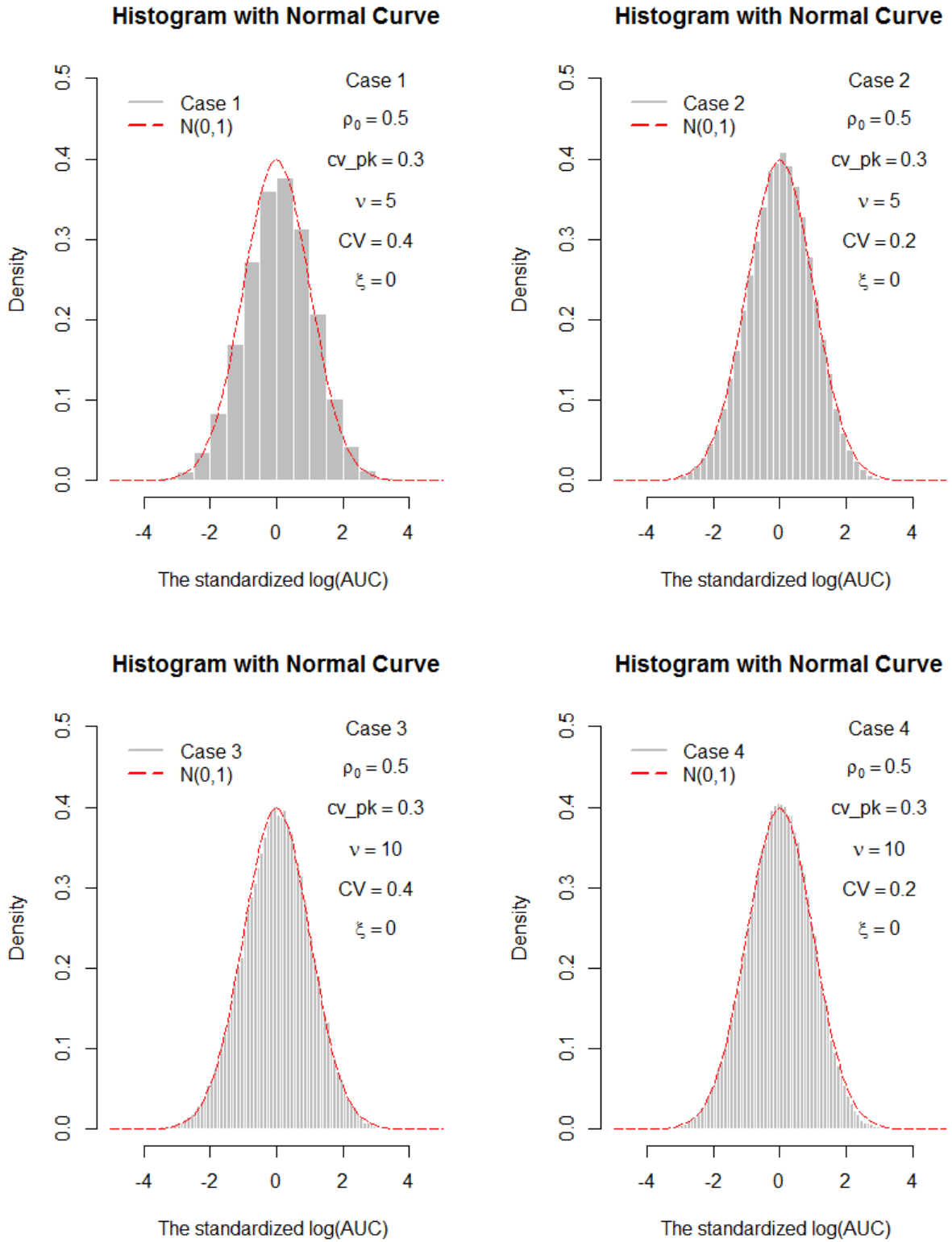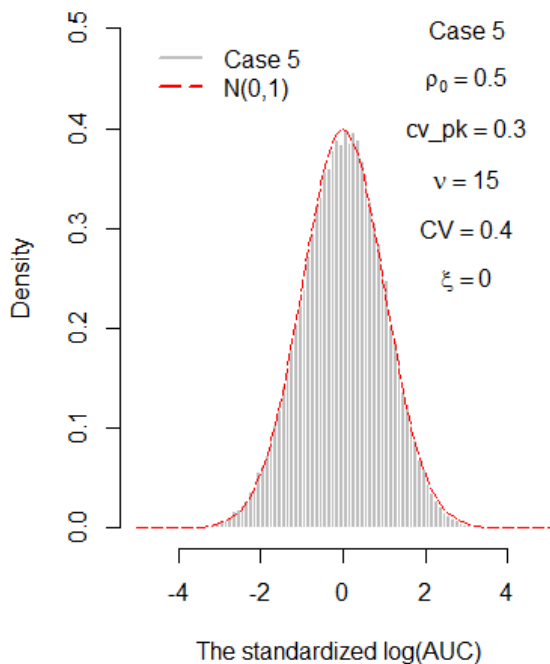
Histogram with Normal Curve — Case 5: $\rho_0 = 0.5$, $cv\_pk = 0.3$, $\nu = 15$, $CV = 0.4$, $\xi = 0$

Histogram with Normal Curve — Case 6: $\rho_0 = 0.5$, $cv\_pk = 0.3$, $\nu = 15$, $CV = 0.2$, $\xi = 0$

Histogram with Normal Curve — Case 7: $\rho_0 = 0.5$, $cv\_pk = 0.3$, $\nu = 20$, $CV = 0.4$, $\xi = 0$

Histogram with Normal Curve — Case 8: $\rho_0 = 0.5$, $cv\_pk = 0.3$, $\nu = 20$, $CV = 0.2$, $\xi = 0$

Histogram with Normal Curve

Case 9
$\rho_0 = 0.5$
cv_pk = 0.3
$\nu = 5$
CV = 0.4
$\xi = -\log(0.5)$

Histogram with Normal Curve

Case 10
$\rho_0 = 0.5$
cv_pk = 0.3
$\nu = 5$
CV = 0.2
$\xi = -\log(0.5)$

Histogram with Normal Curve

Case 11
$\rho_0 = 0.5$
cv_pk = 0.3
$\nu = 10$
CV = 0.4
$\xi = -\log(0.5)$

Histogram with Normal Curve

Case 12
$\rho_0 = 0.5$
cv_pk = 0.3
$\nu = 10$
CV = 0.2
$\xi = -\log(0.5)$
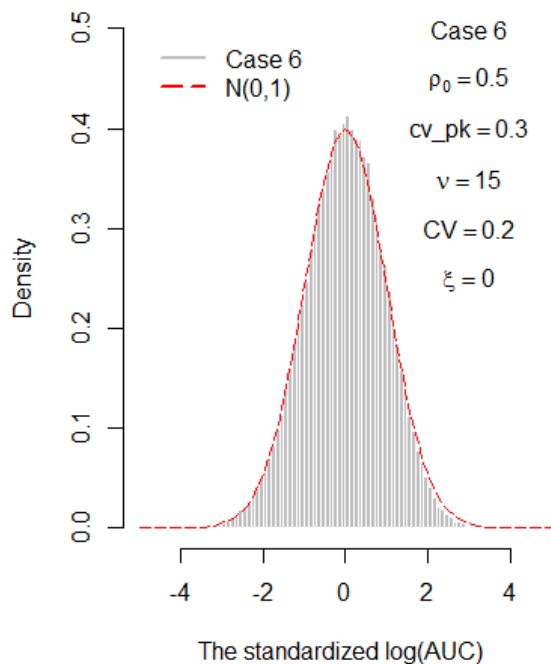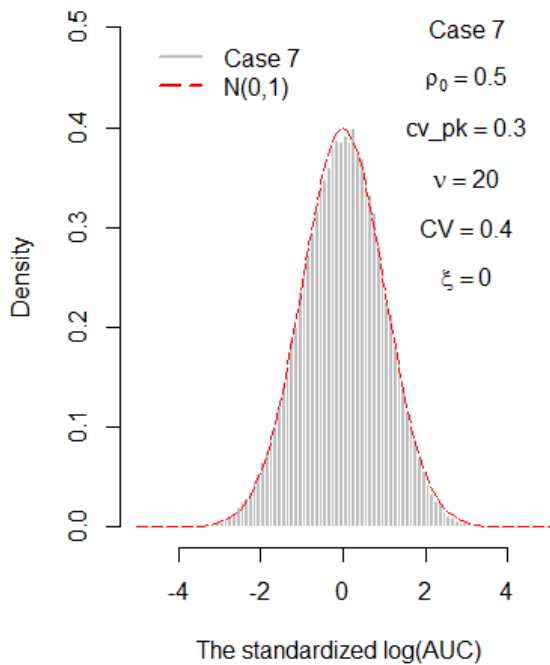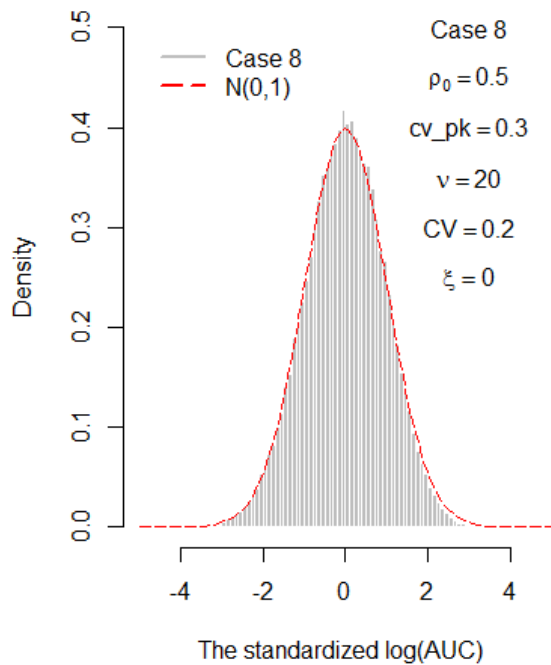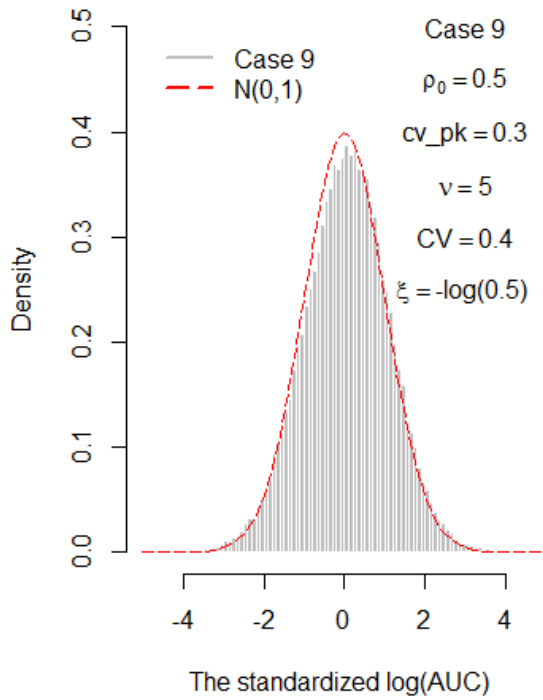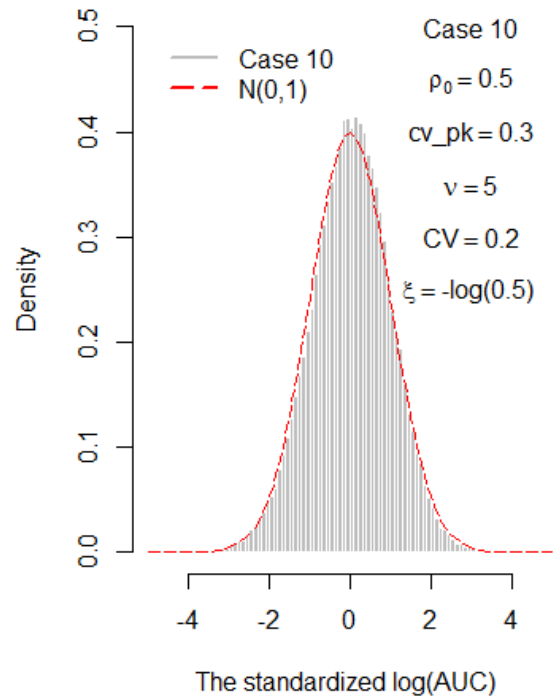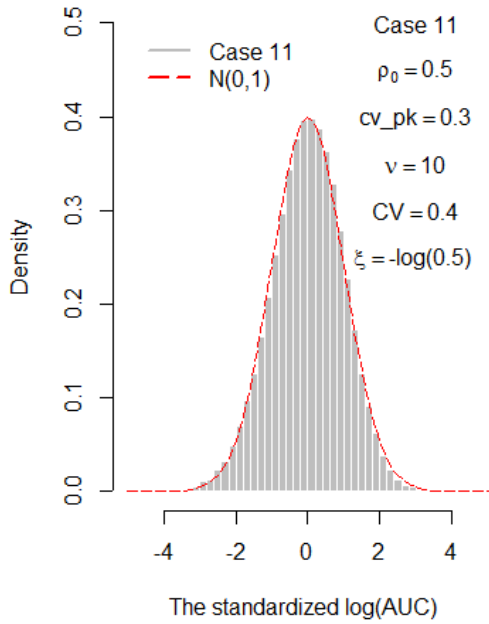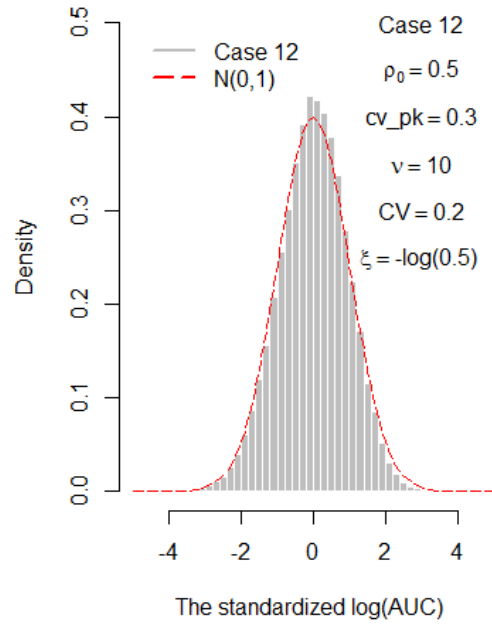
Figure 2.9 Histograms of the standardized log(*Cmax*) for cases in Table 2.3 compared to *N(*0,1)

Histogram with Normal Curve

Case 5

$\rho_0 = 0.5$

cv_pk = 0.3

$\nu = 15$

CV = 0.4

Legend: Case 5, N(0,1)

Y-axis: Density

X-axis: The standardized log(Cmax)

Histogram with Normal Curve

Case 6

$\rho_0 = 0.5$

cv_pk = 0.3

$\nu = 15$

CV = 0.2

Legend: Case 6, N(0,1)

Y-axis: Density

X-axis: The standardized log(Cmax)

Histogram with Normal Curve

Case 7

$\rho_0 = 0.5$

cv_pk = 0.3

$\nu = 20$

CV = 0.4

Legend: Case 7, N(0,1)

Y-axis: Density

X-axis: The standardized log(Cmax)

Histogram with Normal Curve

Case 8

$\rho_0 = 0.5$

cv_pk = 0.3

$\nu = 20$

CV = 0.2

Legend: Case 8, N(0,1)

Y-axis: Density

X-axis: The standardized log(Cmax)

43

Histogram with Normal Curve

Case 9
$\rho_0 = 0.5$
cv_pk $= 0.3$
$\nu = 5$
CV $= 0.4$
$\xi = $ -log(0.5)

Case 9
N(0,1)

Density

The standardized log(Cmax)

Histogram with Normal Curve

Case 10
$\rho_0 = 0.5$
cv_pk $= 0.3$
$\nu = 5$
CV $= 0.2$
$\xi = $ -log(0.5)

Case 10
N(0,1)

Density

The standardized log(Cmax)

Histogram with Normal Curve

Case 11
$\rho_0 = 0.5$
cv_pk $= 0.3$
$\nu = 10$
CV $= 0.4$
$\xi = $ -log(0.5)

Case 11
N(0,1)

Density

The standardized log(Cmax)

Histogram with Normal Curve

Case 12
$\rho_0 = 0.5$
cv_pk $= 0.3$
$\nu = 10$
CV $= 0.2$
$\xi = $ -log(0.5)

Case 12
N(0,1)

Density

The standardized log(Cmax)

44

Figure 2.10 CDF of the standardized log(*Tmax*) for cases in Table 2.3 compared to CDF of *N(*0,1)

CDF with Normal CDF Curve

Case 9
N(0,1)

Case 9
$\rho_0 = 0.5$
cv_pk = 0.3
$\nu = 5$
CV = 0.4
$\xi = -\log(0.5)$

Fn(x)

The standardized log(Tmax)

CDF with Normal CDF Curve

Case 10
N(0,1)

Case 10
$\rho_0 = 0.5$
cv_pk = 0.3
$\nu = 5$
CV = 0.2
$\xi = -\log(0.5)$

Fn(x)

The standardized log(Tmax)

CDF with Normal CDF Curve

Case 11
N(0,1)

Case 11
$\rho_0 = 0.5$
cv_pk = 0.3
$\nu = 10$
CV = 0.4
$\xi = -\log(0.5)$

Fn(x)

The standardized log(Tmax)

CDF with Normal CDF Curve

Case 12
N(0,1)

Case 4
$\rho_0 = 0.5$
cv_pk = 0
$\nu = 10$
CV = 0.2
$\xi = -\log(0.5)$

Fn(x)

The standardized log(Tmax)

2.6.2 Mixed population: one subgroup with a slower absorption process

We assume in this subsection that the population of subjects consists of two subgroups. One subgroup has a slower absorption process. This could happen if coating on the pill took longer to digest for some people. To illustrate how the distribution of the standardized log(*AUC*) or the standardized log(*Cmax*) for large samples of simulated subjects deviates from normality, given that there is a subgroup with a slower absorption rate, we will compare th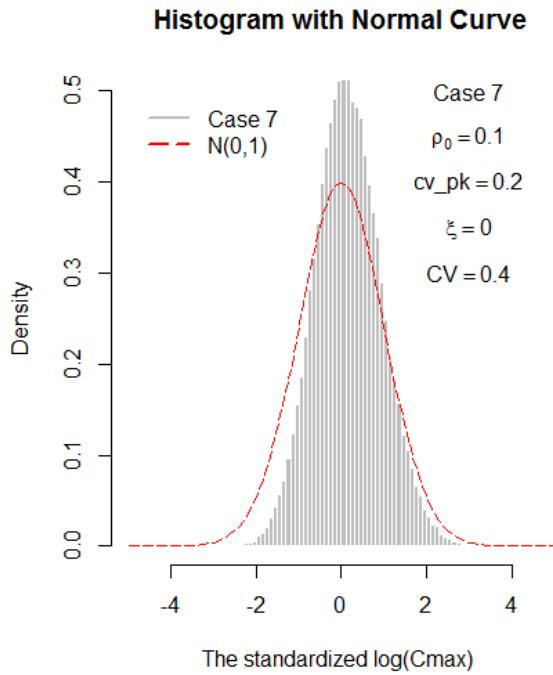e histogram of the standardized log(*AUC*) or the standardized log(*Cmax*) to the standard normal density curve. In this subsection we simulate the pharmacokinetic plasma-concentration profiles from two subpopulations: 70% of population has the mean ka=1.5 *hr⁻¹* and the rest of population has the mean ka=0.2 *hr⁻¹*. We still assume that $e_{i1},...,e_{im}$ are independent

and identically distributed normal variables for each i by letting $\xi = 0$, CV =0.2, and

the pharmacokinetic parameters' cv values and correlation at $cv_{pk} = (0.2, 0.2, 0.2, 0.2)'$,

and $\rho_0 = 0.3$.

   The left graph in Figure 2.11 shows that the histogram of the standardized

$\log(AUC)$ for 100,000 simulated subjects is very close to the standard normal density

curve. The right graph in Figure 2.11 shows that the histogram of the standardized

$\log(Cmax)$ for 100,000 simulated subjects appears to be quite different from a bell

shaped curve.

Figure 2.11 Comparison of standard normal density and sample histogram of the standardized $\log(AUC)$ and $\log(Cmax)$ from 100,000 subjects simulated from one-compartment model (1) with $CV=0.2$, ξ =0, and $\rho_0 = 0.3, cv_{pk} = (0.2, 0.2, 0.2, 0.2)'$ for a mixed population: 70% of population with the mean vector $\left(k_{a,i}, k_{e,i}, F_i/(1-F_i), V_{a,i}\right)' = (1.5, 0.12, 1, 525)'$ and 30% of population with the mean vector $\rho_0 = 0.5, cv_{pk} = (0.3, 0.3, 0.3, 0.3)'$, and $\left(k_{a,i}, k_{e,i}, F_i/(1-F_i), V_{a,i}\right)' = (0.2, 0.12, 1, 525)'$

47
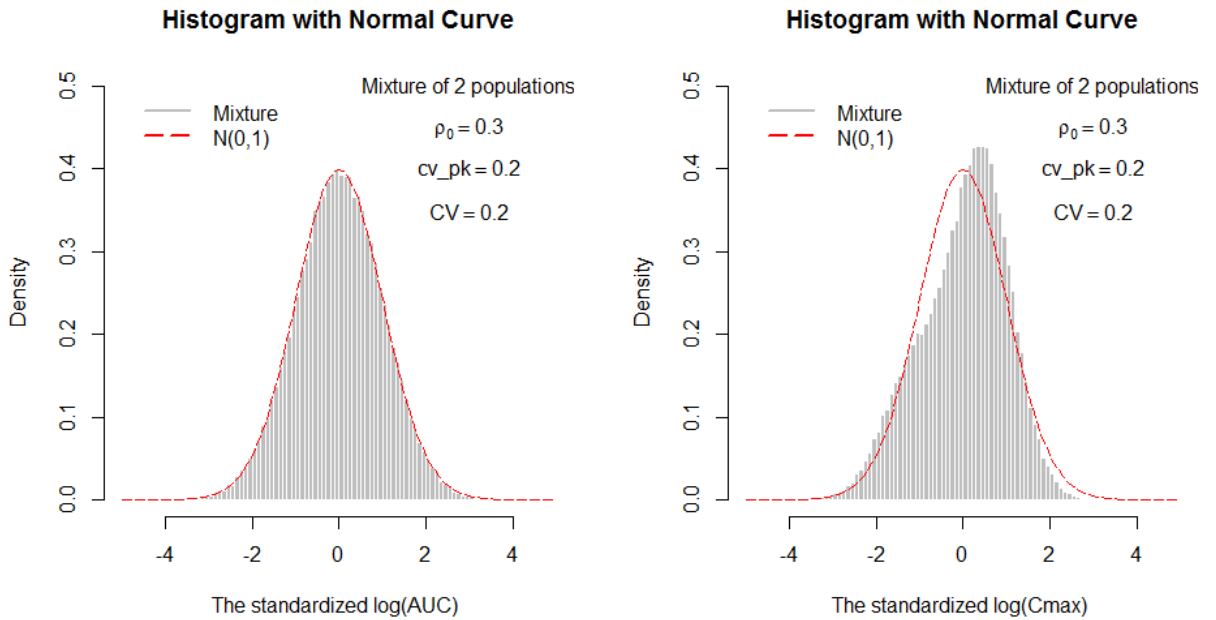
2.6.3 Two-compartment pharmacokinetic models with normal measurement errors

To contrast the results of previous simulations with those of analogous simulations from a two-compartment pharmacokinetic model, we would ideally consider the two-compartment model for the same drug Ropinirole for which we simulated the one-compartment model in Section 2.4. However, the pharmacokinetic parameters for the two-compartment model of Ropinirole are not available in the literature. Hence we have to switch to a different drug, Digoxin, whose pharmacokinetic parameters for the two-compartment pharmacokinetic model were published in [33]. The cardiac glycoside Digoxin [33] has a low therapeutic index and serious side-effects.

In a two-compartment model, Compartment 1 represents the central compartment, compartment 2 the "tissue" or peripheral compartment, $V_{a,k}$ the apparent volume of the kth compartment, and k_{jk} the apparent first-order rate constant for transfer of drug from the jth to the kth compartment (k = 0 represents an elimination process). The equation in [33] describing the time course of drug concentration in the central compartment of this model for the ith subject at time t_{ij} after an intravenous bolus injection (dose D) is

$$E(C_{ij} | \boldsymbol{\beta_i}) = \mu_{C_{ij}} = f^*(t_{ij}, \boldsymbol{\beta_i}) = \frac{D}{V_{a,i,1} \cdot (\lambda_i - \gamma_i)} \left((k_{21i} - \gamma_i) e^{-\gamma_i \cdot t_{ij}} - (k_{21i} - \lambda_i) e^{-\lambda_i \cdot t_{ij}} \right) \quad (2.6),$$

where $\lambda_i = 0.5 \left[(k_{12i} + k_{21i} + k_{10i}) + \sqrt{(k_{12i} + k_{21i} + k_{10i})^2 - 4 k_{21i} k_{10i}} \right]$,

$\gamma_i = 0.5 \left[(k_{12i} + k_{21i} + k_{10i}) - \sqrt{(k_{12i} + k_{21i} + k_{10i})^2 - 4 k_{21i} k_{10i}} \right]$, and $\boldsymbol{\beta_i} = (k_{12i}, k_{21i}, k_{10i}, V_{i1})'$.

Based on Models 2.3 to 2.5, we will simulate the concentration profiles in the central compartment of two–compartment pharmacokinetic model (2.6) for an

average person assumed to weigh 70 kilograms, who receives 1 mg Digoxin by rapid
bolus injection.

The following steps will give the details of how to simulate the plasma
concentration profiles:

1) Obtain the means of pharmacokinetic parameters $(k_{12}, k_{21}, k_{10}, V_{a,1})'$ from [33]. In
   [33], the volume of distribution $(V_{a,1})$ is 53.69 Liters (for an average person
   weighing 70 Kg); k_{12} is 0.76 hr^{-1}, k_{21} is 0.12 hr^{-1}, and k_{10} is 0.29 hr^{-1}.

2) Several sets of coefficients of variation of pharmacokinetic parameters are
   assumed after consideration of the estimated values in [33].

$$\text{Let } \boldsymbol{\eta} = E\begin{pmatrix} k_{12} \\ k_{21} \\ k_{10} \\ V_{a,1} \end{pmatrix} = \begin{pmatrix} 0.76 \\ 0.12 \\ 0.29 \\ 53.69 \end{pmatrix}, \text{ and } \mathbf{cv} = \begin{pmatrix} cv\_k_{12} \\ cv\_k_{21} \\ cv\_k_{10} \\ cv\_V_a \end{pmatrix}.$$

3) Assume that the vector of pharmacokinetic parameters and transformed
   parameters for the ith subject follows a log-normal distribution, denoted
   as $\begin{pmatrix} k_{12,i} & k_{21,i} & k_{10,i} & V_{a,1,i} \end{pmatrix}' \sim \text{log-normal}(\lambda, \Lambda)$. The correlation matrix of these log-
   transformed pharmacokinetic parameters is assumed to be $\mathbf{R}$ without any
   reference. Here we need to convert the marginal mean ($\eta[j], j \in \{1,2,3,4\}$) and
   coefficient of variation (cv[j], $j \in \{1,2,3,4\}$) for each untransformed
   pharmacokinetic parameter obtained from the reference into the marginal mean
   ($\lambda[j], j \in \{1,2,3,4\}$) of each log-transformed pharmacokinetic parameter and
   variance matrix (Λ) of log-transformed pharmacokinetic parameters by the
   following formulas.

$\lambda[j] = \log(\eta[j]) - 0.5 * \log(1 + (cv_{pk}[j])^2)$, $j \in \{1,2,3,4\}$, and

$$\mathbf{M} = (diag\{\sqrt{\log(1 + (cv_{pk}[j]^2))}\})_{4*4}.$$

Let $\mathbf{R} = \begin{pmatrix} 1 & \rho_0 & \rho_0 & \rho_0 \\ \rho_0 & 1 & \rho_0 & \rho_0 \\ \rho_0 & \rho_0 & 1 & \rho_0 \\ \rho_0 & \rho_0 & \rho_0 & 1 \end{pmatrix}$, it follows that $\mathbf{\Lambda} = \mathbf{M} \cdot \mathbf{R} \cdot \mathbf{M}$.

4) For each of m subjects, generate a random vector of pharmacokinetic parameters

and transformed parameters, $(k_{12,i}, k_{21,i}, k_{10,i}, V_{a,1,i})'$, from the distribution in Step 3.

Here $i=1,2,\ldots,m$.

For a given individual i, simulate concentration profiles (C_{ij}) at time points (in

hours),

$$t = (1/30, 1/15, 1/10, 2/15, 1/6, 7/30, 3/10, 11/30, 1/2, 3/4, 1, 2, 3, 4, 6, 8, 16, 24, 48, 72)'$$

and n_i is 20. We assume n_i is the same for all subjects. The choice of sampling

schedule follows [33].

Let t_{ij} be the jth sampling time point after dosing Subject i and let C_{ij} be the

concentration at t_{ij}, $j=1,2,\ldots,n_i$. Here,

$$E(C_{ij} | \boldsymbol{\beta_i}) = \mu_{C_{ij}} = f^*(t_{ij}, \boldsymbol{\beta_i}) = \frac{D}{V_{a,i,1} \cdot (\lambda_i - \gamma_i)} \left((k_{21i} - \gamma_i)e^{-\gamma_i \cdot t_{ij}} - (k_{21i} - \lambda_i)e^{-\lambda_i \cdot t_{ij}} \right).$$

5) Calculate AUC_i by evaluating the summation $\sum_{j=1}^{n_i-1} 0.5(C_{ij} + C_{i,j+1})(t_{i,j+1} - t_{ij})$ for

Subject i.

6) Obtain $Cmax_i = \max_j(C_{ij})$ from all observed values for Subject i.

7) Obtain $Tmax_i$ (time to reach $Cmax_i$) from all observed values for Subject i.

8) Obtain the goodness of fit of normality of $\log(AUC_i)$, $\log(Cmax_i)$, and $\log(Tmax_i)$, $i=1,2,\ldots,m$.

9) Repeat Steps 2 to 8 for $Snum$=10,000 times.

10) Calculate the rejection rates of goodness of fit tests of log-normality at the 0.05 significance level

by $\dfrac{\sum\limits_{l=1}^{Snum} I\left(pvalue_l \leq 0.05\right)}{Snum}$.

To illustrate how the distribution of the standardized $\log(AUC)$ or the standardized $\log(Cmax)$ deviates from normality if the data is described by a two-compartment model, we examine the sampling distribution of the standardized $\log(AUC)$ (or the standardized $\log(Cmax)$) under various combinations of pharmacokinetic parameters' variation and measurement errors. Since there are more pharmacokinetic parameters that vary from subject to subject in the two-compartment model than those in the one-compartment model, it is even more restrictive than before to assume that their across-subject joint distribution is multivariate lognormal. From Table 2.4, we can easily see that the 25th, 50th, and 75th quantiles for $\log(AUC)$ (or $\log(Cmax)$) simulated from the two-compartment model are correspondingly similar to the 25th, 50th, and 75th quantiles for log(AUC) (or $\log(Cmax)$) simulated from the one-compartment model. Hence the sampling distributions of the standardized $\log(AUC)$ for all cases in Figure 2.11 are similar to the first 8 cases in Figure 2.5 of Section 2.5.2. The sampling distributions of the standardized $\log(Cmax)$ for all cases in Figure 2.12 are similar to the first 8 cases in Figure 2.6 of Section 2.5.2. It seems that the sampling distributions of the standardized response variables for the studied cases here and in

Section 2.5.2 are not greatly affected by the choice of compartment model, but rather

by the distributions of the pharmacokinetic parameters and distribution of the

measurement errors.

Table 2.4 Quantiles of log(AUC) and log($Cmax$) at 0.05 significance level (10,000 simulations) for 40 subjects with $\mathbf{cv} = cv_{pk}[1,1,1,1]$ and residual errors $\mathbf{e_i} = [e_{i1},...,e_{in_i}]'$, ξ=0, and CV=0.4.

Case	$\log \boldsymbol{\eta} \sim N(\boldsymbol{\lambda}, \boldsymbol{\Lambda})$		Quantiles for $N(0,1)$: (25th, 50th, 75th)=(-0.674,0,0.674)					
			the standardized log(AUC)			the standardized log($Cmax$)		
	ρ_0	cv_{pk}	25th	50th	75th	25th	50th	75th
1	0.5	0.3	-0.792	0.044	0.875	-0.446	0.317	1.094
3	0.1	0.3	-0.728	0.011	0.741	-0.419	0.317	1.061
5	0.5	0.2	-0.616	-0.030	0.549	-0.356	0.248	0.878
7	0.1	0.2	-0.572	-0.048	0.475	-0.350	0.240	0.853

Figure 2.12 Histograms (breaks=80) of the standardized log(AUC) for cases in Table 2.4 compared to $N(0,1)$
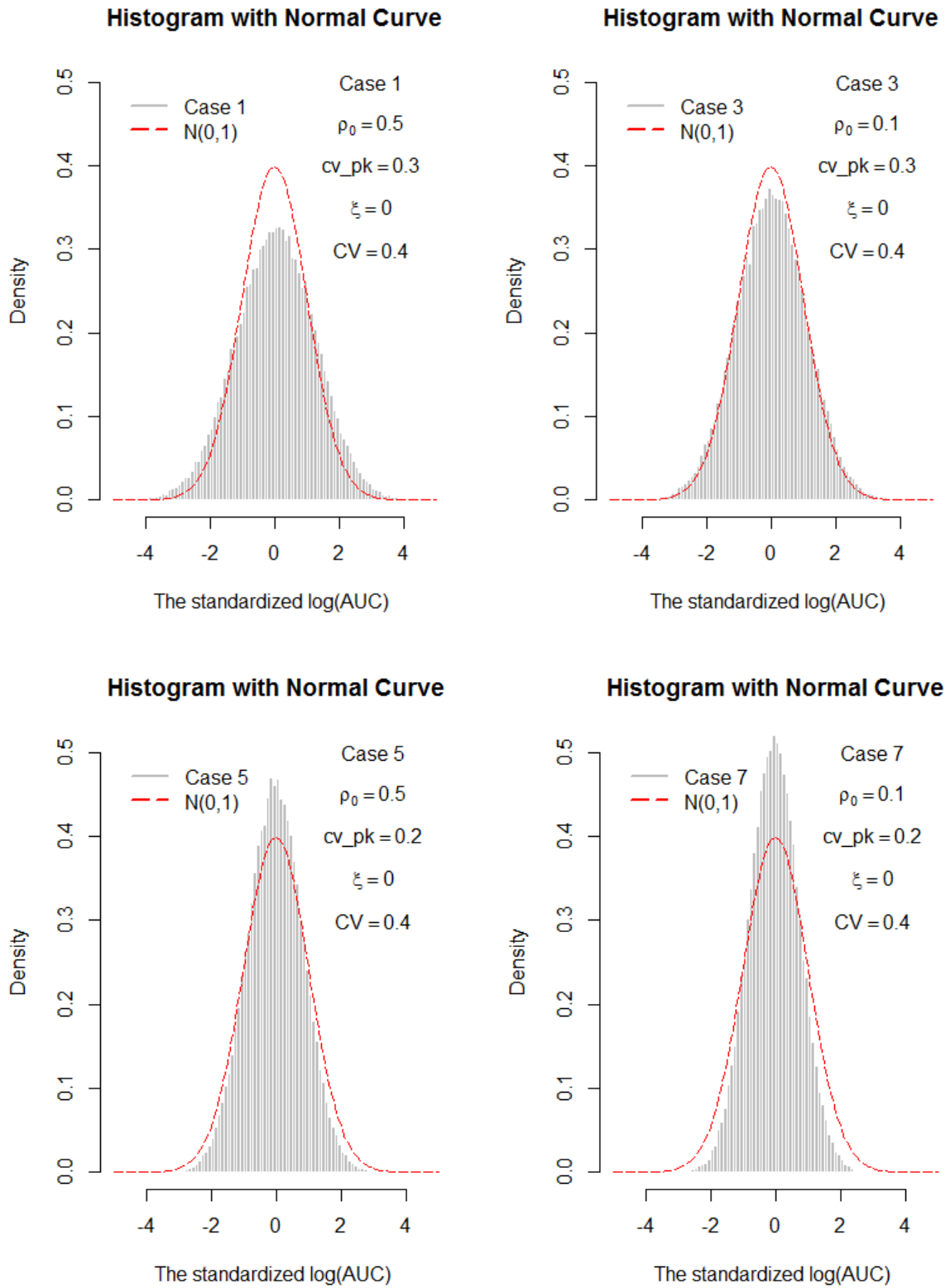
Figure 2.13 Histograms (breaks=80) of the standardized log($Cmax$) for cases in Table 2.4 compared to $N(0,1)$
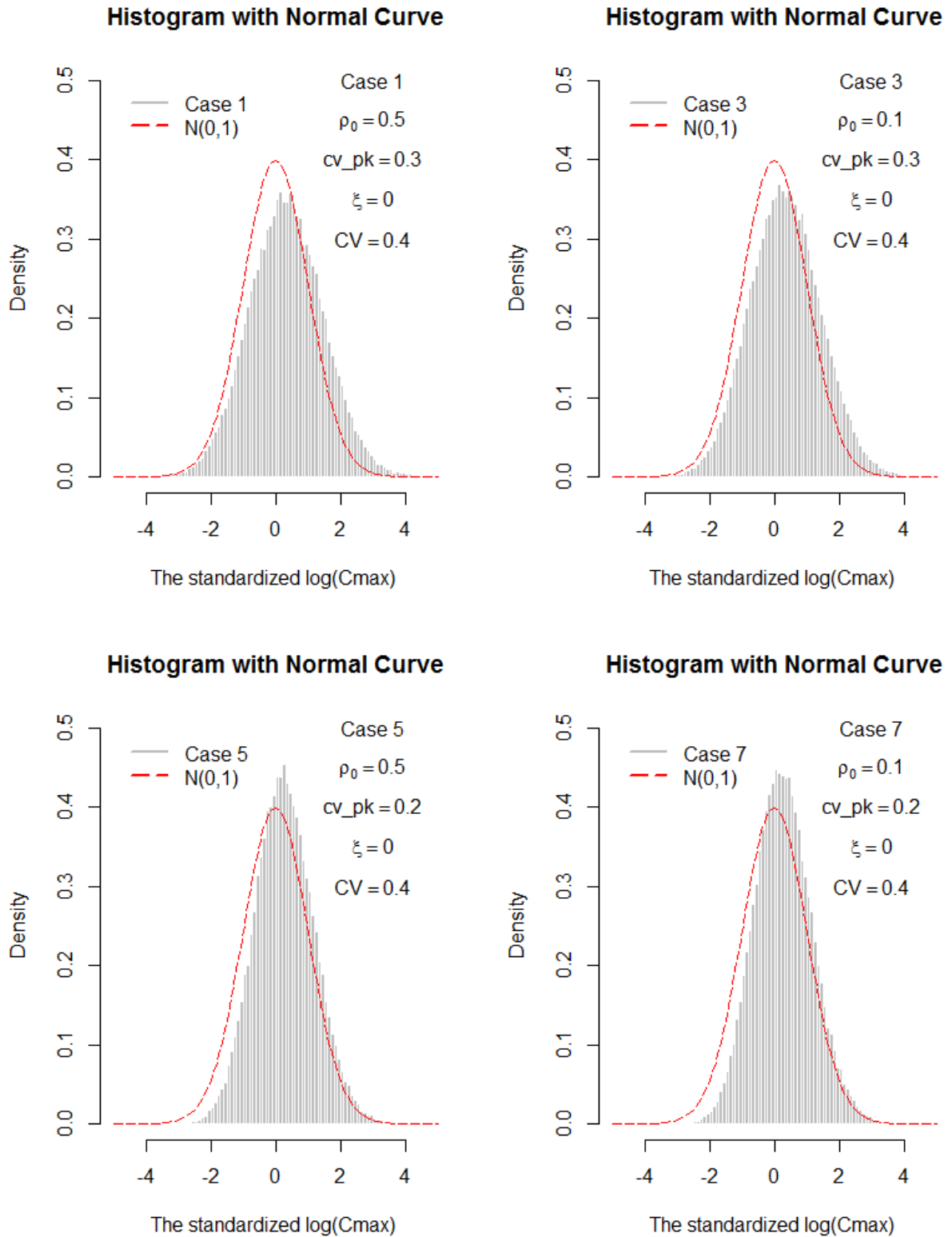
2.7 One real case

We would like to examine the sampling distributions of *AUC*, *Cmax*, and *Tmax* from a pharmacokinetic study of an orally administered agent.

This was a single-dose, randomized, open-label, two-period, two-sequence, two-treatment, crossover, comparative bioavailability study of the generic product to the innovative product. The products were studied using a crossover design with 40 normal, healthy volunteers being administered a single oral dose under fasting conditions. There were 39 subjects in the study because one subject withdrew from the study. Plasma concentration sampling times are pre-dose and at 0.25, 0.5, 0.75, 1.0, 1.25, 1.5, 1.75, 2.0, 2.5, 3.0, 3.5, 4.0, 5.0, 6.0, 8.0, 10.0, 12.0, 16.0, 24.0, and 36.0 hours post-dose.

Figures 2.14 and 2.15 respectively indicate that there are some deviations in the sampling density curve of the standardized log(*AUC*) and log(*Cmax*) from the standard normal density curve. Figure 2.16 indicates that there are a fixed number of discrete points in the cumulative distribution curve of the standardized log(*Tmax*), which is very different from standard normal cumulative distribution curve.

The null hypothesis of normality of log(*AUC*) (or log(*Cmax*)) is not rejected at 0.05 significance level because P-values from Shapiro-Wilk normality test of log(*AUC*) and log(*Cmax*) respectively are 0.7103 and 0.0981; both are larger than 0.05. The null hypothesis of the normality of log(*Tmax*) is rejected at 0.05 significance level because the p-value from Shapiro-Wilk normality test of log(*Tmax*) is 0.0183, smaller than 0.05.

Figure 2.14 Comparison of the standard normal density, empirical density and sample histogram of the standardized log(AUC) from 39 subjects of the real data set

Histogram with Normal Curve
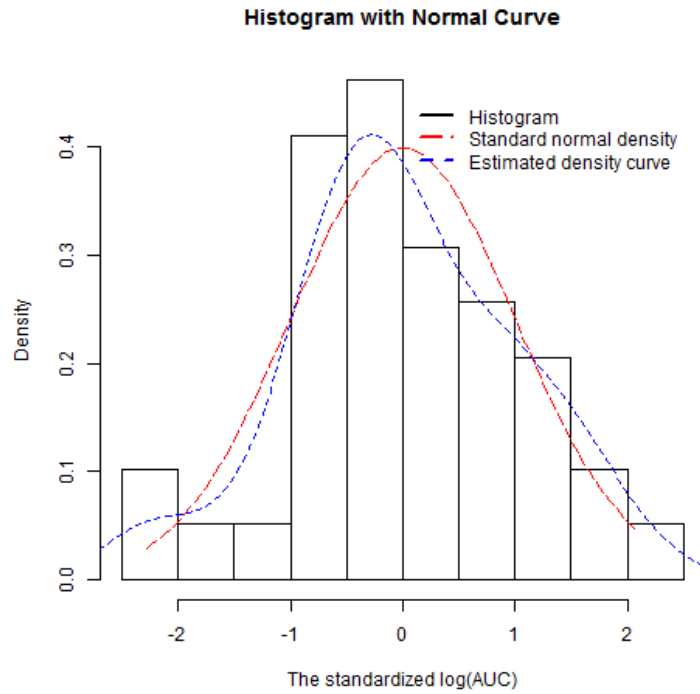
Figure 2.15 Comparison of standard normal density, empirical density and sample histogram of log(*Cmax*) from 39 subjects of the real data set

Histogram with Normal Curve

56
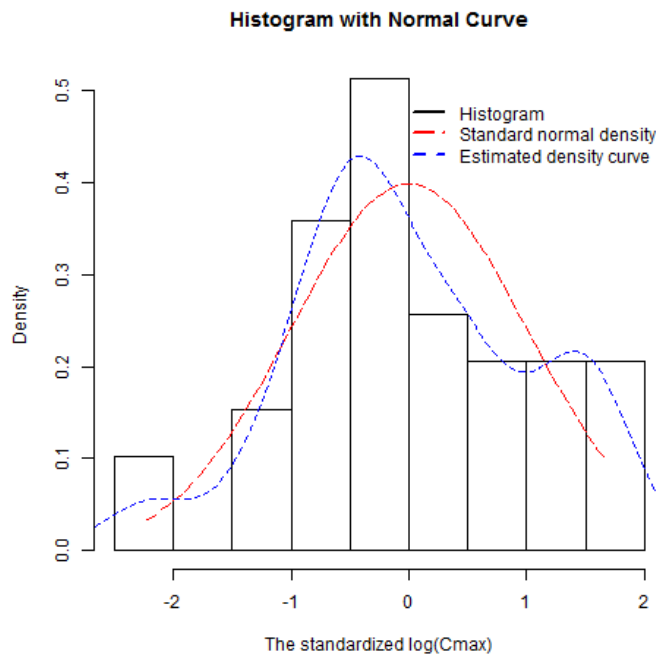
Figure 2.16 Comparison of standard normal distribution function and cumulative distribution function of standardized log(*Tmax*)

CDF with Normal Curve
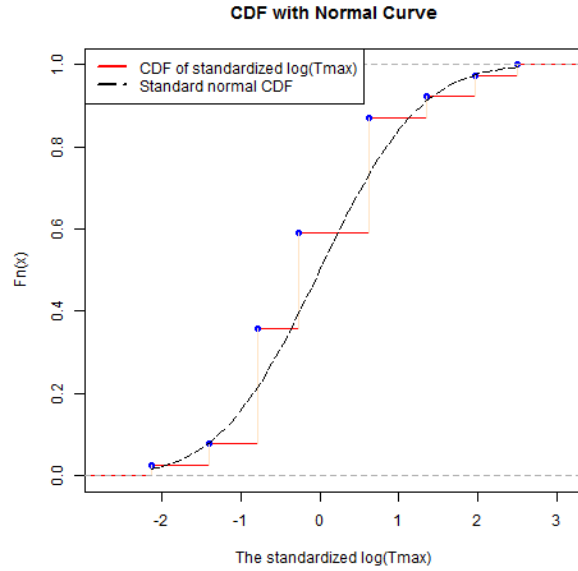
2.8 Discussion and conclusions

Strictly based on the statistical theory, AUC_i is not a log-normal variable even under the assumption that the plasma concentration (C_{ij}) at each time point (t_{ij}) is a log-normal random variable since $AUC = \sum_{j=1}^{n_i-1} 0.5(C_{ij} + C_{i,j+1})(t_{i,j+1} - t_{ij})$, which is a weighted sum of log-normal random variables. The examination of the sampling distributions of the standardized log(*AUC*) for 9 cases in Section 2.5 with large samples shows that the sampling distribution of the standardized log(*AUC*) sometimes has heavy tails compared to the normal distribution. But Figures 2.8 and 2.9 in Section 2.6.1 show that the normality assumption of log(*AUC*) seems reasonable if the independent and identical measurement errors are assumed to be *t* with more than 10 degrees of freedom, moderately heavy tails.

57

Since $Cmax_i = \max_j\left(C_{ij}\right)$ from all observed values for Subject i, $Cmax_i$ should

depend on the expected concentration curve. Furthermore, the possible time points to

obtain $Cmax_i$ are limited by the sampling schedule. So $Cmax_i$ should not be log-

normal variable even if C_{ij} is a log-normal random variable. The examination of the

sampling distributions of the standardized log(*Cmax*) for 9 cases in Section 2.5 with

large samples shows that the distribution of log(*Cmax*) is obviously skewed to the

right or the left. The 25th, 50th, and 75th percentiles of log(*Cmax*) are all different from

the 25th, 50th, and 75th percentiles for the standard normal distribution, respectively.

The close examination of the sampling distribution of the response variable for large

sample size study provides insight into its skewness and kurtosis. The sampling

distribution of log(*Cmax*) is skewed even if the independent and identical

measurement errors are assumed to be *t*. Figure 2.11 in Section 2.6.2 shows that the

normality assumptions of log(*Cmax*) is severely violated if there is a subpopulation

with a slower absorption process. Figures 2.12 and 2.13 in Section 2.6.3 show that the

sampling distribution of the sampling distribution of log(*Cmax*) is skewed for data

from two-compartment model under the combinations of ξ, ρ_0, and *CV* in Table 2.4.

However, the rejection rate of the Shapiro-Wilk normality test for many small sample

size studies does not provide the significant evidence for these cases. In other words,

the Shapiro-Wilk normality test for any small sample size study should not be

recommended in practice as statistical evidence for proving of the validity of the

normality of log(*AUC*) or log (*Cmax*).

Comparing the cases in Section 2.6.3 compared to those in Section 2.5.2, it appears that the sampling distribution of the response variable ($\log(AUC)$ or $\log(Cmax)$) is not greatly affected by the choice of compartment modeling.

The real example in Section 2.7 illustrates that histogram of $\log(AUC)$ or $\log(Cmax)$ shows some departure from normality.

The CDF of the standardized $\log(Tmax)$ has a fixed number of mass points and is above the standard normal CDF for almost every scenario. Since the number of possible values of $Tmax$ is limited by the number of sampling points, distribution of $Tmax$ most likely is discrete, not normally distributed.

In conclusion, the sampling distribution of $\log(AUC)$ with large samples often has heavy tails. The sampling distribution of $\log(Cmax)$ is skewed either to the left or to the right and is not robust to many perturbations studied in this chapter. Our examinations of the sampling distributions of $\log(AUC)$ (or $\log(Cmax)$) for a large number of simulated subjects helps to identify the nature of non-normality of $\log(AUC)$ (or $\log(Cmax)$). On the contrast, the rejection rate of Shapiro-Wilk normality test, which is not sensitive for many small samples (e.g., 40 subjects), cannot provide such insight. Hence it is necessary to examine the sampling distribution of the response variable for many more large sample size simulations with more extensive variation of pharmacokinetic parameters and distributions so that the nature of the distribution for $\log(AUC)$ (or $\log(Cmax)$) can be further evaluated. We must point out the limitation of our investigation since it is based on the simulations generated from the pharmacokinetic compartmental models and a lognormal measurement error structure, while the real concentration-time profile may

not fit the pharmacokinetic compartmental model and measurement error structure may be very different from lognormal.

Chapter 3 Background of the two one-sided tests for univariate bioequivalence testing

3.1 Introduction

In a typical bioavailability/ bioequivalence study, a test product (T) and a reference product (R) are administered to subjects. The reference product could be a marketed innovator's product previously approved by Food and Drug Administration (FDA) and the test product a potential generic substitute manufactured by a different pharmaceutical company. The test and reference products could also be different formulations, but manufactured by the same pharmaceutical company.

Assuming the response variable (*log(AUC)* or *log(Cmax)*) is normally distributed, Schuirmann [9] compared the two one-sided tests procedure and the power approach which tests the hypothesis of no difference, $H_0 : \mu_T = \mu_R$, for assessing the equivalence of average bioavailability in terms of rejection regions. In the bioequivalence evaluation of pharmacokinetic studies, the normality of the response variable (log(*AUC*) or log(*Cmax*)) is often assumed in practice. However the histograms for log(*Cmax*) are skewed for many simulation scenarios in Chapter 2.

Denote the population mean bioavailability (as measured by log*(AUC)* or log*(Cmax)*) of the test product by μ_T and the population mean bioavailability of the reference product by μ_R. In order to conclude the bioequivalence of the test product and the reference product, we should reject the null hypothesis in the context of the following pairs of hypotheses:

$$H_0 : \mu_T - \mu_R \leq \theta_1 \text{ or } \mu_T - \mu_R \geq \theta_2$$
$$H_1 : \theta_1 < \mu_T - \mu_R < \theta_2$$

(3.1).

Here θ_1 and θ_2 are pre-specified constants, also called equivalence margins, and $\theta_1 <$ θ_2.

The null hypothesis, H_0, states that μ_T and μ_R are not equivalent. The alternative hypothesis, H_1, states that they are equivalent.

The statistical hypotheses H_0 and H_1 given above are referred to as the "interval hypotheses" in the literature [10, 11]. The interval hypotheses H_0 and H_1 can be decomposed into two sets of one-sided hypotheses as shown in Section 3.3.

The purpose of this chapter is to provide a general background for the two one-sided testing and the power approach procedures. The power approach in practice usually consists of testing the hypothesis of no difference, $H_0 : \mu_T = \mu_R$, at level 0.05 and a lack of significance is often used to incorrectly infer equivalence. For assessing the equivalence of average bioavailability of the test and reference products, the power approach switches what ought to be the null and alternative hypotheses. The power approach too often frequently fails to reject the hypothesis of no difference with a small study size and/or larger variance, in essence rewarding the investigator or company for a less than adequate study. It is recognized that the power approach will reject $H_0 : \mu_T = \mu_R$ in large studies with smaller variability, but sometimes $H_0 : \mu_T = \mu_R$ is rejected although the two products are likely quite comparable. In Section 3.2, the linear mixed effect model for a two-period two-treatment crossover study is described in great detail. In Sections 3.3 and 3.4, the hypotheses and test statistics for two one-sided tests procedure and for the power approach are discussed,

respectively. In Section 3.5, the exact power functions for these two approaches are derived.

3.2 Linear mixed effect model for a two-period two-treatment crossover bioequivalence study

According to the 2001 Food and Drug Administration's Guidance for Industry [1]: Statistical Approaches to Establishing Bioequivalence, the Center for Drug Evaluation and Research (CDER) recommended that a standard in vivo bioequivalence (BE) study design be based on the administration of either single or multiple doses of the T and R products to healthy subjects on separate occasions, with random assignment to the two possible sequences of drug product administration. Hence the crossover design for an *in vivo* BE study is the primary design employed in regulatory trials.

In a two-period two-treatment crossover design, a group of n_1 subjects (Sequence 1) receives the reference drug, and a profile of the drug concentration within blood plasma over time for each subject is obtained. After a washout period for removal of any carryover effect, this group receives the test drug and drug plasma concentration-time profiles are again obtained. A second group of n_2 subjects (Sequence 2) receives the drugs in the reverse order.

As long as the two periods are sufficiently far apart it is reasonable to assume that there is no interaction between formulation and sequence and no carry-over effect. In the 2-period 2-treatment crossover study, the carry-over and interaction between formulation and sequence are not separately identifiable. The carry-over effect can occur when the active ingredients of a drug given in the first period are still present in

the second period. A washout period is incorporated into the study design to allow a drug administered in the first period to be washed out of the body before the other drug in the second period is taken. In that way, the plasma concentration of a drug administered in the first period will be reduced to a negligible level in the second period. Hence it is reasonable to assume no carry-over effect in this crossover design. However, one must assume no interaction between formulation and sequence for the analysis since the interaction between the formulation and period is not estimable in the 2-period 2-treatment crossover study. In practice, if these assumptions are suspect an alternate study design is used where different subjects are exposed to test and reference product.

Let Y_{ijk} be the response (e.g., $\log(AUC)$) of the kth subject in the jth period of the ith sequence in the 2-period 2-treatment crossover study, where $i=1, 2, j=1, 2$ and $k=1,..,$ n_i. Then Y_{ijk} is modeled by the linear mixed effect model

$$Y_{ijk} = \gamma + S_{ik} + P_j + F_{ij} + \varepsilon_{ijk}, \tag{3.2}$$

where γ is the overall mean; P_j is the fixed effect of period j; F_{ij} is the fixed effect of the formulation administered in period j of sequence i; S_{ik} is the random effect of subject k in sequence i (note: If subjects each have individual identifiers, rather than being labeled within sequences as in model 3.2, then S_{ik} is replaced by S_k.); and ε_{ijk} is the random error. From the treatment assignments, we know that $F_{11} = F_{22} = \mu_R$ and $F_{12} = F_{21} = \mu_T$. The parameters can be estimated only subject to restrictions: $P_1 + P_2$ $= \mu_T + \mu_R = 0$, since otherwise the parameters P_1, P_2, μ_T, and μ_R are not separately identifiable. The S_{ik} and the ε_{ijk} are all independent normal random variables with

mean 0. The variance of S_{ik} is σ_S^2 and the variance of ε_{ijk} is σ_T^2 if $i \neq j, \forall i, j = 1,2$ for the test formulation and σ_R^2 if $i = j, \forall i, j = 1,2$ for the reference formulation.

Let $\theta^* = \mu_T - \mu_R$, where μ_T and μ_R are the true means of the test and reference formulations, respectively. An estimator of θ^* is given as

$$\hat{D} = \frac{\overline{Y}_{12\bullet} - \overline{Y}_{11\bullet} + \overline{Y}_{21\bullet} - \overline{Y}_{22\bullet}}{2} \qquad (3.3).$$

Since

$$\hat{D} = \frac{\overline{Y}_{12\bullet} - \overline{Y}_{11\bullet} + \overline{Y}_{21\bullet} - \overline{Y}_{22\bullet}}{2}$$

$$= \frac{1}{2}\left[\begin{array}{l}\left(\gamma + \overline{S}_{1\bullet} + P_2 + F_{12} + \overline{\varepsilon}_{12\bullet}\right) - \left(\gamma + \overline{S}_{1\bullet} + P_1 + F_{11} + \overline{\varepsilon}_{11\bullet}\right) \\ + \left(\gamma + \overline{S}_{2\bullet} + P_1 + F_{21} + \overline{\varepsilon}_{21\bullet}\right) - \left(\gamma + \overline{S}_{2\bullet} + P_2 + F_{22} + \overline{\varepsilon}_{22\bullet}\right)\end{array}\right],$$

$$= \frac{F_{12} + F_{21} - F_{11} - F_{22}}{2} + \frac{\left(\overline{\varepsilon}_{12\bullet} - \overline{\varepsilon}_{11\bullet} + \overline{\varepsilon}_{21\bullet} - \overline{\varepsilon}_{22\bullet}\right)}{2}$$

then

$$E(\hat{D}) = \frac{F_{12} + F_{21} - F_{11} - F_{22}}{2} + \frac{E\left(\overline{\varepsilon}_{12\bullet} - \overline{\varepsilon}_{11\bullet} + \overline{\varepsilon}_{21\bullet} - \overline{\varepsilon}_{22\bullet}\right)}{2} = \mu_T - \mu_R \text{ and}$$

$$Var(\hat{D}) = \frac{Var\left(\overline{\varepsilon}_{12\bullet} - \overline{\varepsilon}_{11\bullet} + \overline{\varepsilon}_{21\bullet} - \overline{\varepsilon}_{22\bullet}\right)}{4} = \frac{\sigma_T^2 + \sigma_R^2}{4}\left(\frac{1}{n_1} + \frac{1}{n_2}\right).$$

For this crossover design, $\hat{D}$ is a normally distributed unbiased estimate of $\theta^* = $

$F_T - F_R = \mu_T\text{-}\mu_R$ with variance $\sigma_{\hat{D}}^2 = \dfrac{\sigma_R^2 + \sigma_T^2}{4}\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right).$

The standard error of $\hat{D}$ is $\hat{\sigma}_{\hat{D}} = S\dfrac{1}{2}\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}$, where

$$S^2 = \frac{1}{n_1 + n_2 - 2}\left[\sum_{k=1}^{n_1}(Y_{12k} - Y_{11k} - (\overline{Y}_{12\bullet} - \overline{Y}_{11\bullet}))^2 + \sum_{k=1}^{n_2}(Y_{21k} - Y_{22k} - (\overline{Y}_{21\bullet} - \overline{Y}_{22\bullet}))^2\right] \quad (3.4).$$

The estimate $\hat{D}$ is the average of the averages of the intra-subject difference between the test and the reference for the two sequences and S^2 is a pooled estimate of the variance of an intra-subject difference. $\dfrac{(n_1 + n_2 - 2)S^2}{\sigma_T^2 + \sigma_R^2}$ is distributed as χ^2 with $\nu = n_1 + n_2 - 2$ degrees of freedom.

Plugging the model value Y_{ijk} into S^2, we rewrite S^2 as:

$$S^2 = \frac{1}{n_1 + n_2 - 2}\left[\sum_{k=1}^{n_1}\left(\varepsilon_{12k} - \varepsilon_{11k} - (\bar{\varepsilon}_{12\bullet} - \bar{\varepsilon}_{11\bullet})\right)^2 + \sum_{k=1}^{n_2}\left(\varepsilon_{21k} - \varepsilon_{22k} - (\bar{\varepsilon}_{21\bullet} - \bar{\varepsilon}_{22\bullet})\right)^2\right].$$

$\dfrac{\sum_{k=1}^{n_1}\left(\varepsilon_{12k} - \varepsilon_{11k} - (\bar{\varepsilon}_{12\bullet} - \bar{\varepsilon}_{11\bullet})\right)^2}{\sigma_T^2 + \sigma_R^2}$ has a Chi-squared distribution with degrees of

freedom $n_1 - 1$ due to the fact that $\varepsilon_{12k} - \varepsilon_{11k}$ is distributed as $N\left(0, \sigma_T^2 + \sigma_R^2\right)$.

Similarly, $\dfrac{\sum_{k=1}^{n_2}\left(\varepsilon_{21k} - \varepsilon_{22k} - (\bar{\varepsilon}_{21\bullet} - \bar{\varepsilon}_{22\bullet})\right)^2}{\sigma_T^2 + \sigma_R^2}$ has a Chi-squared distribution with degrees

of freedom $n_2 - 1$. It is easy to see that $\dfrac{(n_1 + n_2 - 2)S^2}{\sigma_T^2 + \sigma_R^2}$ is distributed as $\chi^2_{n_1 + n_2 - 2}$.

So $E(S^2) = \dfrac{(n_1 - 1)\left(\sigma_T^2 + \sigma_R^2\right) + (n_2 - 1)\left(\sigma_T^2 + \sigma_R^2\right)}{n_1 + n_2 - 2} = \sigma_T^2 + \sigma_R^2$.

Model 3.2 is a typical model formulation for the 2-period 2-treatment bioequivalence study in the literature [2, 3]. Jones and Kenward's model [3] is:

$Y_{ijk} = \gamma + s_{ik} + P_j + F_{ij} + I(j = 2)\lambda_i + \varepsilon_{ijk}$, where λ_i is the carry-over effect of the ith sequence; this is the same as Model 3.2 if λ_i is assumed to be zero. Another variant of this model is assuming that s_{ik} is a fixed effect.

3.3 The two one-sided tests procedure

The interval hypotheses H_0 and H_1 in Equation 3.1 can be decomposed into two

sets of one-sided hypotheses:

$$H_{01} : \mu_T - \mu_R \le \theta_1$$
$$H_{11} : \mu_T - \mu_R > \theta_1$$

(3.5)

and

$$H_{02} : \mu_T - \mu_R \ge \theta_2$$
$$H_{12} : \mu_T - \mu_R < \theta_2 \quad .$$

(3.6)

The two one-sided tests procedure consists of rejecting the interval hypothesis H_0,

and thus concluding equivalence of μ_T and μ_R, if and only if both H_{01} and H_{02} are

rejected at a chosen nominal level of significance α.

From Model (3.2), the test statistics for hypothesis testing of H_0 in (3.1) are

defined as: $T_1 = \dfrac{\hat{D} - \theta_1}{\dfrac{S}{2}\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$,

(3.7)

$$T_2 = \dfrac{\hat{D} - \theta_2}{\dfrac{S}{2}\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} \quad .$$

(3.8)

Assuming that (1) the bioavailability is normally distributed, (2) the number of

subjects in each sequence are equal ($n_1 = n_2 = n/2$), (3) there are no missing

observations, and (4) the variance of test and reference products are equal (this

assumption need only be approximately correct.), the test statistics for hypothesis

testing of H_0 can be rewritten as: $T_1 = \dfrac{\hat{D} - \theta_1}{S\sqrt{1/n}}$ and $T_2 = \dfrac{\hat{D} - \theta_2}{S\sqrt{1/n}}$. We would conclude

the bioequivalence of the test and reference product if

$$T_1 = \frac{\hat{D} - \theta_1}{S\sqrt{1/n}} \geq t_\alpha(v) \text{ and } T_2 = \frac{\hat{D} - \theta_2}{S\sqrt{1/n}} \leq t_{1-\alpha}(v), \text{ where } \theta_1 \text{ and } \theta_2 \text{ are equivalence}$$

margins such that $\theta_1 < \theta_2$, and $t_\alpha(v)$ is the upper quantile of student t distribution with

α upper tail probability, and $v = n - 2$.

3.4 Power approach

The usual hypothesis testing of no difference as null hypothesis and nonzero

difference as alternative hypothesis at the nominal level of significance α (e.g., 0.05)

is called the power approach. This hypothesis is defined as:

$$H_0' : \mu_T - \mu_R = 0$$
$$H_1' : \mu_T - \mu_R \neq 0$$

(3.9)

The test statistic is defined as $T_p = \dfrac{\hat{D}}{S\sqrt{1/n}}$. One cannot reject H_0' if

$\left| \dfrac{\hat{D}}{S\sqrt{1/n}} \right| \leq t_{\alpha/2}(v)$, where $v = n-2$ is the number of degrees of freedom. The

bioequivalence of test and reference product is inferred.

3.5 Power functions for two one-sided tests procedure and power approach

3.5.1 Power function for the two one-sided tests procedure

In the two one-sided tests procedure (See Section 3.3), the power (P_1) is the

probability of rejecting both H_{01} and H_{02} given both H_{11} and H_{12} are true. Define

$X = \dfrac{vS^2}{2\sigma^2}$. Then $S = \sqrt{\dfrac{2X}{v}}\sigma$, and X is a chi-squared random variable with degrees of

freedom, v. Assume that $n_1 = n_2 = n/2$ and $\sigma_T^2 = \sigma_R^2 = \sigma^2$, we determine the power

function $P_1(\theta^*, n, \sigma)$ as follows:

$$P_1(\theta^*, n, \sigma) = P\left(\text{Reject } H_0 \mid \mu_T - \mu_R = \theta^*, \theta^* \in (\theta_1, \theta_2), \sigma\right)$$

$$= P\left(\frac{\hat{D} - \theta_1}{S\sqrt{1/n}} \geq t_\alpha(v), \frac{\hat{D} - \theta_2}{S\sqrt{1/n}} \leq t_{1-\alpha}(v) \middle| \mu_T - \mu_R = \theta^*, \theta^* \in (\theta_1, \theta_2), \sigma\right)$$

$$= P\left(\theta_1 + t_\alpha(v)S\sqrt{1/n} \leq \hat{D} \leq \theta_2 - t_\alpha(v)S\sqrt{1/n} \middle| \mu_T - \mu_R = \theta^*, \theta^* \in (\theta_1, \theta_2), \sigma\right)$$

Plugging $S = \sqrt{\dfrac{2X}{v}}\sigma$ into the above equation, we have

$$P_1(\theta^*, n, \sigma) = P\left(\theta_1 + t_\alpha(v)\sqrt{\frac{2x}{vn}}\sigma \leq \hat{D} \leq \theta_2 - t_\alpha(v)\sqrt{\frac{2x}{vn}}\sigma \middle| \mu_T - \mu_R = \theta^*, \theta^* \in (\theta_1, \theta_2), \sigma\right)$$

$$= E\left[P\left(\frac{\theta_1 - \theta^* + t_\alpha(v)\sqrt{\frac{2x}{vn}}\sigma}{\sigma\sqrt{2/n}} \leq \frac{\hat{D} - \theta^*}{\sigma\sqrt{2/n}} \leq \frac{\theta_2 - \theta^* - t_\alpha(v)\sqrt{\frac{2x}{vn}}\sigma}{\sigma\sqrt{2/n}} \middle| \theta^* \in (\theta_1, \theta_2), \sigma, X = x\right)\right]$$

$$= E\left[P\left(\frac{\theta_1 - \theta^*}{\sigma\sqrt{2/n}} + t_\alpha(v)\sqrt{\frac{x}{v}} \leq Z \leq \frac{\theta_2 - \theta^*}{\sigma\sqrt{2/n}} - t_\alpha(v)\sqrt{\frac{x}{v}} \middle| \theta^* \in (\theta_1, \theta_2), \sigma, X = x\right)\right]$$

$$= \int_0^c \left(\Phi\left(\frac{\theta_2 - \theta^*}{\sigma\sqrt{2/n}} - t_\alpha(v)\sqrt{\frac{x}{v}}\right) - \Phi\left(\frac{\theta_1 - \theta^*}{\sigma\sqrt{2/n}} + t_\alpha(v)\sqrt{\frac{x}{v}}\right)\right) \frac{1}{2^{v/2}\Gamma(v/2)} x^{v/2-1} e^{-x/2} dx$$

(3.10)

Here $c = \dfrac{n(\theta_2 - \theta_1)^2 v}{8 t_\alpha^2(v)\sigma^2}$, $\theta^* \in (\theta_1, \theta_2)$, $t_\alpha(v)$ is the 1-α quantile of the $t(v)$, and

$\Phi(\cdot)$ is the standard normal cumulative distribution function.

3.5.2 Power function for the power approach

In the power approach (Section 3.4), the power ($P_1^{Power}(\theta^*, \sigma)$) is the probability

of rejecting H_0' given H_1' is true when $n_1 = n_2 = n/2$ and $\sigma_T^2 = \sigma_R^2 = \sigma^2$; that is

69

$$P_1^{Power}(\theta^*, \sigma) = P\left(\text{Reject } H_0' \middle| \mu_T - \mu_R = \theta^*, \theta^* \neq 0, \sigma\right)$$

$$= P\left(\left|\frac{\hat{D}}{S\sqrt{1/n}}\right| \geq t_{\alpha/2}(\nu) \middle| \mu_T - \mu_R = \theta^*, \theta^* \neq 0, \sigma\right).$$

Plugging $S = \sqrt{\dfrac{2X}{\nu}}\sigma$ into the above equation, where X is a chi-squared random

variable with ν degrees of freedom, we have

$$P_1^{Power}(\theta^*, \sigma) = E\left[1 - P\left(-t_{\alpha/2}(\nu)\sigma\sqrt{2x/n\nu} \leq \hat{D} \leq t_{\alpha/2}(\nu)\sigma\sqrt{2x/n\nu} \middle| \theta^* \neq 0, X = x\right)\right]$$

$$= 1 - E\left[P\left(\frac{-\theta^* - t_{\alpha/2}(\nu)\sigma\sqrt{2x/n\nu}}{\sigma\sqrt{2/n}} \leq \frac{\hat{D} - \theta^*}{\sigma\sqrt{2/n}} \leq \frac{-\theta^* + t_{\alpha/2}(\nu)\sigma\sqrt{2x/n\nu}}{\sigma\sqrt{2/n}} \middle| \theta^* \neq 0, X = x\right)\right]$$

$$= 1 - E\left[P\left(\frac{-\theta^* - t_{\alpha/2}(\nu)\cdot\sigma\sqrt{2x/n}}{\sigma\sqrt{2/n}} \leq Z \leq \frac{-\theta^* + t_{\alpha/2}(\nu)\sigma\sqrt{2x/n\nu}}{\sigma\sqrt{2/n}} \middle| \theta^* \neq 0, X = x\right)\right]$$

$$= \int_0^\infty \left[1 - \Phi\left(\frac{-\theta^*}{\sigma\sqrt{2/n}} + t_{\alpha/2}(\nu)\sqrt{\frac{x}{\nu}}\right) + \Phi\left(\frac{-\theta^*}{\sigma\sqrt{2/n}} - t_{\alpha/2}(\nu)\sqrt{\frac{x}{\nu}}\right)\right]\frac{1}{2^{\nu/2}\Gamma(\nu/2)}x^{\nu/2-1}e^{-x/2}dx$$

(3.11)

Here $\theta^* \neq 0$, and $\Phi(\cdot)$ is a standard normal cumulative distribution function.

3.6 Discussion and conclusions

This chapter provides the general background of two one-sided tests procedure for

bioequivalence testing. Although Shuirmann [9] compared numerical value of power

of the two one-sided tests procedure and that of the power approach for assessing the

equivalence of average bioavailability, he did not provide the explicit power function

for two one-sided tests procedure. Hence this chapter derives the explicit power

functions for two one-sided tests procedure and the power approach. The two one-

sided tests procedure is much more widely used than the power approach and is the

only one recommended in the FDA guidance [1]. In addition, this chapter provides a

fundamental understanding of the difference between the two one-sided tests

procedure and the power approach.

Chapter 4 Exact calculation of power and sample size

in bioequivalence studies using two one-sided tests

4.1 Introduction

In Chapter 3, two approaches for testing bioequivalence in the context of a crossover design are described. We indicate that the approach based on two one sided tests is most commonly used in practice for the purpose of establishing bioequivalence.

As we described in Chapter 3, we reject the null hypothesis in the following hypothesis test (4.1) so as to conclude the bioequivalence of the test product and the reference product:

$$H_0 : \mu_T - \mu_R \leq \theta_1 \text{ or } \mu_T - \mu_R \geq \theta_2$$
$$H_a : \theta_1 < \mu_T - \mu_R < \theta_2 \tag{4.1}.$$

In this chapter, we describe various approaches to determine power and sample size in this setting.

Since Owen's Q function has been widely used, we would like to give a brief description about Owen's Q function and how to use it for calculating the probability of rejecting the null for the two one-sided tests procedure for the crossover bioequivalence study. Owen's Q function [35] is defined as:

$$Q_f(t, \delta; 0, c) = \frac{\sqrt{2\pi}}{\Gamma(f/2)2^{(f-2)/2}} \int_0^c G\left[\frac{t \cdot x}{\sqrt{f}} - \delta\right] x^{f-1} G'(x) dx, \text{ where } G'(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \text{ and}$$

$G(x) = \int_{-\infty}^{x} G'(x)dx$, f is degrees of freedom, δ is noncentrality parameter, and c is

upper integration limit. With Owen's Q function, the probability of rejecting the null

for the two one-sided tests procedure for the crossover bioequivalence study is equal

to $Q_v(t_\alpha(v), \delta_2; 0, c) - Q_v(-t_\alpha(v), \delta_1; 0, c)$ (see Equation (11) on Page 439 of [35]) under

$n_1 = n_2 = n/2$ and $\sigma_T^2 = \sigma_R^2 = \sigma^2$. Here σ_T^2 and σ_R^2 are the variance for the test product

and the reference product, respectively; n_1 and n_2 are respectively the number of

subjects in Sequence 1 and Sequence 2; θ^* is the true mean difference; $v = n - 2$;

$$\delta_1 = \frac{\theta^* - \theta_1}{\sigma\sqrt{2/n}}; \; \delta_2 = \frac{\theta^* - \theta_2}{\sigma\sqrt{2/n}}; \text{ and } c = \frac{n(\theta_2 - \theta_1)^2 v}{8t_\alpha^2(v)\sigma^2}.$$

In 1990, Phillips [36] calculated the power and the sample size for the two one-

sided tests procedure with equivalence margins (θ_1 =-20% and θ_2 =20%) for

bioequivalence assessment of normal data based on Owen's special case of bivariate

noncentral t-distribution [35] and presented some sample size tables and power

graphs. We note that Phillips did not provide the formula for calculating R (an

argument in Owen's Q function) in his method. (See pages 138 and 139 of [36].)

Phillips' [36] calculations were for equivalence margins $\theta_2 = -\theta_1$ =0.20, different

from the values $\theta_2 = -\theta_1 = \log 1.25 = 0.2231$ recommended in [1], the FDA

Guidance for Industry. In 1991, Diletti, Hauschke, and Steinijans [37] determined the

sample size for the two one-sided tests procedure with equivalence margins (θ_1

=log(0.8) and θ_2 =log(1.25)) for bioequivalence assessment of one log-transformed

response variable that is assumed to be normally distributed on the basis of Owen's

special case of bivariate noncentral t-distribution. Again we note that Diletti et al did

not provide the formula for calculating c (an argument in Owen's Q function) in their method. (See Page 5 of [37].) They presented the graphs of the power against the number of subjects for various coefficients of variations (standard deviation divided by the mean) of the untransformed data. There are graphs for the above probability but no explicit formula for this probability [36, 37]. In 1992, Liu and Chow [38] derived approximate sample size formulas for two one-sided tests procedure with equivalence margins $\theta_1 = \log(0.8)$ and $\theta_2 = \log(1.25)$ for bioequivalence assessment of one log-transformed response variable that is assumed to be normally distributed. These margins correspond to the margin in the 2001 Food and Drug Administration's Guidance for Industry [1]. Hauschke, Steinijans, Diletti, and Burke [39] compared the sample size based on Owen's bivariate noncentral t-distribution and an approximate formula in the case of a multiplicative model, in which period, treatment effect, subject effect, and residual acts proportionally on AUC or $Cmax$. Taking logarithms of both sides in the multiplicative model transforms the multiplicative model on the original scale to the additive model (in which the different components affected the response variable additively) on the logarithmic scale. In 1999, Kieser and Hauschke [40] proposed a unifying approach to approximate sample size determination for different types of hypotheses formulated in terms of ratio of two means and difference of two means for the situations of testing noninferiority, superiority, or equivalence when the response variable is normally distributed. In 2000, Kieser and Hauschke [41] proposed an approximate sample size formula using both inter-subject and intra-subject variability for demonstrating equivalence in crossover trials based on the ratio of two location parameters. In 2001, Chow and Wang [42] derived

74

approximate formulas for sample size calculation under a balanced crossover design with equal variances of the test and reference products and a parallel design with normally distributed raw data or normally distributed log-transformed data.

In 2007, Hauschke et al [43] published an explicit formula using Owen's Q function [35] with properly defined noncentrality parameters and R under $n_1 = n_2 = n/2$ and $\sigma_T^2 = \sigma_R^2 = \sigma^2$. In 2009, Phillips [44] published an explicit formula for the probability of rejecting the null for the two one-sided tests procedure for the crossover bioequivalence study under $n_1 = n_2 = n/2$ and $\sigma_T^2 = \sigma_R^2 = \sigma^2$. Hauschke et al [43] and Phillips [44] provided a clearer explanation for what these same authors had essentially done in 1991 and 1990, respectively. However, Phillips [44] is newly augmented with references to a specific readily available R-package [45]. There have also been several published approximations [38-42] that have been used in practice. In 2008, the formula for power calculation on the left side below Figure 4 on Page 252 of [46] is the difference of cumulative noncentral t function, which is approximate. In 2014, the formula in Line 5 in Section 2.2 of [46] multiplied by Equation (2) in Section 4 of [46] for sample size calculation at nonzero true mean difference of the test and reference products same as Chow and Wang's approximation is still used. Equation (1) in Section 2.2 of [47] multiplied by Equation (2) in Section 4 of [47] for sample size calculation at zero true mean difference of the test and reference products is also the same as Chow and Wang's approximation. In 2014, Shen, Russek-Cohen, and Slud published the exact power formula for calculating the probability of rejecting the null for the two one-sided tests procedure for the crossover bioequivalence study under general parameters [48] based on the

work in this chapter. The main purpose of [48] is to explain clearly just how

untenable the approximations are for realistic parameter combinations and how easily

the self-contained power expressions, that have been essentially known since Owen's

1965 paper [35] are to implement in R.

We note that except for [48] all of these approaches, both exact and approximate,

assume the variances $\sigma_T^2 = \sigma_R^2$ are homogeneous, that the design is balanced

($n_1 = n_2$), and that measures of bioavailability (log(AUC) and log($Cmax$)) are

normally distributed. In this chapter, we compare approaches when the data is

normally distributed and in Chapter 6 we will address the sensitivity of the exact

method to the assumption of normality.

In Section 4.2, we briefly derive the joint density function of test statistics and the

exact formula for the power of the two one-sided tests procedure for testing

bioequivalence based on a univariate normally distributed response variable. Our

derivation for the exact power under general parameters just serves for completeness

and for expository purposes since the explicit power formula assuming the equal

variances for the test and reference products and balanced design was published in

[43, 44]. However, by allowing for unequal variances, it might allow the user to

assess the robustness of the power and sample size determinations when this

assumption is violated. Modest differences in variability may not cause concern when

the bioequivalence question is being addressed.

In Section 4.3, we compare the simulated values with the numerical values and

indicate how numerical integration easily provides accurate numerical values for

power and sample size. In Section 4.4, we compare the numerical results of the exact

method with the graphed power values of Diletti et al. [37], the power values graphed

by Phillips [36] for $\theta_2 = -\theta_1 = 0.20$, and the sample size and power values generated by

the approximate method of Chow and Wang [42].

4.2 Exact power function and the joint probability density function of test statistics
(T₁ and T₂)

From Section 3.2 in Chapter 3, we recall that $\hat{D}$ is a normally distributed unbiased

estimate of $\theta^* = \mu_T - \mu_R$ with variance $\sigma_{\hat{D}}^2 = \dfrac{\sigma_R^2 + \sigma_T^2}{4}\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)$, and

$\dfrac{(n_1 + n_2 - 2)S^2}{\sigma_T^2 + \sigma_R^2}$ is distributed as χ^2 with $\nu = n_1 + n_2 - 2$.

We use the test statistics $T_1 = \dfrac{\hat{D} - \theta_1}{\dfrac{S}{2}\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$ and $T_2 = \dfrac{\hat{D} - \theta_2}{\dfrac{S}{2}\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$ defined in Section

3.3 in Chapter 3 testing for hypotheses (4.1). To correct the mis-statement of Liu and

Li [49] that the test statistics (T_1 and T_2) do not have a joint density (see first two lines

after Equation (3) in [49]), we simply observe that the function mapping ($\hat{D}, S^2$) to

(T_1, T_2), with domain $(-\infty, \infty) \times (0, \infty)$ and range $\{ (t_1, t_2): t_1 > t_2 \}$, is both

differentiable and differentiably invertible, while the independent variables $\hat{D}$ and

S^2 have respective normal and Gamma densities, as described above. (For an

exposition of the change-of-variable formula for differentiable one-to-one

transformations of random vectors with a joint density, see Formula 4.3.2 on Page

158 and Theorem 2.1.8 on page 53 of [50]). The joint density for (T_1, T_2) can also be

obtained directly by differentiation with respect to t_1 and t_2 on the region $t_1 > t_2$ of

the joint cumulative distribution function derived and published by Owen [35, Sec. 5]).

The joint pdf of (T_1, T_2) is given by

$$f_{T_1,T_2}(t_1,t_2) = |J| \cdot f_{\hat{D}_1,Y}(h_1(t_1,t_2), h_2(t_1,t_2))$$

$$= \frac{1}{t_1 - t_2} \frac{2^{v/2+1}}{\Gamma(v/2) \cdot \left(\frac{\sigma_T^2 + \sigma_R^2}{v}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right)^{v/2}} \left(\frac{\theta_2 - \theta_1}{t_1 - t_2}\right)^{1+v}$$

$$\times \exp\left(-\frac{2v}{(\sigma_T^2 + \sigma_R^2)}\left(\frac{\theta_2 - \theta_1}{t_1 - t_2}\right)^2 \left(\frac{n_1 n_2}{n_1 + n_2}\right)\right) \qquad \text{, where } t_1 > t_2. \quad (4.2)$$

$$\times \sqrt{\frac{2n_1 n_2}{\pi(\sigma_T^2 + \sigma_R^2)(n_1 + n_2)}} \exp\left(-\frac{2\left(\frac{t_1\theta_2 - t_2\theta_1}{t_1 - t_2} - \theta^*\right)^2}{\sigma_T^2 + \sigma_R^2} \frac{n_1 n_2}{n_1 + n_2}\right)$$

Although the exact power function can easily be developed as a double integral

from the joint pdf of T_1 and T_2, we will derive a simpler form of the exact power

function by integrating the conditional power given S^2. This alternative formula, in

the same spirit as formula (11) in Section 5 of [35], is simpler because it involves

only a univariate integral over a bounded interval, after recognizing that the

conditional power given S^2 is a readily evaluated normal tail probability.

The exact power function can be written, in terms of the α and 1-α quantiles $t_v(\alpha)$

and $t_v(1-\alpha) = -t_v(\alpha)$ of the t distribution with v degrees of freedom, as:

$$P_1\left(\theta^*, n_1, n_2, \sigma_T, \sigma_R\right) = P\left(\text{Reject } H_0 \mid \mu_T - \mu_R = \theta^*, \theta^* \in (\theta_1, \theta_2), \sigma_T, \sigma_R\right)$$

$$= P\left(\frac{\hat{D} - \theta_1}{\dfrac{S}{2}\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} \geq t_\alpha(\nu) \cap \frac{\hat{D} - \theta_2}{\dfrac{S}{2}\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} \leq t_{1-\alpha}(\nu) \Big| \mu_T - \mu_R = \theta^*, \theta^* \in (\theta_1, \theta_2), \sigma_T, \sigma_R\right)$$

$$= P\left(\theta_1 + t_\alpha(\nu)\frac{S}{2}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \leq \hat{D} \leq \theta_2 - t_\alpha(\nu)\frac{S}{2}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \Big| \mu_T - \mu_R = \theta^*, \theta^* \in (\theta_1, \theta_2), \sigma_T, \sigma_R\right)$$

This last expression is the expectation over S of the conditionally normal probability given S, using the fact that $\dfrac{\nu \cdot S^2}{\sigma_T^2 + \sigma_R^2} \sim \chi^2(\nu)$ and $\hat{D} \sim N\left(\theta^*, \sigma_{\hat{D}}^2\right)$ are independent:

$$E\left[\Phi\left(\frac{\theta_2 - \theta^*}{\sigma_{\hat{D}}} - t_\alpha(\nu)\frac{S}{\sqrt{\sigma_T^2 + \sigma_R^2}}\right) - \Phi\left(\frac{\theta_1 - \theta^*}{\sigma_{\hat{D}}} + t_\alpha(\nu)\frac{S}{\sqrt{\sigma_T^2 + \sigma_R^2}}\right), I\left(S \leq \frac{\theta_2 - \theta_1}{t_\alpha(\nu)}\left(\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right)^{-1}\right)\right], \text{ where}$$

$t_\alpha(\nu)$ is the 1-α quantile of the *t*-distribution with $\nu = n_1 + n_2 - 2$ degrees of freedom and $\Phi(\cdot)$ is the standard normal cumulative distribution function, $I(\cdot)$ is the indicator function, and $\sigma_{\hat{D}}^2 = \dfrac{\sigma_R^2 + \sigma_T^2}{4}\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)$. Using the fact that $\dfrac{\nu \cdot S^2}{\sigma_T^2 + \sigma_R^2} \sim \chi^2(\nu)$, we write the above expectation over S explicitly as

$$P_1\left(\theta^*, n_1, n_2, \sigma_T, \sigma_R\right)$$
$$= \int_0^{c_1}\left(\Phi\left(\frac{\theta_2 - \theta^*}{\sigma_{\hat{D}}} - t_\alpha(\nu)\sqrt{\frac{x}{\nu}}\right) - \Phi\left(\frac{\theta_1 - \theta^*}{\sigma_{\hat{D}}} + t_\alpha(\nu)\sqrt{\frac{x}{\nu}}\right)\right)\frac{1}{2^{\nu/2}\Gamma(\nu/2)}x^{\nu/2-1}e^{-x/2}dx \qquad (4.3)$$

Here $c_1 = \dfrac{(\theta_2 - \theta_1)^2 \nu}{4 t_\alpha^2(\nu)\sigma_{\hat{D}}^2}$ and $\sigma_T, \sigma_R > 0$.

The gamma density and integrand written in terms of the normal cdf are readily evaluated, so the integral (4.3) is easily evaluated in any good statistical computing package: R code for it is given in Appendix 4.1.

In the two one-sided tests procedure for one single variable, the power (P_1) is the probability of rejecting H_0 when H_a in (4.1) is true. For sample size determination and power of two one-sided tests procedure in the bioequivalence literature, Schuirmann [9], Phillips [36], Diletti et al. [37], and Chow and Wang [42] all assumed that $n_1 = n_2 = n/2$ and $\sigma_T^2 = \sigma_R^2 = \sigma^2$. Following this convention, we will compare the exact power with Chow and Wang's approximate power under the assumption that $n_1 = n_2 = n/2$ and $\sigma_T^2 = \sigma_R^2 = \sigma^2$. The exact power function becomes:

$$P_1\left(\theta^*, n, \sigma\right) = \int_0^{c_2} \left(\Phi\left(\frac{\theta_2 - \theta^*}{\sigma\sqrt{2/n}} - t_\alpha(v)\sqrt{\frac{x}{v}} \right) - \Phi\left(\frac{\theta_1 - \theta^*}{\sigma\sqrt{2/n}} + t_\alpha(v)\sqrt{\frac{x}{v}} \right) \right) \frac{1}{2^{v/2}\Gamma(v/2)} x^{v/2-1} e^{-x/2} dx . \quad (4.4)$$

Here $c_2 = \dfrac{n(\theta_2 - \theta_1)^2 v}{8t_\alpha^2(v)\sigma^2}$ and $\sigma > 0$.

4.3 Comparison of power values from the exact power function and Monte Carlo simulations

The power values computed numerically from (4.4) using a standard numerical integration routine, **integrate** in R [51], were carefully checked both in terms of their own estimated error bounds and by comparing with Monte Carlo simulations of rejections in the two one-sided tests procedure. Table 4.1 lists the results from the exact power function and Monte Carlo simulations. From Table 4.1, we found using 10^6 Monte Carlo replications (with corresponding simulation standard errors less than 0.0005), for a combination of cases of θ^* equal to 0, 0.1, 0.2 and log(1.25), and of σ = 0.2 and 0.3, that the simulated and exact values were always within 0.001 of one another, and that the numerical integration error bounds were less than 0.0001 (usually by one or more orders of magnitude).

Since $\nu/t_\alpha^2(\nu)$ is an increasing function of ν (see the proof in Appendix 4.2), both

the upper limit of integration and the integrand in the integral formula (4.4) are

directly seen to be monotone increasing as a function of n, so the integral (4.4) itself

is also monotone increasing in n. Hence we use a bisection search or other numerical

root-finder to find the required sample size n by first solving for the continuous value

n at which the exact power (4.4) with all parameters held fixed is equal to power $(1-\beta)$

and then rounding it up to the smallest even number $n \geq n^*$. Code lines for doing this

in **R** are also supplied in Appendix 4.1.

Table 4.1 Comparison of power values from the exact power function and Monte Carlo simulations for $n=40$

θ^*	σ	Power	
		Exact power (upper bound on absolute error) by Equation (4.4)	Monte Carlo Simulations (10^6)
0	0.2	0.9988604 (1.8e-6)	0.9989
0.1	0.2	0.8552369 (1.8e-6)	0.8554
0.2	0.2	0.1278706 (3.4e-7)	0.1285
$log(1.25)$	0.2	0.0500 (7.3e-8)	0.0496
0	0.3	0.8950818 (6.1e-8)	0.8946
0.1	0.3	0.5617662 (1e-7)	0.5622
0.2	0.3	0.09578144 (7.4e-8)	0.0955
$log(1.25)$	0.3	0.04999948 (8.9e-5)	0.0499

4.4 Comparison of exact power with power values from graphs and approximate power

4.4.1 Comparison of exact power function with approximate power function of Chow and Wang

4.4.1.1 Approximate power formulas of Chow and Wang

Chow and Wang [42] focused on the log-normally distributed data which can be

modeled by the linear mixed effect model (3.2). When $\theta^* > 0$ and θ^* is large relative

to σ, Chow and Wang assumed that $P\left\{\dfrac{\hat{D}-\theta^*}{S\sqrt{1/n}} \leq t_\alpha(n-2)+\dfrac{\theta_1-\theta^*}{S\sqrt{1/n}}\right\} \approx 0$. Based on

the two one-sided tests procedure (4.1), Chow and Wang [42] approximated the

power function

$$P_1\left(\theta^*, n, \sigma\right) = P\left(\frac{\theta_1 - \theta^*}{S\sqrt{1/n}} + t_\alpha(v) \le \frac{\hat{D} - \theta^*}{S\sqrt{1/n}} \le \frac{\theta_2 - \theta^*}{S\sqrt{1/n}} - t_\alpha(v) \middle| \mu_T - \mu_R = \theta^*, \theta^* \in (\theta_1, \theta_2), \sigma\right)$$

by $P^{CW} = P\left\{\dfrac{\hat{D} - \theta^*}{S\sqrt{1/n}} \le \dfrac{\theta_2 - \theta^*}{S\sqrt{1/n}} - t_\alpha(v) \middle| \mu_T - \mu_R = \theta^*, \theta^* \in (\theta_1, \theta_2), \sigma\right\}$. Comparing the

right sides of the two formulas, we can easily see that Chow and Wang's power

function overestimates the exact power since $P\left\{\dfrac{\hat{D} - \theta^*}{S\sqrt{1/n}} \le t_\alpha(v) + \dfrac{\theta_1 - \theta^*}{S\sqrt{1/n}}\right\} \ge 0$.

Chow and Wang replaced S by the $\sigma\sqrt{2}$ in the right side of inequality of P^{CW}, so

they further approximated the power function (4.4) when $\theta^* > 0$ and θ^* is relatively

big by $P^{CW} \approx P\left\{\dfrac{\hat{D} - \theta^*}{S\sqrt{1/n}} \le \dfrac{\theta_2 - \theta^*}{\sigma\sqrt{2/n}} - t_\alpha(n-2) \middle| \mu_T - \mu_R = \theta^*, \theta^* \in (\theta_1, \theta_2), \sigma\right\}$.

(4.5)

Hence the sample size can be determined by

$$n \ge \frac{2\sigma^2\left(t_\alpha(n-2) + t_\beta(n-2)\right)^2}{\left(\theta_2 - \theta^*\right)^2} \text{ if } \theta^* > 0.$$

Chow and Wang used similar treatments for obtaining approximate power when

$\theta^* < 0$. Hence Chow and Wang concluded that the sample size can be determined by

$$n \ge \frac{2\sigma^2\left(t_\alpha(n-2) + t_\beta(n-2)\right)^2}{\left(\theta_2 - \left|\theta^*\right|\right)^2} \text{ if } \theta^* \neq 0 \quad (4.6).$$

When $\theta^* = 0$, the power function (4.4) is approximated by

$$P^{CW} = P\left\{ \left| \frac{\hat{D}}{S\sqrt{1/n}} \right| \leq \frac{\theta}{\sigma\sqrt{2/n}} - t_\alpha(v) \right\} \text{ if } \theta_2 = -\theta_1 = \theta. \text{ Hence the sample size is}$$

determined by

$$n \geq \frac{2\sigma^2 \left(t_\alpha(n-2) + t_{\beta/2}(n-2) \right)^2}{\theta_2^2} \text{ if } \theta^* = 0 \quad (4.7).$$

4.4.1.2 Comparison of exact power and approximate power of Chow and Wang

The approximate power function of Chow and Wang [42] tends to overestimate

power by removing the upper-tail bound on $\dfrac{\hat{D} - \theta^*}{S\sqrt{1/n}}$, but due to the further

replacement of sample standard deviation by true σ in the inequality for $\dfrac{\hat{D} - \theta^*}{S\sqrt{1/n}}$,

their approximation does not overestimate for all possible parameter combinations.

The differences between exact power and approximate powers are illustrated in

Figures 4.1 to 4.4 for wide ranges of standard deviations and true mean differences.

As defined in Section 4.2, σ^2 is the variance of log-transformed data from the

reference product. It is well known that the relationship between the coefficient of

variation (CV) in the untransformed data and the standard deviation (σ) of the log-

transformed is $CV = \sqrt{e^{\sigma^2} - 1}$.

When $\sigma =0.3$ for log-transformed data, $CV=0.307$ for untransformed data. It is well

known that the product is a highly variable drug when CV for untransformed AUC or

$Cmax$ is greater than 0.3 [52, 53]. Regardless of the magnitude of the standard

deviation σ from examination of many numerical results, Chow and Wang's

approximate power seems to over-estimate the exact power when both n and θ^* are

very small. Chow and Wang's power curve has a peak since Chow and Wang used two different formulas for power when θ^* is zero and nonzero. The difference between the Chow and Wang's approximate power and the exact power will decrease as the total sample size increases for same σ and n, and also decreases as the true difference (θ^*) increases. When σ increases from 0.2 to 0.7, there are more cases when Chow and Wang's approximate power exceeds the exact power.

Thus, sample sizes estimated by Chow and Wang's approximate power for the combination of small n and small θ^* are underestimated. When studies are underpowered, they may fail to meet the study objectives. For combination of large θ^* and small n, Chow and Wang's approximate power can underestimate the true power slightly, which results in having a few more subjects than necessary.

Previous authors [36, 37, 42] often focused on CV<0.3. However, errors of approximation may also be important for the large σ values of highly variable drugs.

From Figures 4.1 to 4.4 there are more combinations of θ^* and n for which Chow and Wang's approximate power over-estimates the exact power as σ increases from 0.2 to 0.7.

From Figure 4.1, it is seen that Chow and Wang's approximate power is very close to the exact power for $n \geq 18$ and $\theta^* > 0.04$ when $\sigma = 0.2$.

From Figure 4.2, it is seen that Chow and Wang's approximate power over-estimates the exact power for many combinations of θ^* and n when $\sigma = 0.3$. For example, Chow and Wang's approximate power is 81.77%, our exact power is 74.13%, and Chow and Wang's approximate power over-estimates the exact power by 7.64% for $n=30$ and $\theta^*=0.02$ when $\sigma = 0.3$.

From Figure 4.3, it is seen that Chow and Wang's approximate power over-estimates the exact power for many combinations of θ^* and n when $\sigma= 0.4$. For example, Chow and Wang's approximate power is 80.35%, our exact power is 71.63%, and Chow and Wang's approximate power over-estimates the exact power by about 9% for $n=50$ and $\theta^*=0.02$ when $\sigma= 0.4$.

From Figure 4.4, it is seen that Chow and Wang's approximate power over-estimates the exact power for many more combinations of θ^* and n when $\sigma= 0.7$ than those when $\sigma= 0.2$. For instance, Chow and Wang's approximate power over-estimate the exact power by 10% for $n<140$ and small θ^* when $\sigma= 0.7$.

Figure 4.1 Comparison of Chow and Wang's approximate power and exact power numerically calculated from Equation (4.4) against and true mean difference, θ^* when $\sigma=0.2$, at different total sample sizes, n
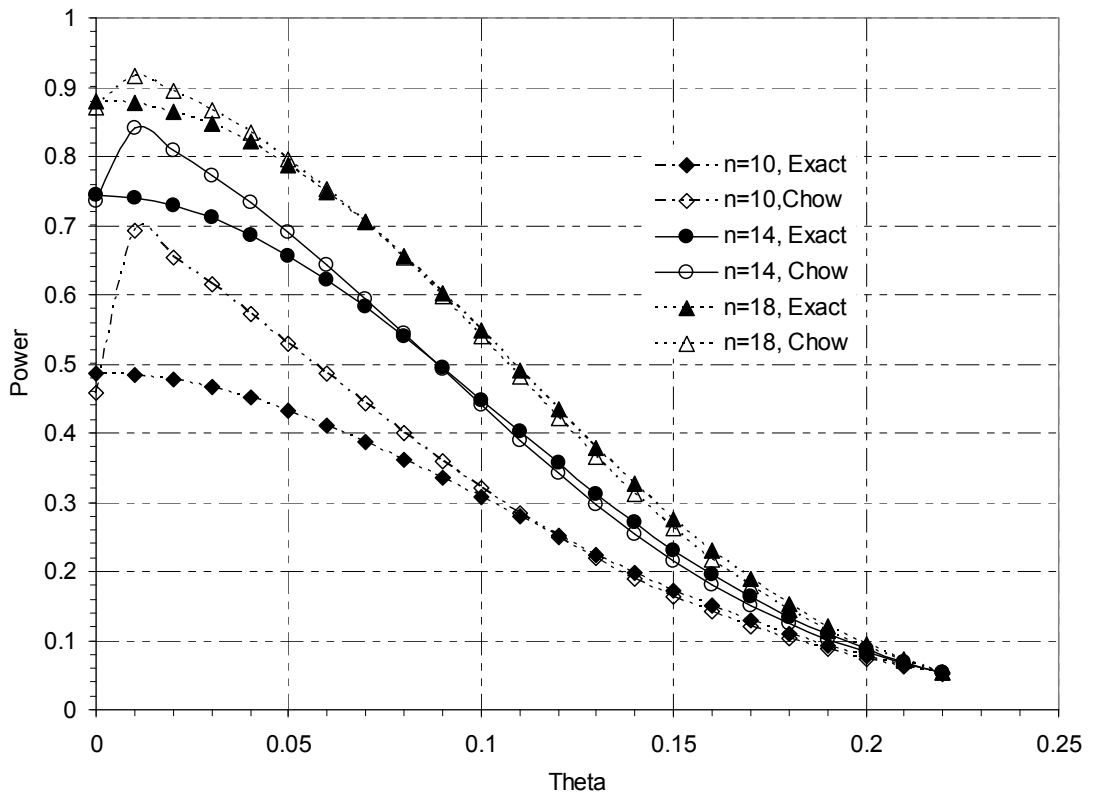
Figure 4.2 Comparison of Chow and Wang's approximate power and exact power numerically calculated from Equation (4.4) against and true mean difference, θ^* when σ=0.3, at different total sample sizes, n
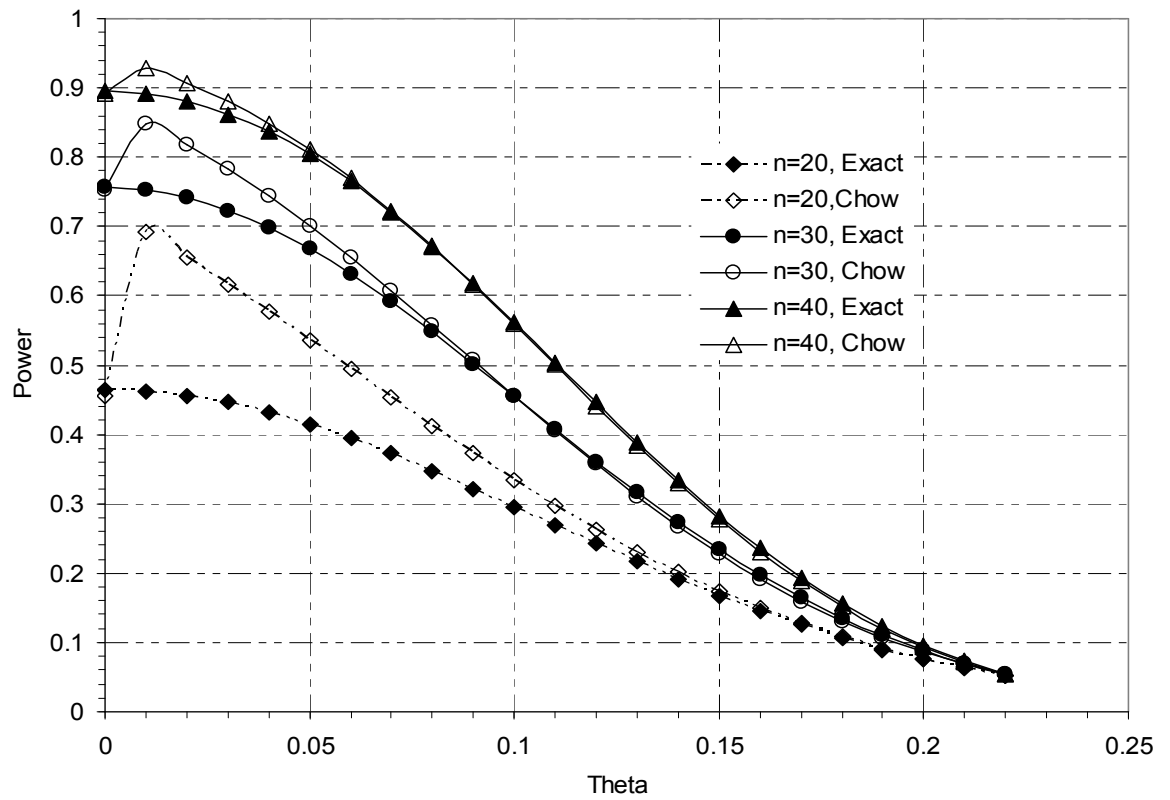
Figure 4.3 Comparison of Chow and Wang's approximate power and exact power numerically calculated from Equation (4.4) against and true mean difference, θ^* when σ=0.4, at different total sample sizes, n
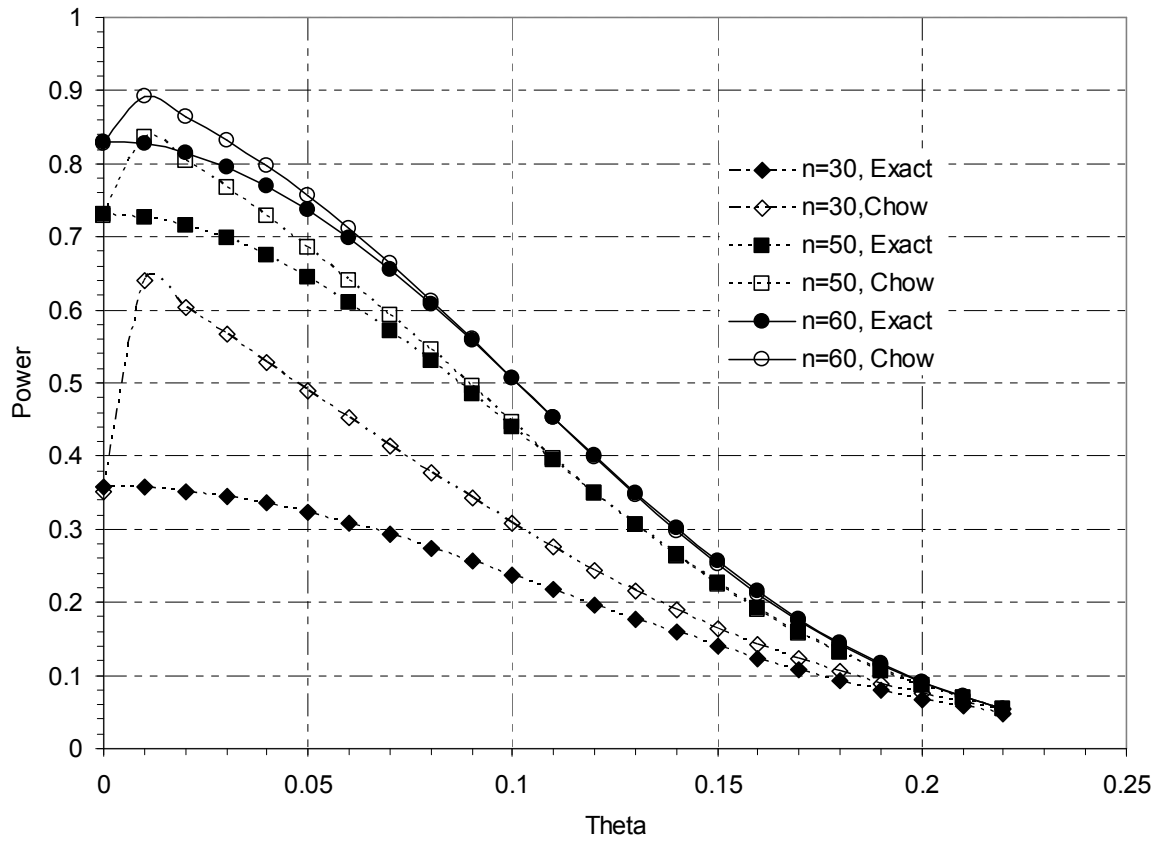
Figure 4.4 Comparison of Chow and Wang's approximate power and exact power numerically calculated from Equation (4.4) against and true mean difference, θ^* when $\sigma=0.7$, at different total sample sizes, *n*
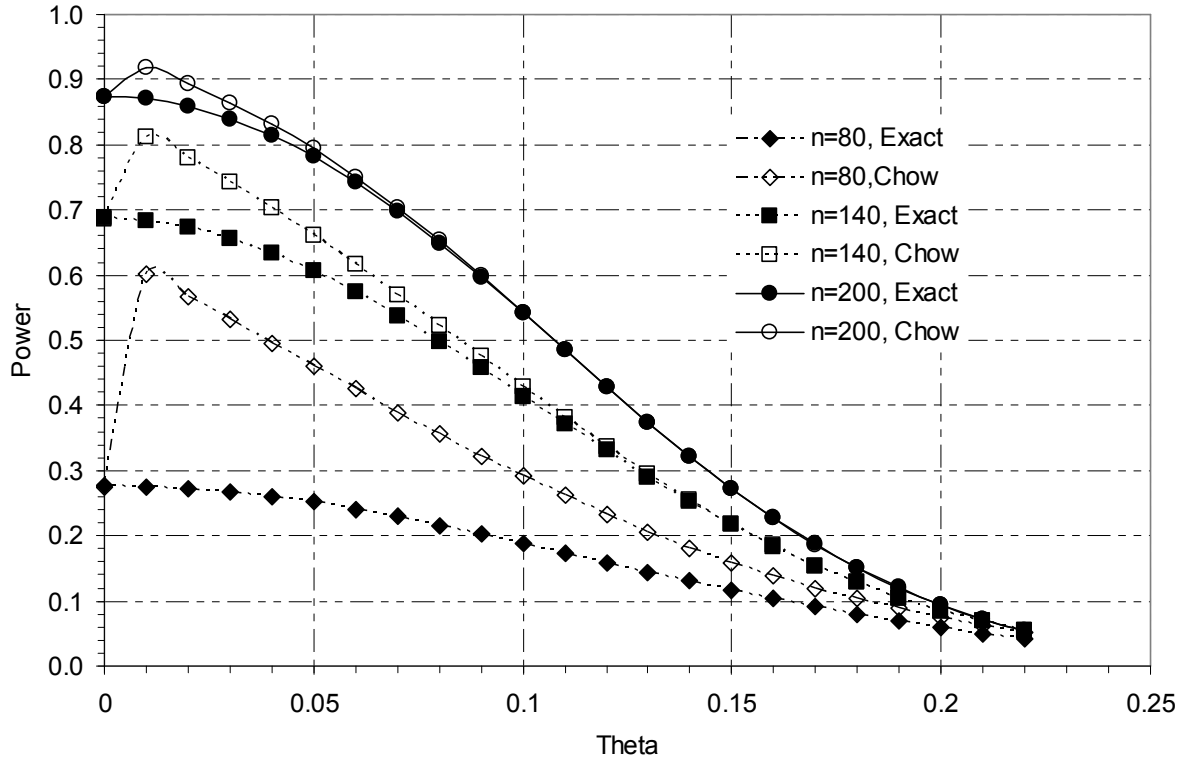
4.4.1.3 Comparison of exact sample size and approximation of Chow and Wang

In Table 4.2, we compare the total sample sizes from Equation (4.4) and Chow and Wang's approximate power [42] for achieving 80% power. The total sample size is rounded to next even number. Table 4.2 shows that the difference in the total sample sizes increases as σ increases, and decreases as θ^* increases, for each given θ^*. For example, the total sample size from the Chow-Wang approximate power is about 10% less than that from exact power when θ^*=0.03 and σ=0.2; about 15% less when θ^*=0.02 and σ=0.4; and about 20% less than that from exact power when θ^*=0.01 and σ=0.3.

Table 4.2 Comparison of sample size from exact power given by Equation (4.4) and Chow and Wang's approximate power for achieving 80% power at different combination of σ and θ^*

σ	θ^*=0.01		θ^*=0.02		θ^*=0.03		θ^*=0.04	
	Exact	Chow and Wang	Exact	Chow and Wang	Exact	Chow and Wang	Exact	Chow and Wang
0.1	6	6	6	6	6	6	6	6
0.2	16	14	16	14	18	16	18	18
0.3	34	28	34	30	36	32	38	36
0.4	58	46	60	50	62	56	66	62
0.5	90	70	92	78	94	86	100	94
0.6	128	100	130	110	136	122	144	136
0.7	172	136	176	150	184	164	194	184

4.4.2 Comparison of the exact power and Diletti et al.'s power

Since Diletti et al [37] did not provide the explicit mathematical formula for power calculation, we read three power values from Fig. 1c of [37]. The exact power values and power values from Fig. 1c of [37] when CV=20% and n=24 are listed below.

Table 4.3 Comparison of the exact power given by Equation (4.4) and power values read from Fig. 1c in Diletti et al's paper when CV=20% and n=24.

Method	$\theta^* = \mu_T - \mu_R$		
	0	Log(1.05)	Log(1.1)
Diletti et al's power	0.97	0.9	0.7
Exact power, $P_1(\theta^*, n, \sigma)$	0.9679	0.902	0.696

From Table 4.3, the power is very close to the exact power by Equation (4.4).

However, the graphs and tables present a limited number of cases. For instance,

Diletti et al. [37] did not provide any power value for CV>30%, as would be the case

with highly variable drugs. The exact and explicit power function allows such

situations to be considered.

4.4.3 Comparison of the exact power and Phillips' power

Since Phillips [36] did not provide the explicit mathematical formula for his power

calculation, we also read three power values from Fig. 3 of [36]. The parameters we

choose and comparison of the exact power values by Equation (4.4) and power values

from Fig. 3 in Phillips' paper under these parameters are listed in Table 4.4. In

Phillips's paper, θ_1 =-20% and θ_2 =20%. In order to compare the exact power,

$P_1(\theta^*, n, \sigma)$, with Phillips's power, θ_1 =-20% and θ_2 =20% are assumed in the

calculation of exact power function, $P_1(\theta^*, n, \sigma)$.

Table 4.4 Comparison of the exact power given by Equation (4.4) and power values read from Fig. 3 in Phillips' paper when CV=20% and n=24.

Method	$\theta^* = \mu_T - \mu_R$		
	0.015	0.05	0.1
Phillips' power	0.9	0.8	0.5
Exact power, $P_1(\theta^*, n, \sigma)$	0.909	0.809	0.517

From Table 4.4, Phillips' power is very close to the exact power by Equation (4.4). However, the graphs and tables present a limited number of cases as Phillips [36] did. The exact and explicit power function, such as P_1 in equations (4.3)-(4.4), allows any situation to be considered.

4.5 Discussion and Conclusions

The exact power can be derived from the joint density function of two highly correlated test statistics (T_1 and T_2). Our derivation for the exact power under the general parameters just serves for completeness and for expository purposes since the explicit power formula under the equal variance for the test and reference product and a balanced design was published in [43, 44]. The exact power numerically integrated from Equation (4.4) is corroborated by the results from 10^6 Monte Carlo simulations. Since suitable free software, such as R software, is available, exact power for the two one-sided tests procedure can be readily obtained from the power function by numerical solutions. Our R code for power function is attached in the Appendix 4.1. Exact sample size calculation is then easy using the exact power for any parameter combinations in bioequivalence studies based on two one-sided tests. This chapter is to remind the readers that approximate methods are still in use [46, 47]. The exact power will pave the way for sample size calculation for any parameters' combination (e.g., unequal variance of the test and reference products) in bioequivalence study. The fact that Chow and Wang's approximate power markedly over-estimates the exact power for many combinations of θ^* and n as shown in Figures 4.1 to 4.4 and Table 4.1 demonstrates the preferability of the exact power function $P_1(\theta^*, n, \sigma)$ in planning of bioequivalence study, especially when θ^* is small relative to σ. Mean

differences in log(*AUC*) or log(*Cmax*) as small as 0.03 are of practical interest based on the review by Davit et al. [54] of 12 years of bioequivalence studies from the FDA, which found that more than 50% of studies have mean difference less than of 0.05 between generic product and innovator. (These are mean differences in log(*AUC*) or log(*Cmax*), when the corresponding mean differences in *AUC* and *Cmax* between generic and innovator products are respectively 3.56% and 4.35%.)

While the exact power agrees closely with [36, 37] in the few cases displayed in those papers and the explicit formula under the equal variance for the test and reference product and balanced design was published in [43, 44], it is still important to have exact powers and sample sizes in all bioequivalence study settings, including those for unequal variance of the test and reference products and to show just how far off the approximate formulas that are still in use [46, 47] can be for many practical settings.

Appendix 4.1: R code for exact power and sample size

The power $P_1(\theta^*, n_1, n_2, \sigma_T, \sigma_R)$ in (4.3) is calculated by the following lines of R code. Here the arguments n1 and n2 respectively represent the sample sizes n_1, n_2 of sequences 1 and 2; sigT , sigR represent σ_T and σ_R; th1 and th2 respectively represent θ_1 and θ_2; and tstar represents θ^*. In fact, $(n_1, n_2, \sigma_T, \sigma_R)$ enters formula (4.3) only through the quantities $\sigma_{\hat{D}}^2 = (\sigma_R^2 + \sigma_T^2)(1/n_1 + 1/n_2)/4$ and $v = n_1 + n_2 - 2$, and arguments sigD and nu represent $\sigma_{\hat{D}}$ and v. Pow1$value is the power in formula (4.3) (with Pow1$abs.error the estimated absolute error of integration), and SSiz denotes the sample size obtained by equating formula (6) to 1-

beta. We use two standard R functions: **integrate** to perform univariate numerical integration, and **uniroot** to perform root-finding or inversion.

```
Pow1 = function(sigD, nu, th2,th1,tstar, alpha=.05){

    Integrand = function(x)    {

        (pnorm((th2-tstar)/sigD- qt(1-alpha, nu)*sqrt(x/nu))-

        pnorm((th1-tstar)/sigD+ qt(1-alpha, nu)*sqrt(x/nu)))*dchisq(x,nu)

    }

    integrate(Integrand, lower = 0, upper =

        nu*((th2-th1)/(2*qt(1-alpha, nu)*sigD))^2)[1:2]

}


    SSiz = function(beta, sigT,sigR,th2,th1,tstar, alpha=.05, uppern=300){

        Pow2 = function(n) {

            Pow1(sqrt((sigT^2+sigR^2)/n), n-2, th2, th1,  tstar, alpha)$value-1+beta }

        uniroot(Pow2,c(3,uppern))$root

    }
```

Appendix 4.2: Proof of monotonicity of $\sqrt{v}/t_\alpha(v)$

Define $t(v)/\sqrt{v} = Z/\sqrt{Y(v)}$, where Z is the standard normal random variable and $Y(v)$ is the chi-squared random variable with v degrees of freedom. Since

$$Y(v) = \sum_{i=1}^{v} Z_i^2,$$ where $Z_1, Z_2,..., Z_v$ are independent and identical standard normal random variables, then obviously $Y(v+1)$ is stochastically larger than $Y(v)$, which

means for any real value x, $P(Y(v+1) \geq x) > P(Y(v) \geq x)$. If a function $f(z,y)$ is

monotone decreasing in its second argument and Z is a random variable independent

of (Y_1, Y_2) with Y_1 stochastically larger than Y_2, then $f(Z, Y_2)$ is stochastically larger

than $f(Z, Y_1)$. From this it follows that $t(v+1)/\sqrt{v+1}$ is stochastically smaller than

$t(v)/\sqrt{v}$, from which it follows that. $t_\alpha(v+1)/\sqrt{v+1} < t_\alpha(v)/\sqrt{v}$. Therefore,

$\sqrt{v}/t_\alpha(v)$ is a strictly increasing function of v.

Chapter 5 Two-stage sample size re-estimation for bioequivalence crossover studies

5.1 Introduction

As previously discussed in Chapter 3, the goal of a bioequivalence study is to establish that two product formulations (T for the test product and R for the reference product) are sufficiently similar with respect to the mean bioavailability (*log(AUC)* or *log(Cmax)*). In this setting, we should reject the null hypothesis in the following hypothesis test (5.1) so as to conclude the bioequivalence of the test product and the reference product:

$$H_0 : \mu_T - \mu_R \leq \theta_1 \text{ or } \mu_T - \mu_R \geq \theta_2$$
$$H_a : \theta_1 < \mu_T - \mu_R < \theta_2 \qquad\qquad (5.1)$$

In this chapter as in previous chapters, we assume the basic study involves a crossover design in which each patient is exposed to each formulation, though the order of which formulation comes first is randomized. The sample size calculation to achieve a specific power is described in Equation (4.3) of Chapter 4 for a single planned stage. In general planning these studies requires specification of the assumed mean difference between the test and the reference products and variances of the test and reference products. Misspecification of either mean difference or variances can result in an under-powered or over-powered trial. Both the assumed mean difference between the test product and the reference product and variances of the test and reference products used in sample size calculations often come from previous trials in

which some factors could be different from those in the current study. So sample size

re-estimation can be important in improving the chances for success in a study. Most

sample size re-estimation procedures for superiority clinical trials [55, 56] and for

bioequivalence crossover trials [46, 47] are based on the updated variance estimates

from an internal pilot study (Stage 1) since naturally Stage 1, a subset of the current

trial's data, is more similar to the current data than any historical data. Stein's classic

two-stage design [55] assumed that the number of subjects in Stage 2 ought to

depend on the variance estimated from Stage 1 and Stein's procedure [55] used only

the Stage 1 variance estimate S_1^2 in the final statistic. This provided a guarantee of

being able to control the type I error rate while the type II error rate would be no

higher than a specified level irrespective of the true variance. Wittes et al [56] used

the Naïve t for the combined data from Stage 1 and Stage 2 in which the number of

subjects depended on the unblinded variance estimate from Stage 1 and they found

out that the type I error rate was inflated. Potvin et al. [46] considered a simple naïve

sample size re-estimation method using the nominal α (e.g., 0.05) at whichever stage

the test is carried out and using data from both stages (if two stages are carried out) to

compute the final variance estimate for two-period two-treatment crossover

bioequivalence trials. Their limited simulations [46] quickly confirmed that the

overall type I error rate can be much larger than the nominal level α. To investigate

the type I error of the unmodified t-test after a blinded sample size revision in

crossover bioequivalence trials, Golkowski et al. [47] defined $Z_1, V_1, Z_2,$ and V_2^* as

96

$$Z_i = \sqrt{n_{i\bullet}} \, \frac{\hat{D}_i - \theta_1}{\sqrt{\sigma_T^2 + \sigma_R^2}}, \; i=1,2, \; V_1 = \frac{(n_{1\bullet} - 2)S_1^2}{\sigma_T^2 + \sigma_R^2}, \; \text{and}$$

$$V_2^* = \frac{(n_{2\bullet} - 2)S_2^2}{\sigma_T^2 + \sigma_R^2} + \frac{n_{1\bullet} n_{2\bullet}}{n_{1\bullet} + n_{2\bullet}} \left[\left(\frac{\bar{d}_{11} - \bar{d}_{21}}{\sqrt{\sigma_T^2 + \sigma_R^2}} \right)^2 + \left(\frac{\bar{d}_{12} - \bar{d}_{22}}{\sqrt{\sigma_T^2 + \sigma_R^2}} \right)^2 \right].$$

Here n_{li} is the sample size of the ith sequence at the lth stage $\forall i, l = 1,2$,

$n_{l\bullet} = n_{l1} + n_{l2}$ is the total sample size at the lth stage, $\hat{D}_l$ is the average of the intra-subject difference between the test and the reference over subjects and crossover

sequences in Stage l, S_l^2 is a pooled estimate of $\sigma_T^2 + \sigma_R^2$, $\bar{d}_{1i} \left(= \sum_{k=1}^{n_{1i}} (Y_{i2k} - Y_{i1k}) \middle/ n_{1i} \right)$ is

the mean period difference of Sequence i at Stage 1, and $\bar{d}_{2i} \left(= \sum_{k=n_{1i}+1}^{n_{1i}+n_{2i}} (Y_{i2k} - Y_{i1k}) \middle/ n_{2i} \right)$

is the mean period difference of Sequence i at Stage 2. Golkowski et al. proposed to

re-estimate $n_{2\bullet}$ with the following formulas in which $\sigma_T^2 + \sigma_R^2$ was replaced by the

blinded variance estimate $\hat{\sigma}_{OS}^2$:

$$n_{2\bullet} = \max \left[2 \frac{(t_{1-\alpha}(n_{1\bullet}) + t_{1-\beta}(n_{1\bullet}))^2}{(\theta_0 - \theta_2)^2} \hat{\sigma}_{OS}^2 - n_{1\bullet}, 0 \right] \text{ for } \theta_0 > 0 \text{ and}$$

$$n_{2\bullet} = \max \left[2 \frac{(t_{1-\alpha}(n_{1\bullet}) + t_{1-\beta/2}(n_{1\bullet}))^2}{\theta_2^2} \hat{\sigma}_{OS}^2 - n_{1\bullet}, 0 \right] \text{ for } \theta_0 = 0.$$

Here $\hat{\sigma}_{OS}^2 = \frac{1}{2(n_{1\bullet} - 1)} \sum_{i=1}^{2} \sum_{k=1}^{n_{1i}} (d_{1ik} - \bar{d}_1)^2$, θ_0 (the assumed mean difference) is in practice

assumed to lie between $\log(0.9)$ and $\log(1.1)$, θ_2 (the equivalence margin) equals $\log(1.25)$,

d_{1ik} is the difference between Period 2 and Period 1 for the kth subject in the ith

sequence at Stage 1, and $\bar{d}_1$ is the mean of the period differences of two sequences at Stage 1. Golkowski et al. [47] calculated the t-test statistic from the combined data of Stage 1 and Stage 2 by $\left(\dfrac{\sqrt{n_{1\bullet}/(n_{1\bullet}+n_{2\bullet})}Z_1 + \sqrt{n_{2\bullet}/(n_{1\bullet}+n_{2\bullet})}Z_2}{\sqrt{(V_1+V_2^*)/(n_{1\bullet}+n_{2\bullet}-2)}} \right)$, a function of four components $Z_1, V_1, Z_2,$ and V_2^* whose distributions were discussed in detail on page 1003 of [57]. The type I error rate $\alpha_{act}^{H_{01}^D}$ for which they claimed as the actual type I error rate under the null hypothesis $H_{01}^D : \mu_T - \mu_R \leq \delta_1^D$ (see pages 1004 and 1005 of [45]) is an approximation, not the exact value due to the following two reasons: 1) The distribution of V_2^* given (Z_1, V_1) in [57] is approximately Chi-squared; 2) The boundaries of the integrations over the rejection region for calculating the maximum type I error rate are unclearly specified. Golkowski et al. [47] showed that their proposed sample size re-estimation procedure, although blinded, can lead to some inflation of the type I error rate. Furthermore, Golkowski et al. [47] proposed to adjust the actual type I error rate such that for the particular sample size of the internal pilot study n_1, the maximum type I error rate falls below the nominal α. The actual type I error rate for bioequivalence in Golkowski et al. [47] is approximate (possibly masking a minor type I error level inflation) since the distribution that they use for V_2^* is an approximation due to the fact that $\dfrac{(n_{2\bullet}-2)S_2^2}{\sigma_T^2 + \sigma_R^2}$ and

$$\left[\left(\frac{\bar{d}_{11}-\bar{d}_{21}}{\sqrt{\sigma_T^2+\sigma_R^2}}\right)^2,\left(\frac{\bar{d}_{12}-\bar{d}_{22}}{\sqrt{\sigma_T^2+\sigma_R^2}}\right)^2\right]$$ are still independent, but $\left(\frac{\bar{d}_{11}-\bar{d}_{21}}{\sqrt{\sigma_T^2+\sigma_R^2}}\right)^2$ and

$$\left(\frac{\bar{d}_{12}-\bar{d}_{22}}{\sqrt{\sigma_T^2+\sigma_R^2}}\right)^2$$ are no longer independent conditioning on both V_1 and $\hat{D}_1$ [58].

Let $X=(n_{1\bullet}-2)S_1^2$. The new sample size $(n_{1\bullet}+n_{2\bullet}(X))$ is a random variable

because it is calculated using S_1^2 (the estimated variance) from Stage 1.Therefore

$$T_{21}=(\hat{D}-\theta_1)\Bigg/\left(\frac{S}{2}\sqrt{\frac{1}{n_{11}+n_{21}(X)}+\frac{1}{n_{12}+n_{22}(X)}}\right) \text{ and}$$

$$T_{22}=(\hat{D}-\theta_2)\Bigg/\left(\frac{S}{2}\sqrt{\frac{1}{n_{11}+n_{21}(X)}+\frac{1}{n_{12}+n_{22}(X)}}\right) \text{ will not follow the standard } t\text{-}$$

distribution under the null hypothesis. Here $\hat{D}$ and S are respectively the average of

the intra-subject difference between the test and the reference over subjects and

crossover sequences and the pooled estimate of the variance of an intra-subject

difference from the combined data of both Stage 1 and Stage 2. The quantities

$n_{11}+n_{21}(X)$ and $n_{12}+n_{22}(X)$ are respectively the number of subjects in Sequence 1

and Sequence 2 of the combined data of Stage 1 and Stage 2. Consequently the type I

error rate in sample size re-estimation [46] cannot be controlled at the nominal level

α if the naïve $t_\alpha(n_{1\bullet}+n_{2\bullet}(X)-2)$ is used when combining data of two stages.

In the literature and in practice, the point estimate for the variance obtained at stage 1

is often used for sample size re-estimation. Following this practice, we propose that

the number of subjects added to the second stage should be based on the point

estimate of the variance from the first stage. Inspired by Gould and Shih's idea of

using the pooled variance of Stage 1 and Stage 2 in the final statistics [59], we

propose to replace S in T_{21} and T_{22}, by S^* which is equal to $\sqrt{\dfrac{X+(n_{2\bullet}(X)-2)S_2^2}{n_{1\bullet}+n_{2\bullet}(X)-4}}$ so

that the exact critical value for the two-stage study can be derived analytically. Two

new test statistics for the hypothesis testing based on the combined data of Stage 1

and Stage 2 are $T_{21}^* = (\hat{D}-\theta_1)\Big/\left(\dfrac{S^*}{2}\sqrt{\dfrac{1}{n_{11}+n_{21}(X)}+\dfrac{1}{n_{12}+n_{22}(X)}}\right)$ and

$T_{22}^* = (\hat{D}-\theta_2)\Big/\left(\dfrac{S^*}{2}\sqrt{\dfrac{1}{n_{11}+n_{21}(X)}+\dfrac{1}{n_{12}+n_{22}(X)}}\right).$

We will compare T_{22} and T_{22}^* in Section 5.7 and provide justification for the use of

T_{22}^*.

Let S_0^2 be the initial value for $\sigma_T^2 + \sigma_R^2$ from the historical data. Throughout this

chapter, we assume that S_0^2 is a constant, not a random variable. Let θ^* be the true

mean difference between the test product and the reference product ($\mu_T - \mu_R$). We

also let θ_0 be the assumed mean difference between the test product and the reference

product for power calculation. Since the probability of rejecting H_0 under H_a in

Equation (5.1) when we assume that S_0^2 is the true variance ($\sigma_T^2 + \sigma_R^2$) and $\theta^* = \theta_0$ is

p_0 (targeted power), $n_{1\bullet}$ is calculated for any given allocation ratio of two sequences

($e.g., n_{11}/n_{12}=1$) by solving the following equation:

$$p_0 = \int_0^{c_0} \Phi\left(\frac{\theta_2 - \theta_0}{\sqrt{\dfrac{S_0^2}{4}\left(\dfrac{1}{n_{11}} + \dfrac{1}{n_{12}}\right)}} - t_\alpha(v_1)\sqrt{\dfrac{x}{v_1}} \right) - \Phi\left(\frac{\theta_1 - \theta_0}{\sqrt{\dfrac{S_0^2}{4}\left(\dfrac{1}{n_{11}} + \dfrac{1}{n_{12}}\right)}} + t_\alpha(v_1)\sqrt{\dfrac{x}{v_1}} \right) . \quad (5.2)$$

$$\frac{1}{2^{v_0/2}\Gamma(v_1/2)} x^{v_1/2-1} e^{-x/2} dx$$

Here $c_0 = \dfrac{(\theta_2 - \theta_1)^2 v_1}{t_\alpha^2(v_1)S_0^2}\left(\dfrac{1}{n_{11}} + \dfrac{1}{n_{12}}\right)^{-1}$ and $v_1 = n_{11} + n_{12} - 2$.

If $X \le (n_{1\bullet} - 2)S_0^2$, then we stop at the end of the first stage. We use the following

test statistics $T_{11} = \left(\hat{D}_1 - \theta_1\right) \Big/ \left(\dfrac{S_1}{2}\sqrt{\dfrac{1}{n_{11}} + \dfrac{1}{n_{12}}}\right)$ and $T_{12} = \left(\hat{D}_1 - \theta_2\right) \Big/ \left(\dfrac{S_1}{2}\sqrt{\dfrac{1}{n_{11}} + \dfrac{1}{n_{12}}}\right)$ to

test the hypothesis in Equation (5.1) at Stage 1. Thus the data from the first stage is

analyzed by comparing T_{11} and T_{12} to the critical value $t_\alpha(v_1)$ and $-t_\alpha(v_1)$,

respectively.

To simplify the notation for $\sigma_T^2 + \sigma_R^2$, we define $V = \sqrt{\sigma_T^2 + \sigma_R^2}$.

Let $P_{1s}(\theta_2, n_{11}, n_{12}, V)$ be the joint probability of rejecting H_0 in Equation (5.1) under

H_0 for one-stage study and $X \le (n_{1\bullet} - 2)S_0^2$. $P_{1s}(\theta_2, n_{11}, n_{12}, V)$ is calculated by (5.3) in

which $\theta^* = \theta_2$. The function $P_{1s}(\theta^*, n_{11}, n_{12}, V)$ is expressed as

$$P_{1s}(\theta^*, n_{11}, n_{12}, V)$$
$$= \int_0^{c_1} \Phi\left(2\frac{\theta_2 - \theta^*}{V}\left(\frac{1}{n_{11}} + \frac{1}{n_{12}}\right)^{-1/2} - \frac{t_\alpha(v_1)}{V}\sqrt{\frac{x}{v_1}} \right) - \Phi\left(2\frac{\theta_1 - \theta^*}{V}\left(\frac{1}{n_{11}} + \frac{1}{n_{12}}\right)^{-1/2} + \frac{t_\alpha(v_1)}{V}\sqrt{\frac{x}{v_1}} \right) .$$
$$\frac{1}{2^{v_1/2}\Gamma(v_1/2)V^2}\left(\frac{x}{V^2}\right)^{v_1/2-1} \exp\left(-x/(2V^2)\right)dx$$

$$(5.3)$$

101

Here $c_1 = \min\left(V_1 S_0^2, \dfrac{(\theta_2 - \theta_1)^2 V_1}{t_\alpha^2(V_1)}\left(\dfrac{1}{n_{11}} + \dfrac{1}{n_{12}} \right)^{-1} \right)$ and $V_1 = n_{11} + n_{12} - 2$.

α is the significance level for Stage 1 analysis, and $V_1 = n_{11} + n_{12} - 2$. See its

derivation in Section 5.5.1.

If $X > (n_{1\bullet} - 2)S_0^2$, then $n_{1\bullet} + n_{2\bullet}(X)$ is calculated to assure that the combined study

(using Stages 1 and 2) will have the targeted power p_0. From this point forward,

$n_{2\bullet}(X)$ is a function of X. This assumes that S_1^2 is V^2 and θ_0 is θ^* for any given

allocation ratio of two sequences ($e.g., (n_{11} + n_{21}(X))/(n_{12} + n_{22}(X)) = 1$) and we solve

the following equation for $n_{2\bullet}(X)$: if $X > (n_{1\bullet} - 2)S_0^2$, then

$$p_0 = \int_0^c \left(\Phi\left(\dfrac{\theta_2 - \theta_0}{\hat{\sigma}_1^*(X)} - t_\alpha(V)\sqrt{\dfrac{w}{V}} \right) - \Phi\left(\dfrac{\theta_1 - \theta_0}{\hat{\sigma}_1^*(X)} + t_\alpha(V)\sqrt{\dfrac{w}{V}} \right) \right) \dfrac{1}{2^{V/2}\Gamma(V/2)} (w)^{V/2 - 1} e^{-w/2} dw .$$

(5.4)

Here $c(X) = \dfrac{(\theta_2 - \theta_1)^2 V_1 V}{t_\alpha^2(V)X}\left(\dfrac{1}{n_{11} + n_{21}(X)} + \dfrac{1}{n_{12} + n_{22}(X)} \right)^{-1}$,

$\hat{\sigma}_1^*(X) = \sqrt{\dfrac{X}{4V_1}\left(\dfrac{1}{n_{11} + n_{21}(X)} + \dfrac{1}{n_{12} + n_{22}(X)} \right)}$, and $V = n_{1\bullet} + n_{2\bullet}(X) - 2$.

The solution of Equation (5.4) for V is unique (see Proposition 5 in Section 5.5.3).

Clearly, $n_{2\bullet}(X)$ is a continuous-valued function implicitly defined in Equation (5.4).

Actually in practice $n_{2\bullet}(X)$ is rounded up to next even number. This is a naïve and

approximate way to obtain $n_{1\bullet} + n_{2\bullet}(X)$ since this formula pretends that the first stage

sample variance was not observed and that the second stage sample size was fixed in advance at $n_{2\bullet}(X)$.

To illustrate the idea, we assume that $\theta_0 = 0.05$, $p_0 = 0.9$, $\theta_2 = -\theta_1 = \log(1.25)$, $\alpha = 0.05$, $n_{11} = n_{12}$, and $n_{21}(X) = n_{22}(X)$ throughout this chapter. From Equation (5.4), it can be seen that $n_{2\bullet}$ is a function of X.

Let $P_{2s}(\theta_2, n_{11}, n_{12}, V, p_0, u)$ be the joint probability of rejecting H_0 in Equation (5.1) under H_0 for two-stage study ($n_{2\bullet}(X) > 0$) and $X > (n_{1\bullet} - 2)S_0^2$. Then $P_{2s}(\theta^*, n_{11}, n_{12}, V, p_0, u)$ is expressed in terms of the function $n_{2\bullet} = n_{2\bullet}(X)$ as

$$
\begin{aligned}
&P_{2s}(\theta^*, n_{11}, n_{12}, V, p_0, u) \\
&= \int_b^\infty \int_0^a \left[\Phi\left(\frac{\theta_2 - \theta^* - uc_3(x)\sqrt{x+y}}{c_4(x)V} \right) - \Phi\left(\frac{\theta_1 - \theta^* + uc_3(x)\sqrt{x+y}}{c_4(x)V} \right) \right] \cdot \\
&\quad \frac{1}{2^{(\nu_1+\nu_2)/2}\Gamma(\nu_1/2)\Gamma(\nu_2/2) \cdot V^{\nu_1+\nu_2}} y^{\nu_2/2-1} x^{\nu_1/2-1} \exp\left(-\frac{x+y}{2V^2} \right) dy\, dx
\end{aligned}
\tag{5.5}
$$

See its derivation in Section 5.5.2. Here u is the critical value.

$$
a = a(x) = (n_{1\bullet} + n_{2\bullet}(x) - 4)c_2^2(x) - x.
\tag{5.6}
$$

$$
b = (n_{1\bullet} - 2)S_0^2.
\tag{5.7}
$$

$$
c_2(x) = \frac{\theta_2 - \theta_1}{u} \left(\sqrt{\frac{1}{n_{11} + n_{21}(x)} + \frac{1}{n_{12} + n_{22}(x)}} \right)^{-1}.
\tag{5.8}
$$

$$
c_3(x) = \frac{1}{2} \sqrt{\left(\frac{1}{n_{11} + n_{21}(x)} + \frac{1}{n_{12} + n_{22}(x)} \right) \frac{1}{n_{1\bullet} + n_{2\bullet}(x) - 4}}.
\tag{5.9}
$$

103

$$c_4(x) = \sqrt{\frac{1}{4(n_{1\bullet} + n_{2\bullet}(x))^2} \left\{ n_{1\bullet}^2 \left(\frac{1}{n_{11}} + \frac{1}{n_{12}} \right) + n_{2\bullet}^2(x) \left(\frac{1}{n_{21}(x)} + \frac{1}{n_{22}(x)} \right) \right\}} . \quad (5.10)$$

If $X > (n_{1\bullet} - 2)S_0^2$, we proceed collecting the data from additional $n_{2\bullet}(X)$ subjects and then analyze the combined data from both Stage 1 and Stage 2 using the test statistics: T_{21}^* and T_{22}^* compared to critical values u_α and $-u_\alpha$ respectively. The exact critical value, u_α, is derived as the largest value of u for which the condition:

$\max\limits_{V}[P_{1s}(\theta_2, n_{11}, n_{12}, V) + P_{2s}(\theta_2, n_{11}, n_{12}, V, p_0, u_\alpha)] \leq \alpha$ holds. Therefore u_α is not a function of V or X. This property ensures the experimentwise type I error rate in the sample size re-estimation procedure.

In Section 5.2, we will next apply the linear mixed effect model (3.2) defined in Chapter 3 to both Stage 1 and Stage 2 and derive the distributions of $\hat{D}_1$ and $\hat{D}_2$ given X. In Section 5.3, we present the test statistics $(T_{21}$ and $T_{22})$ based on our two-stage sample size re-estimation procedure and discuss the challenge of controlling the type I error rate with these statistics. In order to develop the analytical solution of type I error rate, we propose two simpler test statistics $(T_{21}^*$ and $T_{22}^*)$. In Section 5.4, we describe the overall strategy for assuring that our proposed unblinded sample size re-estimation procedure has an exact α experimentwise type I error rate. In Section 5.5, we derive the exact power function $P_{2s}(\theta^*, n_{11}, n_{12}, V, p_0, u)$ of T_{21}^* and T_{22}^* for the two-stage study and the exact power function $P_{1s}(\theta^*, n_{11}, n_{12}, V)$ for a one-stage study. In Section 5.6, we derive the exact critical value u_α for T_{21}^* (and $-u_\alpha$ for T_{22}^*) for the two-stage study by assuring the supremum of experimentwise probability of rejecting

H_0 under H_0 for the two-stage study over the whole range of true variances is less than or equal to α. The critical value $t_\alpha(v_1)$ will be used for the one-stage study. In Section 5.7, we will show that our proposed test statistic T_{22}^* performs much like T_{22} if the difference in the period effect between Period 2 and Period 1 is similar for both stages. Specifically, we compare the null and alternative distributions of T_{22} and T_{22}^* for two-stage study.

5.2 Linear mixed effect model

Recall that Y_{ijk} is the response of the kth subject in the jth period of the ith sequence in the 2-period 2-treatment crossover study, here $i=1, 2, j=1, 2$ and $k=1,..,n_{1i}$ for Stage 1 and $k=n_{1i}+1,.., n_{1i}+n_{2i}(X)$ for Stage 2. Then as in Section 3.2, Y_{ijk} is modeled by the linear mixed effect model (3.2).

For Stage 1, $\hat{D}_1 = \dfrac{1}{2}\left(\sum_{k=1}^{n_{11}}(Y_{12k}-Y_{11k})\Big/n_{11} + \sum_{k=1}^{n_{12}}(Y_{21k}-Y_{22k})\Big/n_{12} \right)$ and

$$S_1^2 = \frac{1}{n_{11}+n_{12}-2}\left\{ \sum_{k=1}^{n_{11}}(Y_{12k}-Y_{11k}-\sum_{k=1}^{n_{11}}(Y_{12k}-Y_{11k})\Big/n_{11})^2 + \sum_{k=1}^{n_{12}}\left[Y_{21k}-Y_{22k}-\sum_{k=1}^{n_{12}}(Y_{21k}-Y_{22k})\Big/n_{12} \right]^2 \right\}.$$

From the derivations in Section 3.2, $\hat{D}_1$ is a normally distributed unbiased estimate of θ^* with variance $\sigma_{\hat{D}_1}^2 = \dfrac{V^2}{4}\left(\dfrac{1}{n_{11}} + \dfrac{1}{n_{12}} \right)$, and $\dfrac{X}{V^2}$ is distributed as χ^2 with the degrees of freedom $n_{11}+n_{12}-2$.

For Stage 2, $\hat{D}_2 = \dfrac{1}{2}\left(\displaystyle\sum_{k=n_{11}+1}^{n_{11}+n_{21}(X)}(Y_{12k}-Y_{11k})\Big/n_{21}(X) + \sum_{k=n_{12}+1}^{n_{12}+n_{22}(X)}(Y_{21k}-Y_{22k})\Big/n_{22}(X)\right)$ and

$$S_2^2 = \frac{1}{n_{21}(X)+n_{22}(X)-2}$$
$$\left\{\sum_{k=n_{11}+1}^{n_{11}+n_{21}(X)}\left[Y_{12k}-Y_{11k}-\sum_{k=n_{11}+1}^{n_{11}+n_{21}(X)}(Y_{12k}-Y_{11k})\Big/n_{21}(X)\right]^2 + \sum_{k=n_{12}+1}^{n_{12}+n_{22}(X)}\left[Y_{21k}-Y_{22k}-\sum_{k=n_{12}+1}^{n_{12}+n_{22}(X)}(Y_{21k}-Y_{22k})\Big/n_{22}(X)\right]^2\right\}.$$

For Stage 2, given sample sizes $n_{21}(X)$ and $n_{22}(X)$, $\hat{D}_2$ is a normally distributed

unbiased estimate of θ^* with variance $\sigma^2_{\hat{D}_2|n_{21}(x),n_{22}(x)} = \dfrac{V^2}{4}\left(\dfrac{1}{n_{21}(X)}+\dfrac{1}{n_{22}(X)}\right)$, and

conditional distribution of $\dfrac{(n_{21}(X)+n_{22}(X)-2)\bullet S_2^2}{V^2}$ given X is χ^2 with the degrees

of freedom $n_{21}(X)+n_{22}(X)-2$. Hence the distributions of $\hat{D}_2$ and of

$\dfrac{(n_{21}(X)+n_{22}(X)-2)S_2^2}{V^2}$ can be respectively obtained by averaging the conditional

cumulative distribution function over X with respect to its density.

5.3 Proposed test statistics

As discussed in the introduction of this chapter, random variables $n_{21}(X)$ and

$n_{22}(X)$ are a function of $X = v_1 S_1^2$ from the first stage data, so T_{21} and T_{22} do not

follow the standard *t*-distribution under the null. Consequently, the type I error rate in

sample size re-estimation [46] cannot be controlled at the nominal level α if the

naïve choice $t_\alpha(n_{1\bullet}+n_{2\bullet}(X)-2)$ is used as the critical value for combined data from

both Stage 1 and Stage 2. We propose $T_{21}^* = (\hat{D}-\theta_1)\Big/\left(\dfrac{S^*}{2}\sqrt{\dfrac{1}{n_{11}+n_{21}(X)}+\dfrac{1}{n_{12}+n_{22}(X)}}\right)$ and

$$T_{22}^* = \left(\hat{D} - \theta_2\right) \Bigg/ \left(\frac{S^*}{2} \cdot \sqrt{\frac{1}{n_{11} + n_{21}(X)} + \frac{1}{n_{12} + n_{22}(X)}}\right) \text{ for hypothesis test (5.1) using the}$$

combined data of Stage 1 and Stage 2. Here $S^* = \sqrt{\dfrac{X + (n_{2\bullet}(X) - 2)S_2^2}{n_{1\bullet} + n_{2\bullet}(X) - 4}}$ and T_{21}^* and

T_{22}^* can be decomposed as functions of three components: Z_1, Z_2, X, and $v_2 S_2^2$.

We have $Z_i = \sqrt{n_{i\bullet}} \dfrac{\hat{D}_i - \theta_1}{\sqrt{\sigma_T^2 + \sigma_R^2}}$, $i=1,2$; Z_1 is a standard normal variable;

Z_2 conditioned on X is a standard normal variable, $X/V^2 \sim \chi^2(v_1)$, and

$v_2 S_2^2 / V^2$ conditioned on $X \sim \chi^2(v_2)$.

5.4 Unblinded sample size re-estimation procedure with exact power functions

In general, most bioequivalence trials are unblinded. So we proposed an unblinded

sample size re-estimation procedure described in Section 5.1 in which we compared

T_{21}^* with the exact critical value u_α and T_{22}^* with $-u_\alpha$ for the two-stage study. The

exact critical value, u_α, is derived as the largest value of u for which the condition:

$\max_{V}[P_{1s}(\theta_2, n_{11}, n_{12}, V) + P_{2s}(\theta_2, n_{11}, n_{12}, V, p_0, u_\alpha)] \le \alpha$ holds.

To assure the experimentwise type I error rate ($P_{1s}(\theta_2, n_{11}, n_{12}, V)$ +

$P_{2s}(\theta_2, n_{11}, n_{12}, V, p_0, u)$) at exact α value, we calculate $P_{1s}(\theta_2, n_{11}, n_{12}, V)$ and

$P_{2s}(\theta_2, n_{11}, n_{12}, V, p_0, u)$ from the exact power function using the numerical method

which will be described in Section 5.6.2. We also quantify all levels of errors for the

numerical calculation of $P_{1s}(\theta_2, n_{11}, n_{12}, V)$ and $P_{2s}(\theta_2, n_{11}, n_{12}, V, p_0, u)$ in the bounded

interval $\{V : V_b \leq V \leq V_e\}$ in Section 5.6.3 and make sure the sum of all levels of error

is no larger than 10^{-4}

5.5 Exact power function for the proposed test statistics

In this section, we first derive $P_{1s}\left(\theta^*, n_{11}, n_{12}, V\right)$ and $P_{2s}\left(\theta^*, n_{11}, n_{12}, V, p_0, u\right)$. Then

we prove several properties of the two power functions.

5.5.1 One-stage power function $P_{1s}\left(\theta^*, n_{11}, n_{12}, V\right)$

Recall the previously defined test statistics $T_{11} = \left(\hat{D}_1 - \theta_1\right)\Big/\left(\dfrac{S_1}{2}\sqrt{\dfrac{1}{n_{11}} + \dfrac{1}{n_{12}}}\right)$ and

$T_{12} = \left(\hat{D}_1 - \theta_2\right)\Big/\left(\dfrac{S_1}{2}\sqrt{\dfrac{1}{n_{11}} + \dfrac{1}{n_{12}}}\right)$ for the hypothesis test in Equation (5.1) for Stage 1.

Also recall previously defined random variable $X = \left(n_{1\bullet} - 2\right)S_1^2$.

Since $P_{1s}\left(\theta^*, n_{11}, n_{12}, V\right)$ is the joint probability of rejecting H_0 given $\theta^* \in \left(\theta_1, \theta_2\right)$

and $X \leq \left(n_{1\bullet} - 2\right)S_0^2$, then we derive its exact function as:

$$P_{1s}\left(\theta^*, n_{11}, n_{12}, V\right) = P\left(X \leq \left(n_{1\bullet} - 2\right)S_0^2, \text{Reject } H_0 \mid \mu_T - \mu_R = \theta^*, \theta^* \in \left(\theta_1, \theta_2\right)\right)$$

$$= P\left(\left(\hat{D}_1 - \theta_1\right)\Big/\left(\dfrac{S_1}{2}\bullet\sqrt{\dfrac{1}{n_{11}} + \dfrac{1}{n_{12}}}\right) \geq t_\alpha(v_1),\ \left(\hat{D}_1 - \theta_2\right)\Big/\left(\dfrac{S_1}{2}\bullet\sqrt{\dfrac{1}{n_{11}} + \dfrac{1}{n_{12}}}\right) \leq t_{1-\alpha}(v_1),\ X \leq v_1 S_0^2\right)$$

$$= P\left(\theta_1 + \dfrac{t_\alpha(v_1)}{2}\sqrt{\dfrac{X}{v_1}\left(\dfrac{1}{n_{11}} + \dfrac{1}{n_{12}}\right)} \leq \hat{D}_1 \leq \theta_2 - \dfrac{t_\alpha(v_1)}{2}\sqrt{\dfrac{X}{v_1}\left(\dfrac{1}{n_{11}} + \dfrac{1}{n_{12}}\right)},\ X \leq v_1 S_0^2\right)$$

$$= E\left\{\begin{array}{l} P\left(\theta_1 + \dfrac{t_\alpha(v_1)}{2}\sqrt{\dfrac{X}{v_1}\left(\dfrac{1}{n_{11}} + \dfrac{1}{n_{12}}\right)} \leq \hat{D}_1 \leq \theta_2 - \dfrac{t_\alpha(v_1)}{2}\sqrt{\dfrac{X}{v_1}\left(\dfrac{1}{n_{11}} + \dfrac{1}{n_{12}}\right)}\Big| X\right) \\ I\left(X \leq \min\left(v_1 S_0^2, \dfrac{v_1\left(\theta_2 - \theta_1\right)^2}{t_\alpha^2(v_1)}\left(\dfrac{1}{n_{11}} + \dfrac{1}{n_{12}}\right)^{-1}\right)\right) \end{array}\right\}$$

.

Using the fact that $\dfrac{X}{V^2} \sim \chi^2(v_1)$ and $\hat{D}_1 \sim N\!\left(\theta^*, \sigma_{\hat{D}_1}^2\right)$, this last expression is the

expectation over X of a conditional probability given X, which is expressed simply in

terms of the cumulative normal distribution function as:

$$E\left[\left\{\Phi\!\left(\frac{\theta_2-\theta^*}{\sigma_{\hat{D}_1}}-\frac{t_\alpha(v_1)}{V}\sqrt{\frac{X}{v_1}}\right)-\Phi\!\left(\frac{\theta_1-\theta^*}{\sigma_{\hat{D}_1}}+\frac{t_\alpha(v_1)}{V}\sqrt{\frac{X}{v_1}}\right)\right\}\; I\!\left(X \le \min\!\left(v_1 S_0^2, \frac{(\theta_2-\theta_1)^2}{t_\alpha^2(v_1)}\left(\frac{1}{n_{11}}+\frac{1}{n_{12}}\right)^{-1}\right)\right)\right],$$

where $t_\alpha(v_1)$ is the 1-α quantile of the t-distribution with $v_1 = n_{11}+n_{12}-2$ degrees

of freedom, $\Phi(\cdot)$ is the standard normal cumulative distribution function (cdf), $I(\cdot)$ is

the indicator function, and recall that $\sigma_{\hat{D}_1}^2 = \dfrac{V^2}{4}\left(\dfrac{1}{n_{11}}+\dfrac{1}{n_{12}}\right)$. Using the fact that $\dfrac{X}{V^2}$

$\sim \chi^2(v_1)$, we write the above expectation over X explicitly as:

$$\begin{aligned}
&P_{1s}\!\left(\theta^*, n_{11}, n_{12}, V\right)\\[4pt]
&=\int_0^{c_1}\left(\Phi\!\left(2\frac{\theta_2-\theta^*}{V}\left(\frac{1}{n_{11}}+\frac{1}{n_{12}}\right)^{-1/2}-\frac{t_\alpha(v_1)}{V}\sqrt{\frac{x}{v_1}}\right)-\Phi\!\left(2\frac{\theta_1-\theta^*}{V}\left(\frac{1}{n_{11}}+\frac{1}{n_{12}}\right)^{-1/2}+\frac{t_\alpha(v_1)}{V}\sqrt{\frac{x}{v_1}}\right)\right)\\[4pt]
&\quad\frac{1}{2^{v_1/2}\Gamma(v_1/2)V^2}\left(\frac{x}{V^2}\right)^{v_1/2-1}\exp\!\left(-x/(2V^2)\right)dx
\end{aligned}\qquad (5.3)$$

Here $c_1 = \min\!\left(v_1 S_0^2, \dfrac{(\theta_2-\theta_1)^2 v_1}{t_\alpha^2(v_1)}\left(\dfrac{1}{n_{11}}+\dfrac{1}{n_{12}}\right)^{-1}\right)$ and $v_1 = n_{11}+n_{12}-2$.

5.5.2 Two-stage power function $P_{2s}\!\left(\theta^*, n_{11}, n_{12}, V, p_0, u\right)$

We use the test statistics $T_{21}^* = \left(\hat{D}-\theta_1\right)\Big/\left(\dfrac{S^*}{2}\sqrt{\dfrac{1}{n_{11}+n_{21}(X)}+\dfrac{1}{n_{12}+n_{22}(X)}}\right)$ and

$T_{22}^* = \left(\hat{D}-\theta_2\right)\Big/\left(\dfrac{S^*}{2}\sqrt{\dfrac{1}{n_{11}+n_{21}(X)}+\dfrac{1}{n_{12}+n_{22}(X)}}\right)$ for hypothesis test in Equation (5.1)

after combining data from both Stage 1 and Stage 2. Note $S^* = \sqrt{\dfrac{X + (n_{2\bullet}(X) - 2)S_2^2}{n_{1\bullet} + n_{2\bullet}(X) - 4}}$,

where $X = (n_{1\bullet} - 2)S_1^2$.

Since $n_{2\bullet}(X)$ is a random variable, a function of X in Equation (5.4), then

$$E(\hat{D}|X = x) = E\left(\frac{n_{1\bullet}\hat{D}_1 + n_{2\bullet}(x)\hat{D}_2}{n_{1\bullet} + n_{2\bullet}(x)}\right) = \theta^* \text{ and}$$

$$\sigma_{\hat{D}|X=x} = \sqrt{\frac{V^2}{4(n_{1\bullet} + n_{2\bullet}(x))^2}\left\{n_{1\bullet}^2\left(\frac{1}{n_{11}} + \frac{1}{n_{12}}\right) + n_{2\bullet}^2(x)\left(\frac{1}{n_{21}(x)} + \frac{1}{n_{22}(x)}\right)\right\}}.$$

Since $P_{2s}(\theta^*, n_{11}, n_{12}, V, p_0, u)$ is the joint probability of rejecting H_0 given

$\theta^* \in (\theta_1, \theta_2)$ and $X > v_1 S_0^2$, then we derive its exact function as:

$$P_{2s}(\theta^*, n_{11}, n_{12}, V, p_0, u) = P(X > v_1 S_0^2, \text{Reject } H_0 \mid \mu_T - \mu_R = \theta^*, \theta^* \in (\theta_1, \theta_2))$$

$$= P\left(\begin{array}{l}(\hat{D} - \theta_1)\Big/\left(\dfrac{S^*}{2}\sqrt{\dfrac{1}{n_{11} + n_{21}(X)} + \dfrac{1}{n_{12} + n_{22}(X)}}\right) \geq u, \ (\hat{D} - \theta_2)\Big/\left(\dfrac{S^*}{2}\sqrt{\dfrac{1}{n_{11} + n_{21}(X)} + \dfrac{1}{n_{12} + n_{22}(X)}}\right) \leq -u, \\ X > v_1 S_0^2\end{array}\right).$$

$$= E\left\{P\left(\theta_1 + u\frac{S^*}{2}\sqrt{\frac{1}{n_{11} + n_{21}(X)} + \frac{1}{n_{12} + n_{22}(X)}} \leq \hat{D} \leq \theta_2 - u\frac{S^*}{2}\sqrt{\frac{1}{n_{11} + n_{21}(X)} + \frac{1}{n_{12} + n_{22}(X)}}\Big|X\right)I(X > v_1 S_0^2)\right\}$$

Since

$$\theta_1 + u\frac{S^*}{2}\sqrt{\frac{1}{n_{11} + n_{21}(X)} + \frac{1}{n_{12} + n_{22}(X)}} \leq \hat{D} \leq \theta_2 - u\frac{S^*}{2}\sqrt{\frac{1}{n_{11} + n_{21}(X)} + \frac{1}{n_{12} + n_{22}(X)}},$$

we obtain $uS^*\sqrt{\dfrac{1}{n_{11} + n_{21}(X)} + \dfrac{1}{n_{12} + n_{22}(X)}} \leq \theta_2 - \theta_1$. So

$$\frac{X + (n_{2\bullet}(X) - 2)S_2^2}{n_{1\bullet} + n_{2\bullet}(X) - 4} \leq \left[\frac{\theta_2 - \theta_1}{u}\left(\frac{1}{n_{11} + n_{21}(X)} + \frac{1}{n_{12} + n_{22}(X)}\right)^{-1}\right]^2.$$

To simply the above equation, let

$$c_2(X) = \left[\frac{\theta_2 - \theta_1}{u}\left(\frac{1}{n_{11} + n_{21}(X)} + \frac{1}{n_{12} + n_{22}(X)}\right)^{-1}\right].$$

It is easily seen that $S_2^2 \leq \dfrac{(n_{1\bullet} + n_{2\bullet}(X) - 4)c_2^2(X) - X}{n_{2\bullet}(X) - 2}$. If we define $Y = (n_{2\bullet} - 2)S_2^2$,

then $S^* = \sqrt{\dfrac{X + Y}{n_{1\bullet} + n_{2\bullet} - 4}}$. S_2^2 is the variance of $\hat{D}_2$ given X. Since $\dfrac{X}{V^2}$ is distributed

as $\chi^2(v_1)$ with degrees of freedom $v_1 = n_{11} + n_{12} - 2$ and conditional distribution of

$\dfrac{Y}{V^2}$ given X, is $\chi^2(v_2)$ with degrees of freedom $v_2 = n_{21} + n_{22} - 2$, we can shorten

the power expression for the true mean difference of the test product and the

reference product (θ^*) as the following

$$
\begin{aligned}
&P_{2s}\left(\theta^*, n_{11}, n_{12}, V, p_0, u\right) \\
&= E_X\left\{
\begin{aligned}
&E_Y\left[\Phi\left(\frac{\theta_2 - \theta^* - uc_3(x)\sqrt{x+y}}{\sigma_{\hat{D}|x}}\right) - \Phi\left(\frac{\theta_1 - \theta^* + uc_3(x)\sqrt{x+y}}{\sigma_{\hat{D}|x}}\right)\right]I\left(Y \leq (n_{1\bullet} + n_{2\bullet}(X) - 4)c_2^2(X) - X\right) \\
&I\left(X > v_1 S_0^2\right)
\end{aligned}
\right\} \quad .(5.5) \\
&= \int_b^\infty\int_0^a\left[\Phi\left(\frac{\theta_2 - \theta^* - uc_3(x)\sqrt{x+y}}{c_4(x)V}\right) - \Phi\left(\frac{\theta_1 - \theta^* + uc_3(x)\sqrt{x+y}}{c_4(x)V}\right)\right]\frac{1}{2^{(v_1+v_2)/2}\Gamma(v_1/2)\Gamma(v_2/2)\cdot V^{v_1+v_2}} \\
&\qquad y^{v_2/2 - 1}x^{v_1/2 - 1}e^{-\frac{x+y}{2V^2}}\,dy\,dx
\end{aligned}
$$

Since $Y \leq (n_{1\bullet} + n_{2\bullet}(X) - 4)c_2^2(X) - X$, it is obvious that $(n_{1\bullet} + n_{2\bullet}(x) - 4)c_2^2(x) - x$ is

the upper integration limit for the inner integral with respect to y. For ease of use, we

let $a = a(x) = (n_{1\bullet} + n_{2\bullet}(x) - 4)c_2^2(x) - x$. Since $X \geq (n_{1\bullet} - 2)S_0^2$ on the event whose

probability is calculated in the integral, the lower limit of the integration in x is

$(n_{1\bullet} - 2)S_0^2$.

Here u is the critical value, $c_3(x) = \dfrac{1}{2}\sqrt{\left(\dfrac{1}{n_{11}+n_{21}(x)}+\dfrac{1}{n_{12}+n_{22}(x)}\right)\dfrac{1}{n_{1\bullet}+n_{2\bullet}(x)-4}}$

and $c_4(x) = \sqrt{\dfrac{1}{4(n_{1\bullet}+n_{2\bullet}(x))^2}\left\{n_{1\bullet}^2\left(\dfrac{1}{n_{11}}+\dfrac{1}{n_{12}}\right)+n_{2\bullet}^2(x)\left(\dfrac{1}{n_{21}(x)}+\dfrac{1}{n_{22}(x)}\right)\right\}}$.

If $n_{21}(x) = n_{22}(x) = n_{2\bullet}(x)/2$, then $c_3(x) = \sqrt{\dfrac{1}{(n_{1\bullet}+n_{2\bullet}(x))(n_{1\bullet}+n_{2\bullet}(x)-4)}}$ and

$c_4(x) = \sqrt{\dfrac{1}{n_{1\bullet}+n_{2\bullet}(x)}}$.

5.5.3 Properties of exact power functions

The exact power functions $P_{1s}(\theta_2, n_{11}, n_{12}, V)$ and $P_{2s}(\theta^*, n_{11}, n_{12}, V, p_0, u)$ are calculated

by Equation (5.3) and Equation (5.5), respectively. $a(x)$, b, $c_2(x)$, $c_3(x)$, and $c_4(x)$ are

calculated by Equation (5.6), Equation (5.7), Equation (5.8), Equation (5.9), and Equation

(5.10), respectively. Clearly, all of them except b depend on x, not V, while b does not depend

on either x or V.

Before we develop the numerical method of obtaining the critical value, we will prove

several properties of $P_{1s}(\theta_2, n_{11}, n_{12}, V)$ and $P_{2s}(\theta^*, n_{11}, n_{12}, V, p_0, u)$.

Lemma 1: *Let* $H(b) = \displaystyle\int_c^{f(b)} g(x,b)dx$. *If both* $f(b)$ *and* $g(x,b)$ *are non-increasing*

continuous functions of b, *then* $H(b)$ *is a monotonically non-increasing continuous function*

of b *for a nonnegative* $g(x,b)$ *and* $f(b) \geq c$ *(c is a constant).*

Proof: Since $H(b) = \displaystyle\int_c^{f(b)} g(x,b)dx$, then $H(b)$ is a definite integral where both the limits

of integration and the integrand are functions of b. Leibniz rule [60] allows us to take the first

112

derivative for both sides, which results in $H'(b) = g(f(b),b)f'(b) + \int_c^{f(b)} \dfrac{\partial g(x,b)}{\partial b} dx$.

$H'(b) \leq 0$ since $f'(b) \leq 0$, $\dfrac{\partial g(x,b)}{\partial b} \leq 0$, $g(x,b) \geq 0$, and $f(b) \geq c$.

Proposition 1: $P_{2s}(\theta_2, n_{11}, n_{12}, V, p_0, u) \geq P_{2s}(\theta^*, n_{11}, n_{12}, V, p_0, u)$ for any $\theta^* \geq \theta_2$ or $\theta^* \leq \theta_1$.

Proof: First let us recall Equation (5.5) as follows:

$$P_{2s}(\theta^*, n_{11}, n_{12}, V, p_0, u)$$
$$= \int_b^\infty \int_0^a \left[\Phi\left(\frac{\theta_2 - \theta^* - uc_3(x)\sqrt{x+y}}{c_4(x)V} \right) - \Phi\left(\frac{\theta_1 - \theta^* + uc_3(x)\sqrt{x+y}}{c_4(x)V} \right) \right].$$
$$\frac{1}{2^{(\nu_1+\nu_2)/2} \Gamma(\nu_1/2)\Gamma(\nu_2/2) \cdot V^{\nu_1+\nu_2}} \, y^{\nu_2/2-1} x^{\nu_1/2-1} \exp\left(-\frac{x+y}{2V^2} \right) dy dx$$

Φ is monotonically increasing with derivative increasing on the negative axis, and

the difference of two normal distribution function values given in the integrand of

Equation (5.5) has the form $\Phi(z - K_1) - \Phi(z - K_2)$ for $z = -\theta^* / (c_4(x)V)$, where

$K_1 = -\dfrac{\theta_2 - uc_3\sqrt{x+y}}{c_4 V}$, $K_2 = -\dfrac{\theta_1 + uc_3\sqrt{x+y}}{c_4 V}$, and $K_2 > K_1$ does not involve θ^*.

Since the derivative of $\Phi(z - k)$ with respect to z is positive and increasing for

$z - k < 0$, it follows immediately from the mean value theorem and chain rule that

$$\frac{\partial}{\partial \theta^*} [\Phi(z - K_1) - \Phi(z - K_2)] < 0.$$

By Lemma 1, the decreasing property of the integrand with respect to θ^* gives the

integral the same property.

113

Similarly, since $\dfrac{\theta_2 - \theta^* - uc_3\sqrt{x+y}}{c_4 V} \geq \dfrac{\theta_1 - \theta^* + uc_3\sqrt{x+y}}{c_4 V} \geq 0$ for $\theta^* \leq \theta_1$, then the

derivative of $\Phi(z-k)$ with respect to z is decreasing for $z - k > 0$, it follows

immediately from the mean value theorem and chain rule [60] that

$$\frac{\partial}{\partial \theta^*}\left[\Phi(z-K_1) - \Phi(z-K_2)\right] > 0.$$

By Lemma 1, the increasing property of the integrand with respect to θ^* gives the

integral the same property. Therefore, we have proved that

$$P_{2s}(\theta_1, n_{11}, n_{12}, V, p_0, u) \geq P_{2s}(\theta^*, n_{11}, n_{12}, V, p_0, u) \text{ for any } \theta^* \geq \theta_2 \text{ or } \theta^* \leq \theta_1.$$

Proposition 2: $P_{1s}(\theta_2, n_{11}, n_{12}, V)$ *is a decreasing function of* V .

Proof: Substituting $\theta^* = \theta_2$ in Equation (5.3) and by change of variable, we have

$P_{1s}(\theta_2, n_{11}, n_{12}, V)$ as,

$$
\begin{aligned}
&P_{1s}(\theta_2, n_{11}, n_{12}, V) \\
&= \int_0^{c_1/V^2}\left(\Phi\left(-t_\alpha(v_1) \bullet \sqrt{\frac{x}{v_1}}\right) - \Phi\left(\frac{\theta_1 - \theta_2}{(V/2)\sqrt{1/n_{11} + 1/n_{12}}} + t_\alpha(v_1)\sqrt{\frac{x}{v_1}}\right)\right)\frac{1}{2^{v_1/2}\Gamma(v_1/2)}x^{v_1/2-1}e^{-x/2}dx
\end{aligned}
$$

$$(5.11)$$

where $c_1 = \min\left(v_1 S_0^2, \dfrac{(\theta_2 - \theta_1)^2 v_1}{t_\alpha^2(v_1)}\left(\dfrac{1}{n_{11}} + \dfrac{1}{n_{12}}\right)^{-1}\right)$ and $v_1 = n_{11} + n_{12} - 2$.

Since c_1 is a constant, free of V, then it is easily seen that c_1/V^2 is a decreasing

function of V. Since $\dfrac{\theta_1 - \theta_2}{(V/2)\sqrt{1/n_{11} + 1/n_{12}}} + t_\alpha(v_1)\sqrt{\dfrac{x}{v_1}} \leq 0$, then the integrand is

obviously decreasing in V.

114

Since both the upper integration and integrand bounded between 0 and 1 in the right side of Equation (5.11) are a monotonically decreasing function of V and $c_1 > 0$, which satisfy all conditions in Lemma 1, then $P_{1s}(\theta_2, n_{11}, n_{12}, V)$ is a monotonically decreasing function of V.

Proposition 3: $P_1(\theta_2, n_1, n_2, V)$ *is a decreasing function of V.*

Proof: As we discussed in Chapter 4, $P_1(\theta^*, n_1, n_2, V)$ is the probability of rejecting the null for the two one-sided tests procedure for the crossover bioequivalence study.

Plugging $\theta^* = \theta_2$ and $V = \sqrt{\sigma_T^2 + \sigma_R^2}$ into Equation (4.3), we have

$$P_1(\theta_2, n_1, n_2, V) = \int_0^{c_0} \left(\Phi\left(-t_\alpha(v)\sqrt{\frac{x}{v}} \right) - \Phi\left(\frac{\theta_1 - \theta_2}{(V/2)\sqrt{1/n_1 + 1/n_2}} + t_\alpha(v)\sqrt{\frac{x}{v}} \right) \right) \frac{1}{2^{v/2}\Gamma(v/2)} x^{v/2-1} e^{-x/2} dx,$$

$$(5.12)$$

where $c_0 = \dfrac{(\theta_2 - \theta_1)^2 v}{t_\alpha^2(v)V^2}\left(\dfrac{1}{n_1} + \dfrac{1}{n_2} \right)^{-1}$.

In Proposition 2, we proved that

$$\Phi\left(-t_\alpha(v)\sqrt{\frac{x}{v}} \right) - \Phi\left(\frac{\theta_1 - \theta_2}{(V/2)\sqrt{1/n_1 + 1/n_2}} + t_\alpha(v)\sqrt{\frac{x}{v}} \right)$$

is a monotonically decreasing function of V.

Obviously c_0 is a monotonically decreasing function of V.

Both the integrand, which is bounded between 0 and 1, and the upper integration limit are decreasing continuous functions of V and $c_0 > 0$, so from Lemma 1, we know that $P_1(\theta_2, n_1, n_2, V)$ is a non-increasing function of V.

Proposition 4: *For fixed* $\theta_2, n_{11}, n_{12}, p_0,$ *and* u, *there exists a maximum of*

$P_{1s}(\theta_2, n_{11}, n_{12}, V) + P_{2s}(\theta_2, n_{11}, n_{12}, V, p_0, u)$ *on* $\{V \geq 0\}$.

Proof: Recall $n_{2\bullet}(x)$ is a real valued quantity defined by Equation (5.4) rather than an integer, and so it is a continuous function of x. Substituting $\theta^* = \theta_2$ in Equation (5.5), we have

$$
\begin{aligned}
&P_{2s}(\theta_2, n_{11}, n_{12}, V, p_0, u) \\
&= \int_b^\infty \int_0^a \left[\Phi\left(\frac{\theta_2 - \theta_2 - uc_3(x)\sqrt{x+y}}{c_4(x)V} \right) - \Phi\left(\frac{\theta_1 - \theta_2 + uc_3(x)\sqrt{x+y}}{c_4(x)V} \right) \right] \cdot \\
&\quad \frac{1}{2^{(\nu_1+\nu_2)/2} \Gamma(\nu_1/2)\Gamma(\nu_2/2) \cdot V^{\nu_1+\nu_2}} \, y^{\nu_2/2-1} x^{\nu_1/2-1} \exp\left(-\frac{x+y}{2V^2} \right) dy dx
\end{aligned}
\tag{5.13}
$$

Since $a = a(x) = (n_{1\bullet} + n_{2\bullet}(x) - 4)c_2^2(x) - x$ and $b = (n_{1\bullet} - 2)S_0^2$, both are free of V. From the above equation, $P_{2s}(\theta_2, n_{11}, n_{12}, V, p_0, u)$ is always a nonnegative continuous function of V for any $V \geq 0$. Similarly, from Equation (5.11), $P_{1s}(\theta_2, n_{11}, n_{12}, V)$ is always a nonnegative continuous function of V for any $V \geq 0$.

$P_{2s}(\theta_2, n_{11}, n_{12}, V, p_0, u) \to 0$ as $V \to 0$ since both $\Phi(\cdot)$ values in the integrand of right side in Equation (5.13) go to 0. $P_{1s}(\theta_2, n_{11}, n_{12}, V) \to \alpha$ as $V \to 0$ since

116

$$\Phi\left(\frac{\theta_1 - \theta_2}{(V/2)\sqrt{1/n_{11} + 1/n_{12}}} + t_\alpha(V_1)\sqrt{\frac{x}{V_1}}\right)$$ in the integrand of right side in Equation (5.11)

goes to 0. Therefore, $P_{1s}(\theta_2, n_{11}, n_{12}, V) + P_{2s}(\theta_2, n_{11}, n_{12}, V, p_0, u) \to \alpha$ as $V \to 0$.

$P_{2s}(\theta_2, n_{11}, n_{12}, V, p_0, u) \to 0$ as $V \to \infty$ since both $\Phi(\cdot)$ values in the integrand of

right side in Equation (5.13) go to 1/2. $P_{1s}(\theta_2, n_{11}, n_{12}, V) \to 0$ as $V \to \infty$ since

$$\Phi\left(\frac{\theta_1 - \theta_2}{(V/2)\sqrt{1/n_{11} + 1/n_{12}}} + t_\alpha(V_1)\sqrt{\frac{x}{V_1}}\right) \to \Phi\left(t_\alpha(V_1)\sqrt{\frac{x}{V_1}}\right)$$ in the integrand of right side

in Equation (5.11). Therefore, $P_{1s}(\theta_2, n_{11}, n_{12}, V) + P_{2s}(\theta_2, n_{11}, n_{12}, V, p_0, u) \to 0$ as

$V \to \infty$.

We can find $\delta > 0$ such that $\displaystyle\operatorname*{Sup}_{V \in [0, 1/\delta] \cup [1/\delta, \infty]} P_{1s}(V) + P_{2s}(V, u) < P_{1s}(l) + P_{2s}(l, u)$,

$l \in [\delta, 1/\delta]$ and then the maximum of the continuous function $P_{1s}(V) + P_{2s}(V, u)$ exists

on $[\delta, 1/\delta]$.

Let N be the upper bound for the sample size of n. In Proposition 5, we will bound

the any solution of n of Equation (5.4) is less than or equal to N.

Proposition 5: *If N is greater than or equal to*

$$\max\left(33, 3S_1^2 \Big/ (\theta t_\alpha(v))^2, \frac{S_1}{\theta}\left(3t_\alpha(v) + \Phi^{-1}(0.5(1.001 + p_0))\right)\right), \text{ then any solution of n of}$$

Equation (5.4) is less than or equal to N. We rewrite Equation (5.4) as follows,

$$p_0 = \int_0^{c_0} \left(\Phi\left(\frac{\theta_2 - \theta_0}{S_1\sqrt{1/n}} - t_\alpha(v)\sqrt{\frac{x}{v}} \right) - \Phi\left(\frac{\theta_1 - \theta_0}{S_1\sqrt{1/n}} + t_\alpha(v)\sqrt{\frac{x}{v}} \right) \right) \frac{1}{2^{v/2}\Gamma(v/2)} x^{v/2-1} e^{-x/2} dx, \; where$$

$$\theta_0 = 0, \; p_0 < 0.999, \; c_0 = \frac{(\theta_2 - \theta_1)^2 vn}{4t_\alpha^2(v)S_1^2}, \; n = n_{1\bullet} + n_{2\bullet}(X) \text{ and } v = n - 2.$$

Proof: Let $X(v)$ denote a random variable as χ_v^2, then the right-hand side

expression of the formula we will equate to p_0 is

$$E\left(\left\{ \Phi\left(\frac{\theta_2}{S_1\sqrt{1/n}} - t_\alpha(v)\sqrt{\frac{X(v)}{v}} \right) - \Phi\left(\frac{\theta_1}{S_1\sqrt{1/n}} + t_\alpha(v)\sqrt{\frac{X(v)}{v}} \right) \right\}^+ \right) \qquad (5.14)$$

The proof goes in three steps:

(1) For $v \geq 31$, $P(X(v) \geq 3v) < 10^{-3}$ which follows from the Markov-

Cheybchev inequality and the moment generating function of chi-square,

since $P(X(v) \geq 3v) < E(\exp((3/7)(X - 3v))) = \exp(-(9/7 - 0.5\log(7))v) < \exp(-0.3v)$.

Taking $v \geq 31$ gives the right-hand side $< 10^{-4}$.

(2) When $v \geq 33$, so that $v \geq 31$, also $t_\alpha(v) \leq t_\alpha(31)$, and the integrand in

curly brackets in (5.14) is positive for all $X \leq 3v$ as long as

$$t_\alpha(v)\sqrt{3} < \sqrt{n}\theta / S_1$$

which holds as long as $n \geq 3S_1^2 / (\theta t_\alpha(v))^2$.

(3) Finally, by removing an event of probability at most 10^{-4} from the

expectation on the right of (5.14), we have the curly-bracketed term strictly

118

positive with $X/v \leq 3$ on the complement, which means that (5.14) is

$$\geq -10^{-4} + \Phi\left(\sqrt{n}\frac{\theta_2}{S_1} - \sqrt{3}t_\alpha(v)\right) - \Phi\left(-\sqrt{n}\frac{\theta_2}{S_1} + \sqrt{3}t_\alpha(v)\right), \text{ which is } > p_0 \text{ as}$$

long as n satisfies the bound in (1),(2) above along with

$$2\Phi\left(\sqrt{n}\frac{\theta_2}{S_1} - \sqrt{3}t_\alpha(v)\right) - 1 - 10^{-4} > p_0. \text{ Proof is over.}$$

Let V_b be the lower bound of the search range for V. V_b is obtained from the

following inequality: $v_1 S_0^2 / V_b^2 \geq q\chi^2(1-10^{-5}, v_1)$, where $q\chi^2(1-10^{-5}, v_1)$ is the

quantile of $\chi^2_{1-10^{-5}}(v_1)$. Let V_e be the upper bound of the search range for V. In

Proposition 6, V_e is chosen so large that the RHS of the following inequality $<$

0.510^{-4}. $P_{2s}(V,u) \leq P(a(X) \geq 0) \leq \inf_{v \in [v_1, N]} P\left(\chi_v^2 < v(v+4)(\theta_2 - \theta_1)^2 / (4V^2)\right).$

In Proposition 6, the lower bound V_b and upper bound V_e for the search range of V

can be determined by satisfying the above conditions.

Proposition 6: *We find that* $P_{1s}(\theta_2, n_{11}, n_{12}, V) + P_{2s}(\theta_2, n_{11}, n_{12}, V, p_0, u) < \alpha$

whenever $V < V_b$ *or* $V > V_e$.

Proof: Let $X(v_1) = v_1 S_1^2$.

Step (1) is to recall P_{1s} is decreasing in V and by the Intersection-Union Test size

inequality of Berger and Casella [50] (see Theorem 8.3.24 on Page 396 in Section

8.3.3), $P_{1s}(\theta^*, n_{11}, n_{12}, V > 0) < \alpha$.

Step (2) is to note that uniformly in u,

119

$$P_{2s}(V,u) \le P(S_1 \ge S_0) = P\left(X(v_1)/V^2 \ge v_1 S_0^2/V^2\right)$$

$P_{2s}(V,u)$ goes to 0 when $V \to 0$. Therefore, with $v_1 S_0^2/V_b^2 \ge q\chi^2\left(1-10^{-5}, v_1\right)$, we

find that $P_{1s}(\theta_2, n_{11}, n_{12}, V) + P_{2s}(\theta_2, n_{11}, n_{12}, V, p_0, u) < \alpha$ whenever $V < V_b$.

Step (3) is to note that for large V, $P_{1s}(V) \le P(S_1 < S_0) \to 0$ (and express this

probability in terms of $\chi_{v_1}^2$ quantiles as $p\chi^2\left(v_1 S_0^2/V^2, v_1\right)$), where $p\chi^2\left(v_1 S_0^2/V^2, v_1\right)$

is the probability of $\chi_{v_1}^2$ smaller than or equal to $v_1 S_0^2/V^2$.

Step (4) is to note that for fixed u and $V \to \infty$,

$a(X) = v(v+4)(\theta_2 - \theta_1)^2/4 - V^2\left(X(v)/V^2\right)$ and $P_{2s}(V,u) \le P(a(X) \ge 0)$. This goes to 0

because v does not depend on V and according to Prop.5 is bounded above by N,

while $X(v)/V^2$ is χ_v^2 distributed. In particular,

$$P_{2s}(V,u) \le P(a(X) \ge 0) \le \inf_{v \in [v_1, N]} P\left(\chi_v^2 < v(v+4)(\theta_2 - \theta_1)^2/\left(4V^2\right)\right)$$

and V_e is chosen so large that the RHS $< 0.510^{-4}$. Therefore

$P_{1s}(\theta_2, n_{11}, n_{12}, V) + P_{2s}(\theta_2, n_{11}, n_{12}, V, p_0, u) < \alpha$ whenever $V > V_e$.

Remark 1: In Proposition 6, we established the search range $\{V : V_b \le V \le V_e\}$ for

V. Here we determine the search range for positive u for T_{21}^*. The positive lower

bound for u is Z_α since our calculation indicates that

$P_{1s}(\theta_2, n_{11}, n_{12}, V) + P_{2s}(\theta_2, n_{11}, n_{12}, V, p_0, u) > \alpha$ for some V when $u = Z_\alpha$.

We differentiate Equation (5.13), then

$$\frac{\partial}{\partial u} P_{2s}\left(\theta_2, n_{11}, n_{12}, V, p_0, u\right)$$

$$= \int_b^\infty \int_0^a -\left[\phi\left(\frac{-uc_3(x)\sqrt{x+y}}{c_4(x)V}\right) + \phi\left(\frac{\theta_1 - \theta_2 + uc_3(x)\sqrt{x+y}}{c_4(x)V}\right)\right]\frac{c_3(x)\sqrt{x+y}}{c_4(x)V} \qquad (5.15).$$

$$\frac{1}{2^{(\nu_1+\nu_2)/2}\Gamma(\nu_1/2)\Gamma(\nu_2/2)\cdot V^{\nu_1+\nu_2}} \, y^{\nu_2/2-1} x^{\nu_1/2-1} e^{-\frac{x+y}{2V^2}} \, dy dx$$

It is easy to see that the first partial derivative of $P_{2s}\left(\theta_2, n_{11}, n_{12}, V, p_0, u\right)$ with

respect to u is positive since $c_3(x) > 0$ and $c_4(x) > 0$. Therefore,

$P_{2s}\left(\theta_2, n_{11}, n_{12}, V, p_0, u\right)$ is a decreasing function of positive u for a given positive V.

When $u \to \infty$, $P_{2s}\left(\theta_2, n_{11}, n_{12}, V, p_0, u\right) \to -1$ uniformly for all positive V. Hence

$\exists u_0$ such that $P_{1s}\left(\theta_2, n_{11}, n_{12}, V\right) + P_{2s}\left(\theta_2, n_{11}, n_{12}, V, p_0, u_0\right) < \alpha$ for all positive V.

Let $B = [V_b, V_e] \times [Z_\alpha, u_0]$. B is a compact set. By intermediate value theorem [60],

the exact critical value, u_α, is derived as the largest value of u for which the

condition: $\max_V [P_{1s}\left(\theta_2, n_{11}, n_{12}, V\right) + P_{2s}\left(\theta_2, n_{11}, n_{12}, V, p_0, u_\alpha\right)] \le \alpha$ holds.

5.6 Numerical analysis

The type I error rate for our proposed sample re-estimation procedure is

$P_{1s}\left(\theta_2, n_{11}, n_{12}, V\right) + P_{2s}\left(\theta_2, n_{11}, n_{12}, V, p_0, u\right)$ since the maximum of the probability of

rejecting H_0 under H_0 in the hypothesis (5.1) occurs at the boundary θ_2 or θ_1 (see

Proposition 1).

Section 5.6.1 summarizes the overall method for obtaining the largest u such that

$P_{1s}\left(\theta_2, n_{11}, n_{12}, V\right) + P_{2s}\left(\theta_2, n_{11}, n_{12}, V, p_0, u\right)$ is exact α. We denote this u as u_α. Section 5.6.2

summarizes the numerical method for approximating the integrals in the calculation

of $P_{1s}(\theta_2, n_{11}, n_{12}, V)$ and $P_{2s}(\theta_2, n_{11}, n_{12}, V, p_0, u)$. Section 5.6.3 summarizes all numerical

errors including the numerical approximation error bounds for the integrals at a fixed

V and the upper bound for the difference between two adjacent grid points such as

$[V_k, V_{k+1}]$, k=1,.2, …. Section 5.6.4 presents an example for comparing the numerical

value for $P_{1s}(\theta_2, n_{11}, n_{12}, V) + P_{2s}(\theta_2, n_{11}, n_{12}, V, p_0, u)$ for u at two values.

5.6.1 Numerical method for obtaining critical values

The first part of this section provides the conceptual reasoning to prove there exists

a global maximum of $P_{1s}(\theta_2, n_{11}, n_{12}, V) + P_{2s}(\theta_2, n_{11}, n_{12}, V, p_0, u)$ in the range

$\{V : V_b \leq V \leq V_e\}$ for V.

The second part of this section provides the computational method for obtaining

u_α as the largest u such that $P_{1s}(\theta_2, n_{11}, n_{12}, V) + P_{2s}(\theta_2, n_{11}, n_{12}, V, p_0, u)$ takes a value as

large as α for some V within the interval $\{V : V_b \leq V \leq V_e\}$.

Due to the complexity of the analytical expression for the second derivative with

respect to V of $P_{1s}(\theta_2, n_{11}, n_{12}, V) + P_{2s}(\theta_2, n_{11}, n_{12}, V, p_0, u)$, it is impossible to derive its

maximum in the bounded range $\{V : V_b \leq V \leq V_e\}$. However, the existence of this

maximum in a bounded interval of V can be reasonably assumed since $P(S_1 > S_0)$

approaches 1 for any unusually large V. As $P(S_1 > S_0) \to 1$, the influence of the Stage

1 data is not significantly important and the two-stage study can be approximately

treated as the fixed sample study. Therefore, $P_{2s}(\theta_2, n_{11}, n_{12}, V, p_0, u)$ is a decreasing

function of V for unusually large V since $P_1(\theta_2, n_1, n_2, V)$ is a decreasing function of V

for any $V>0$ (see Proposition 3). We also know that $P_{1s}(\theta_2, n_{11}, n_{12}, V)$ is a

monotonically decreasing function of V for any $V>0$ (see Proposition 2). Hence

$P_{1s}(\theta_2, n_{11}, n_{12}, V) + P_{2s}(\theta_2, n_{11}, n_{12}, V, p_0, u)$ is a decreasing function of V for large

positive V. From Proposition 4, we know that there exists a maximum for

$P_{1s}(\theta_2, n_{11}, n_{12}, V) + P_{2s}(\theta_2, n_{11}, n_{12}, V, p_0, u)$ for any $V>0$. By carefully choosing the search

range $[V_b, V_e]$ for V and $u \geq Z_\alpha$ for u (see Proposition 6 and Remark 1), respectively,

we can assure that there exists a maximum in the bounded interval $[V_b, V_e]$ and this

maximum is the global maximum for any $V>0$ based on the calculation.

The algorithm to obtain u_α is described in the following steps. We calculate the

n_{21} and n_{22} from Equation (5.4), which is bounded by N (see Proposition 5) and

obtain the maximum of $P_{1s}(\theta_2, n_{11}, n_{12}, V) + P_{2s}(\theta_2, n_{11}, n_{12}, V, p_0, u)$ for a fixed u in the

bounded interval $[V_b, V_e]$. We use Z_α as the initial value for u. If the maximum of

$P_{1s}(\theta_2, n_{11}, n_{12}, V) + P_{2s}(\theta_2, n_{11}, n_{12}, V, p_0, u)$ is greater than α, we increase u by 0.001. We

continue this iteration until the difference between α and the maximum of

$P_{1s}(\theta_2, n_{11}, n_{12}, V) + P_{2s}(\theta^*, n_{11}, n_{12}, V, p_0, u)$ is less than or equal to 10^{-4}. The last u is

u_α.

For a fixed u, we use a grid search to obtain the maximum of $P_{1s}(\theta_2, n_{11}, n_{12}, V) +$

$P_{2s}(\theta_2, n_{11}, n_{12}, V, p_0, u)$ among the grid points in the bounded interval $[V_b, V_e]$. The

numerical value for $P_{1s}(\theta_2, n_{11}, n_{12}, V)$ and $P_{2s}(\theta_2, n_{11}, n_{12}, V, p_0, u)$ at a given V can be

obtained by the numerical approximation methods described in Section 5.6.2.

123

The critical value for the two-stage study, u_α, is determined as the largest u such that $P_{1s}(\theta_2, n_{11}, n_{12}, V) + P_{2s}(\theta_2, n_{11}, n_{12}, V, p_0, u)$ takes a value as large as α for some V within the interval $\{V : V_b \leq V \leq V_e\}$. Error bounds for numerical approximation of $P_{1s}(\theta_2, n_{11}, n_{12}, V)$ and $P_{2s}(\theta_2, n_{11}, n_{12}, V, p_0, u)$ can be obtained in Section 5.6.3.1. The maximum difference between two adjacent grid points can be quantified in Section 5.6.3.2.

5.6.2 Numerical calculation of $P_{1s}(\theta^*, n_{11}, n_{12}, V)$ and $P_{2s}(\theta^*, n_{11}, n_{12}, V, p_0, u)$

For a fixed critical value u, we use a grid search in the bounded interval $[V_b, V_e]$ to obtain the maximum of $P_{1s}(\theta_2, n_{11}, n_{12}, V) + P_{2s}(\theta_2, n_{11}, n_{12}, V, p_0, u)$. We divide the interval $[V_b, V_e]$ into many subintervals of length 10^{-3} or 10^{-4}. The typical interval $[V_b, V_e]$ for $n_{11} (= n_{12})$ less than 30 is $[0.1, 0.7]$

For a given V, $P_{1s}(\theta_2, n_{11}, n_{12}, V)$ is calculated by a definite integral $\int_0^{c_1} f^*(x)dx$ (see Equation (5.16)). The numerical value of $\int_0^{c_1} f^*(x)dx$ is calculated by Equation (5.17) which is derived from Simpson's rule. In Simpson's Rule we approximate the function $f^*(x)$ as a quadratic in each interval and require that the quadratic agree with three of the points from each subinterval.

From Equation (5.19), $P_{2s}(\theta_2, n_{11}, n_{12}, V, p_0, u)$ is calculated by an improper integral $\int_b^\infty \int_0^a f(x, y)dydx$. We approximate the improper integral $\int_b^\infty \int_0^a f(x, y)dydx$ by the definite

124

integral $\int_{b}^{b^*}\int_{0}^{a} f(x,y)dydx$ such that $\int_{b}^{\infty}\int_{0}^{a} f(x,y)dydx - \int_{b}^{b^*}\int_{0}^{a} f(x,y)dydx \leq 5\cdot 10^{-7}$, where

$b^* = V^2 \chi^2_{0.999999}(n_{1\bullet} - 2)$. Now we need to compute a two dimensional definite integral.

To ensure a good approximation for the entire integral, we have to approximate the

inner integral first and then use a second approximation to deal with the outer

integral. We use Simpson's rule for estimating $\int_{0}^{a} f(x,y)dy$ by $\hat{I}_1(x)$ (see Equation

(5.20)). We approximate $\int_{b}^{b^*} \hat{I}_1(x)dx$ by Equation (5.24) with the trapezoidal rule.

5.6.3 Numerical approximation errors

This section mainly discusses all levels of numerical errors. The first subsection

discusses the numerical approximation error bounds for the integral at a fixed V and

the second subsection numerically quantifies the maximum difference between two

adjacent grid points using monotone functions in the integrand for any interval

$[V_k, V_{k+1}]$, k=1,.2, … , in the neighborhood of V^* at which the maximum of

$P_{1s}(\theta_2, n_{11}, n_{12}, V^*) + P_{2s}(\theta_2, n_{11}, n_{12}, V^*, p_0, u_\alpha)$ occurs.

5.6.3.1 Numerical approximation error bounds for the integrals at a fixed V

1. Error bound for the numerical approximation of $P_{1s}(\theta^*, n_{11}, n_{12}, V)$

The numerical approximation error bound, $R_S^{P_{1s}}(f^*)$ [61] (see Equation (5.18)), in

approximating the definite integral $\int_0^{c_1} f^*(x)dx$, is an analytical upper bound since it is

calculated from $\max_{0 \le \xi \le c_1} \left| \dfrac{\partial^4 f^*(\xi)}{\partial x^4} \right|$, which is the analytical and uniform upper bound

in x on the bounded interval $[0, c_1]$ for the fourth derivative with respect to x of $f^*(x)$

(see Appendix 5.2 for more detail).

2. Error bounds for the numerical approximation of $P_{2s}(\theta_2, n_{11}, n_{12}, V, p_0, u)$

When calculating the numerical value of $P_{2s}(\theta_2, n_{11}, n_{12}, V, p_0, u)$, the first numerical

error is the difference between the definite integral and an improper integral. This

difference is $5 \cdot 10^{-7}$. The numerical approximation error bound, $R_S^{inner}(x, f)$ (see

Equation (5.21)), in approximating the definite integral $\int_0^a f(x, y)dy$, is an analytical

upper bound since it is calculated from $\max_{0 \le \xi \le a} \left| \dfrac{\partial^4 f(x, \xi)}{\partial y^4} \right|$, which is the analytical

and uniform upper bound in y on the bounded interval $[0, a]$ for the fourth derivative

with respect to y of $f(x, y)$ (see Appendix 5.3 for more detail).

The numerical value for the definite integral of $R_S^{inner}(x, f)$ is approximate since

$\int_b^{b^*} R_S^{inner}(x, f)dx$ is approximated by $\hat{I}_3$ (see Equation (5.22)) using the trapezoidal

rule. The numerical approximation error bound, $R(\hat{I}_3(\zeta))$ (see Equation (5.23)), in

approximating the definite integral $\int_b^{b^*} R_S^{inner}(x,f)dx$, is approximate since we bound

$\left(d^2/dx^2 \left(R_S^{inner}(x,f)\right)\right)$ by the upper bound of second finite differences of the integrand

values $R_S^{inner}(x,f)$ at the grid points (10^4) for the trapezoidal rule. This is necessary

because we can't analytically bound $\left(d^2/dx^2 \left(R_S^{inner}(x,f)\right)\right)$ on the bounded interval

$\left[b,b^*\right]$.

The numerical approximation error bound, $R_S^{outer}\left(\hat{I}_1(\varsigma)\right)$ (see Equation (5.25)), for

approximating the definite integral $\int_b^{b^*} \hat{I}_1(x)dx$, is approximate since we bound

$d^2/dx^2 \left(\hat{I}_1(x)\right)$ by the upper bound of second finite differences of the integrand values

$\hat{I}_1$ at up to 10^4 breakpoints for the trapezoidal rule. As above, we cannot analytically

bound $d^2/dx^2 \left(\hat{I}_1(x)\right)$ on the bounded interval $\left[b,b^*\right]$.

5.6.3.2 Maximum difference between two adjacent grid points

Although one can refine the grid, one cannot arrive at any mathematically provable

bound on the trapezoid-rule errors. We quantify the bound on $P_{1s}(\theta_2,n_{11},n_{12},V)$

$+P_{2s}(\theta_2,n_{11},n_{12},V,p_0,u)$ between grid-points in V so that we can check the smoothness

and continuity with respect to V of $P_{1s}(\theta_2,n_{11},n_{12},V)+P_{2s}(\theta_2,n_{11},n_{12},V,p_0,u)$.

From Equation (5.9), the upper bound for $P_{1s}(\theta_2, n_{11}, n_{12}, V)$ in the bounded interval $[V_k, V_{k+1}]$ can be calculated by bP_1.

$$bP_1 = \int_0^{c_1} \left(\Phi\left(-\frac{t_\alpha(v_1)}{V_{k+1}} \cdot \sqrt{\frac{x}{v_1}} \right) - \Phi\left(\frac{\theta_1 - \theta_2}{\frac{V_k}{2}\sqrt{\frac{1}{n_{11}} + \frac{1}{n_{12}}}} + \frac{t_\alpha(v_1)}{V_k} \cdot \sqrt{\frac{x}{v_1}} \right) \right) \frac{1}{2^{v_1/2} \Gamma(v_1/2) V_k^{v_1}} x^{v_1/2-1} e^{-x/2V_{k+1}^2} dx \cdot$$

By choosing a large number of grid points, the upper bound between bP_1 and $P_{1s}(\theta_2, n_{11}, n_{12} V_k)$ for any interval $[V_k, V_{k+1}]$, $k=1,.2, \ldots$, in the neighborhood of V^* at which the maximum of $P_{1s}(\theta_2, n_{11}, n_{12}, V^*) + P_{2s}(\theta_2, n_{11}, n_{12}, V^*, p_0, u_\alpha)$ occurs is 10^{-5}

From Equation (5.19), we approximate the upper bound for $P_{2s}(\theta_2, n_{11}, n_{12}, V, p_0, u_\alpha)$ in the bounded interval $[V_k, V_{k+1}]$ by BP_2.

$$BP_2 = \int_b^\infty \int_0^a \left[\Phi\left(\frac{\theta_2 - \theta^* - u \cdot c_3 \cdot \sqrt{x+y}}{c_4 \cdot V_k} \right) - \Phi\left(\frac{\theta_1 - \theta^* + u \cdot c_3 \cdot \sqrt{x+y}}{c_4 \cdot V_{K+1}} \right) \right] \frac{1}{2^{(v_1+v_2)/2} \Gamma(v_1/2)\Gamma(v_2/2) \cdot V_k^{v_1+v_2}} y^{v_2/2-1} x^{v_1/2-1} e^{-(x+y)/(2V_{k+1}^2)} dy dx \cdot$$

By choosing a large number of grid points, the maximum difference between bP_1 + BP_2 and $P_{1s}(\theta_2, n_{11}, n_{12} V_k)$ + $P_{2s}(\theta_2, n_{11}, n_{12}, V_k, p_0, u_\alpha)$ for any interval $[V_k, V_{k+1}]$, $k=1,.2, \ldots$, in the neighborhood of V^* at which the maximum of $P_{1s}(\theta_2, n_{11}, n_{12}, V^*)$ + $P_{2s}(\theta_2, n_{11}, n_{12}, V^*, p_0, u_\alpha)$ occurs is 10^{-4}

5.6.3.3 A numerical example

We present an example for illustrating the levels of all error bounds. Using Simpson's rule with 5,000 intermediate points, for the parameter values $(\theta_2, n_{11}, n_{12}, V) = (\log(1.25), 10, 10, 0.275)$, $R_S^{P_{1s}}(f^*)$ is equal to $2.22 \cdot 10^{-16}$. For the

parameter values $(\theta_2, n_{11}, n_{12}, V, u) = (\log(1.25), 10, 10, 0.275, 1.782)$, the difference between

the definite integral and an improper integral for $P_{2s}(\theta_2, n_{11}, n_{12}, V, p_0, u)$ is $5 \cdot 10^{-7}$.

$R_S^{outer}(\hat{I}_1(\varsigma))$ and $R(\hat{I}_3(\zeta))$ from the trapezoidal method with 5,000 intermediate points

are, respectively, $1.87 \cdot 10^{-4}$ and $1.60 \cdot 10^{-9}$. $\hat{I}_3$ is $1.20 \cdot 10^{-7}$. Let $R_S^{P_{2s}}$ be the total error

for $P_{2s}(\theta_2, n_{11}, n_{12}, V, p_0, u)$. $R_S^{P_{2s}} = 5 \cdot 10^{-7} + 1.87 \cdot 10^{-4} + 1.60 \cdot 10^{-9} + 1.20 \cdot 10^{-7} = 1.87 \cdot 10^{-4}$.

Hence the total integration error of $P_{1s}(\theta_2, n_{11}, n_{12}, V) + P_{2s}(\theta^*, n_{11}, n_{12}, V, p_0, u)$ for

$V = 0.275$ is $2.22 \cdot 10^{-16} + 1.87 \cdot 10^{-4} = 1.87 \cdot 10^{-4}$. $bP_1 + BP_2$ in which V is within the

interval [0.2750, 0.2751] is 0.05010652 while the numerical value of $P_{1s}(\theta_2, n_{11}, n_{12}, V)$

$+ P_{2s}(\theta_2, n_{11}, n_{12}, V, p_0, u)$ is 0.04974823 for the parameter values

$(\theta_2, n_{11}, n_{12}, V, u) = (\log(1.25), 10, 10, 0.275, 1.782)$. The upper bound for the difference

between $bP_1 + BP_2$ and $P_{1s}(\theta_2, n_{11}, n_{12}, V) + P_{2s}(\theta_2, n_{11}, n_{12}, V, p_0, u)$ is $3.6 \cdot 10^{-4}$. So the

upper bound for $P_{1s}(\theta_2, n_{11}, n_{12}, V) + P_{2s}(\theta_2, n_{11}, n_{12}, V, p_0, u)$ is 0.0503.

5.6.4 An example for comparing $P_{1s}(\theta_2, n_{11}, n_{12}, V) + P_{2s}(\theta_2, n_{11}, n_{12}, V, p_0, u)$ with
respect to V for two different u values

We present one example to show the behavior of

$P_{1s}(\theta_2, n_{11}, n_{12}, V) + P_{2s}(\theta_2, n_{11}, n_{12}, V, p_0, u)$, with respect to V for two different u values

when $n_{11} = n_{12} = 10$ in Figure 5.1. In Figure 5.1, we use $t_{0.05}(18)$ as the value of u for

the dot line and use u_α as the critical value of u for the solid line. It is clearly shown

that the type I error rate is inflated if $t_{0.05}(18)$ is used as the critical value. With

u=1.782, the type I error rate does not exceed 0.05. The type I error rate becomes larger if $t_{0.05}(v)$ is used for $v > 18$ when more subjects are added in the second stage.

Figure 5.1 Experimentwise type I error rate, $P_{1s}(\theta_2, n_{11}, n_{12}, V) + P_{2s}(\theta_2, n_{11}, n_{12}, V, p_0, u)$, against V for the critical value at 2 levels for $n_{11} = n_{12} = 10$, $p_0 = 0.9$, and $\theta_0 = 0.05$
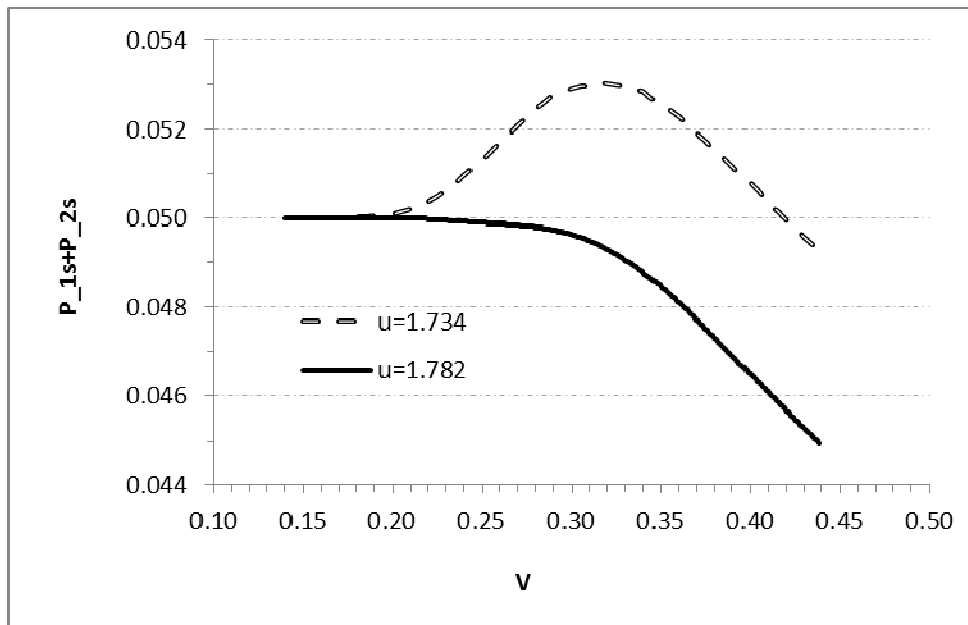
5.7 Comparison of T_{22} and T_{22}^*

Some statistical power for analyzing the combined data of Stage 1 and Stage 2 can be lost since the degrees of freedom for T_{22} are $n_{1\bullet} + n_{2\bullet} - 2$ and the degrees of freedom for T_{22}^* are $n_{1\bullet} + n_{2\bullet} - 4$. In order to evaluate how much power may be lost, we will evaluate the difference in the null distributions of T_{22} and T_{22}^*, and the alternative distributions of T_{22} and T_{22}^*.

Comparing S^2 and $(S^*)^2$, we can easily see that the difference between T_{22} and T_{22}^* is that the contribution from the difference of the averaged period differences for each sequence between Stage 1 and Stage 2 is omitted in T_{22}^*. Let dP be the difference in period effect between Period 2 and Period 1. It is reasonable to assume that dP is the same for both stages since they belong to the same study.

5.7.1 Comparison of the null distributions between T_{22} and T_{22}^*

To evaluate the null distributions of T_{22} and T_{22}^*, we use the following simulation schemes assuming that $\theta^* = \log(1.25)$ and $dP = 0.05$:

1) Simulate d_{11k}, k=1,2...,n_{11}, from a normal distribution with $E(d_{11k}) = \theta^* + dP$ and $V(d_{11k}) = V^2$, and d_{12k}, k=1,2...,n_{12}, from a normal distribution with $E(d_{12k}) = -\theta^* + dP$ and $V(d_{12k}) = V^2$.

2) Compute S_1^2 from Stage 1 data generated in Step 1.

3) Calculate the new total sample size ($n_{1\bullet} + n_{2\bullet}$) from the power function (5.4) assuming that $p_0 = 0.9$, $\theta^* = \log(1.25)$, $\theta_0 = 0.05$ and S_1^2 is used as V^2 if $S_1^2 > S_0^2$. If $n_2 = 0$, then stop here. Otherwise continue to next step.

4) Simulate d_{21k} (the difference in Y between Period 2 and Period 1 for the kth subject in Sequence 1 at Stage 2), k=1,2...,n_{21}, from a normal distribution with $E(d_{21k}) = \theta^* + dP$, and $V(d_{21k}) = V^2$, and d_{22k} (the difference in Y between Period 2 and Period 1 for the kth subject in Sequence 2 at Stage 2), l=1,2...,n_{22}, from a normal

distribution with $E(d_{22k}) = -\theta^* + dP$ and $V(d_{22k}) = V^2$. Note that

$n_{21} = (n_{1\bullet} + n_{2\bullet})/2 - n_{11}$ and $n_{22} = (n_{1\bullet} + n_{2\bullet})/2 - n_{12}$.

5) Compute S^*, S, T_{22}, and T_{22}^*.

6) Repeat Steps 1 to 5 many times, for example 50,000 times.

The null distributions of T_{22} and T_{22}^* under different parameters are compared in Tables 5.1 and 5.2. From these two tables, we can see that the quantiles of the null distributions of T_{22} and T_{22}^* match at the first decimal except for extreme quantiles and those quantiles become larger with smaller $n_{1\bullet}$ under the same other parameters. Comparing Table 5.1 and Table 5.2, we can see that the quantiles of the null distributions of T_{22} and T_{22}^* become larger with larger V^2 under the same n_1 and other parameters.

Table 5.1 Comparison of the null distributions under $\theta^* = \log(1.25)$ and $dP = 0.05$ between T_{22} and T_{22}^* when $S_0 = 0.2563$ ($n_{11} = n_{12} = 10$)

	V									
	0.25		0.3		0.4		0.45		0.5	
	T_{22}	T_{22}^*	T_{22}	T_{22}^*	T_{22}	T_{22}^*	T_{22}	T_{22}^*	T_{22}	T_{22}^*
0.1%	-2.810	-2.721	-3.222	-3.180	-3.426	-3.432	-3.435	-3.415	-3.118	-3.117
0.5%	-2.338	-2.238	-2.627	-2.582	-2.797	-2.785	-2.781	-2.781	-2.689	-2.693
1%	-2.107	-2.025	-2.366	-2.322	-2.507	-2.501	-2.471	-2.477	-2.446	-2.445
2%	-1.860	-1.799	-2.060	-2.024	-2.184	-2.174	-2.163	-2.163	-2.141	-2.143
5%	-1.469	-1.434	-1.637	-1.606	-1.717	-1.714	-1.703	-1.699	-1.731	-1.734
10%	-1.147	-1.115	-1.266	-1.239	-1.327	-1.331	-1.332	-1.335	-1.335	-1.337
50%	-0.002	0.000	0.011	0.008	0.000	0.000	0.001	0.001	0.002	0.002
90%	1.171	1.149	1.290	1.262	1.334	1.331	1.327	1.334	1.329	1.332
95%	1.493	1.467	1.659	1.627	1.701	1.704	1.718	1.720	1.717	1.714
98%	1.897	1.854	2.098	2.051	2.118	2.109	2.166	2.169	2.146	2.152
99%	2.142	2.083	2.374	2.309	2.414	2.408	2.477	2.479	2.443	2.452
99.5%	2.378	2.323	2.625	2.557	2.721	2.741	2.758	2.765	2.703	2.712
99.9%	2.753	2.780	3.182	3.140	3.442	3.453	3.349	3.388	3.360	3.382

Table 5.2 Comparison of the null distributions under $\theta^* = \log(1.25)$ and $dP = 0.05$ between T_{22} and T_{22}^* when $S_0 = 0.3665$ ($n_{11} = n_{12} = 20$)

Cumulative probability	V							
	0.3		0.4		0.5		0.52	
	T_{22}	T_{22}^*	T_{22}	T_{22}^*	T_{22}	T_{22}^*	T_{22}	T_{22}^*
0.1%	-2.473	-2.394	-3.037	-3.058	-3.286	-3.269	-3.252	-3.235
0.5%	-2.066	-2.040	-2.537	-2.503	-2.722	-2.697	-2.702	-2.697
1%	-1.843	-1.781	-2.301	-2.286	-2.443	-2.428	-2.379	-2.379
2%	-1.649	-1.607	-2.022	-1.989	-2.119	-2.128	-2.146	-2.139
5%	-1.329	-1.294	-1.612	-1.589	-1.685	-1.685	-1.684	-1.682
10%	-1.028	-1.006	-1.252	-1.232	-1.302	-1.301	-1.312	-1.312
50%	-0.010	-0.009	0.003	0.008	-0.005	-0.005	-0.002	-0.002
90%	1.013	1.006	1.229	1.210	1.317	1.317	1.311	1.313
95%	1.285	1.272	1.621	1.610	1.697	1.695	1.688	1.686
98%	1.616	1.580	1.993	1.970	2.111	2.116	2.122	2.133
99%	1.851	1.777	2.264	2.233	2.408	2.399	2.403	2.410
99.5%	2.017	1.958	2.495	2.460	2.723	2.731	2.715	2.718
99.9%	2.424	2.414	2.911	2.878	3.353	3.346	3.261	3.273

5.7.2 Comparison of the alternative distributions between T_{22} and T_{22}^*

To evaluate the alternative distributions of T_{22} and T_{22}^*, we use the same simulation schemes in Section 5.7.1 but assume that $\theta^* = 0.05$ and $dP = 0.05$. The results in Tables 5.3 and 5.4 indicate that the quantiles of the alternative distributions of T_{22} and T_{22}^* match at the first decimal except for 0.1% or 99.9% quantile and those quantiles become larger with smaller $n_{1\bullet}$ under same other parameters. Comparing Table 5.3 and Table 5.4, we can see that the quantiles of the alternative distributions of T_{22} and T_{22}^* become larger with larger V^2 under the same n_1 and other parameters.

134

Table 5.3 Comparison of the alternative distributions under $\theta^* = 0.05$ and $dP = 0.05$ between T_{22} and T_{22}^* when $S_0 = 0.2563$ ($n_{11} = n_{12} = 10$)

	V									
	0.25		0.3		0.4		0.45		0.5	
	T_{22}	T_{22}^*	T_{22}	T_{22}^*	T_{22}	T_{22}^*	T_{22}	T_{22}^*	T_{22}	T_{22}^*
0.1%	-6.211	-6.044	-6.408	-6.318	-6.715	-6.685	-6.725	-6.695	-6.627	-6.637
0.5%	-5.629	-5.627	-5.874	-5.790	-5.978	-5.982	-5.933	-5.977	-5.952	-5.936
1%	-5.379	-5.317	-5.610	-5.548	-5.666	-5.690	-5.648	-5.648	-5.650	-5.659
2%	-5.126	-5.065	-5.292	-5.254	-5.340	-5.370	-5.299	-5.306	-5.292	-5.304
5%	-4.736	-4.682	-4.861	-4.842	-4.868	-4.875	-4.828	-4.833	-4.821	-4.834
10%	-4.382	-4.347	-4.471	-4.445	-4.467	-4.478	-4.433	-4.446	-4.401	-4.404
50%	-3.190	-3.189	-3.145	-3.147	-3.056	-3.068	-3.010	-3.018	-2.993	-3.000
90%	-2.012	-2.069	-1.842	-1.866	-1.654	-1.671	-1.605	-1.612	-1.598	-1.598
95%	-1.674	-1.749	-1.474	-1.511	-1.255	-1.279	-1.213	-1.219	-1.189	-1.197
98%	-1.296	-1.355	-1.047	-1.098	-0.774	-0.791	-0.737	-0.747	-0.733	-0.743
99%	-1.071	-1.122	-0.772	-0.815	-0.454	-0.481	-0.415	-0.422	-0.450	-0.455
99.5%	-0.870	-0.923	-0.495	-0.549	-0.175	-0.183	-0.169	-0.172	-0.157	-0.155
99.9%	-0.405	-0.455	-0.018	-0.086	0.424	0.397	0.380	0.377	0.396	0.412

Table 5.4 Comparison of the alternative distributions under $\theta^* = 0.05$ and $dP = 0.05$ between T_{22} and T_{22}^* when $S_0 = 0.3665$ ($n_{11} = n_{12} = 20$)

Cumulative probability	V							
	0.3		0.4		0.5		0.52	
	T_{22}	T_{22}^*	T_{22}	T_{22}^*	T_{22}	T_{22}^*	T_{22}	T_{22}^*
0.1%	-5.648	-5.601	-6.091	-6.001	-6.383	-6.345	-6.283	-6.288
0.5%	-5.236	-5.231	-5.644	-5.586	-5.755	-5.784	-5.707	-5.730
1%	-4.994	-4.990	-5.422	-5.370	-5.545	-5.543	-5.455	-5.468
2%	-4.754	-4.733	-5.150	-5.102	-5.206	-5.208	-5.157	-5.165
5%	-4.406	-4.385	-4.705	-4.671	-4.748	-4.752	-4.736	-4.732
10%	-4.110	-4.090	-4.320	-4.308	-4.360	-4.361	-4.348	-4.345
50%	-3.073	-3.070	-3.042	-3.050	-3.013	-3.013	-2.992	-2.994
90%	-2.055	-2.074	-1.773	-1.790	-1.660	-1.663	-1.646	-1.643
95%	-1.761	-1.781	-1.422	-1.439	-1.275	-1.276	-1.238	-1.238
98%	-1.414	-1.449	-1.015	-1.032	-0.826	-0.830	-0.823	-0.822
99%	-1.203	-1.233	-0.721	-0.745	-0.507	-0.507	-0.537	-0.536
99.5%	-0.981	-1.034	-0.477	-0.493	-0.292	-0.292	-0.272	-0.274
99.9%	-0.546	-0.672	0.051	-0.007	0.186	0.185	0.350	0.351

5.8 Discussion and conclusions

Our simulations show that the null distributions of T_{22} and T_{22}^* are close to each other and the alternative distributions of T_{22} and T_{22}^* are also close to each other. Our simulations also show that the powers using (T_{21}, T_{22}) and (T_{21}^*, T_{22}^*) are matched at the second digit for all cases and the type I error rates using (T_{21}, T_{22}) and (T_{21}^*, T_{22}^*) are matched at the second digit for most of cases. Hence we do not lose much information if the simplified test statistics (T_{21}^*, T_{22}^*) is used instead of (T_{21}, T_{22}).

If $S_1 \leq S_0$, the critical value $t_\alpha(v_1)$ for T_{11} (or $-t_\alpha(v_1)$ for T_{12}) is used for the one-stage

study. Otherwise, the exact critical value u_α for T_{21}^* (or $-u_\alpha$ for T_{22}^*) is used for two-stage study. We calculate $P_{1s}(\theta_2, n_{11}, n_{12}, V)$ and $P_{2s}(\theta_2, n_{11}, n_{12}, V, p_0, u)$ for any given V and u with the numerical integration methods. The sum of all calculated levels of numerical error bounds at grid points and between grid points is about 10^{-4}. We find the bounded interval for V.

To search for u_α, we use the following strategy. The initial value for u is Z_α. If the maximum of $P_{1s}(\theta_2, n_{11}, n_{12}, V) + P_{2s}(\theta_2, n_{11}, n_{12}, V, p_0, u)$ in the bounded interval $[V_b, V_e]$ is greater than α, we increase u by 0.001. We continue this iteration until the difference between α and the maximum of $P_{1s}(\theta_2, n_{11}, n_{12}, V) + P_{2s}(\theta^*, n_{11}, n_{12}, V, p_0, u)$ is less than or equal to 10^{-4}. Due to all levels of errors, we can make $\alpha = 0.0495$.

Appendix 5.1 Details for Numerical integration method

1. Numerical method for calculation of $P_{1s}(\theta_2, n_{11}, n_{12}, V)$ and its accuracy

From Equation (5.11), we know $P_{1s}(\theta_2, n_{11}, n_{12}, V)$ is calculated by one-dimension integration. Now we use the Simpson's rule to obtain the numerical value for $P_{1s}(\theta_2, n_{11}, n_{12}, V)$ re-expressed as:

$$I = \int_0^{c_1} f^*(x)dx \cdot \qquad (5.16)$$

Here

$$f^*(x) = \left(\Phi\left(-\frac{t_\alpha(v_1)}{V} \bullet \sqrt{\frac{x}{v_1}} \right) - \Phi\left(\frac{\theta_1 - \theta_2}{\frac{V}{2}\sqrt{\frac{1}{n_{11}} + \frac{1}{n_{12}}}} + \frac{t_\alpha(v_1)}{V} \bullet \sqrt{\frac{x}{v_1}} \right) \right) \frac{1}{2^{v_1/2}\Gamma(v_1/2)V^{v_1}} x^{v_1/2-1} e^{-x/2V^2} \; .$$

Let $R_S^{P_1 s}(f^*)$ denote the numerical error for the integration. $\hat{I}$ is the numerical estimate for I.

In the Simpson's rule, we divide the inner interval $[0, c_1]$ into the even number (n_1) steps of length h_1 and obtain $I = \int_0^{c_1} f^*(x)dx = \hat{I} + R_S^{P_1 s}(f^*)$, where

$$\hat{I} = \frac{h_1}{3}\left(f^*(x_0) + 4\sum_{m=0}^{n/2-1} f^*(x_{2m+1}) + 2\sum_{m=1}^{n/2-1} f^*(x_{2m}) + f^*(x_n) \right), \qquad (5.17)$$

$$R_S^{P_1 s}(f^*) = \frac{c_1}{180} h_1^{\,4} \, \max_{0 \le \xi \le c_1} \left| \frac{\partial^4 f^*(\xi)}{\partial x^4} \right|, \qquad (5.18)$$

Here $x_0 = 0$, $x_i = i * h_1 \;\; \forall i = 1, 2, ..., n_1$. $R_S^{P_1 s}(f^*)$ calculated by (5.18) is an analytical

upper bound for the remainder [61] since it is calculated from $\max_{0 \le \xi \le c_1} \left| \frac{\partial^4 f^*(\xi)}{\partial x^4} \right|$,

which is the analytical and uniform upper bound in x on the bounded interval $[0, c_1]$

for the fourth derivative with respect to x of $f^*(x)$. $\max_{0 \le \xi \le c_1} \left| \frac{\partial^4 f^*(\xi)}{\partial x^4} \right|$ will be derived

in Appendix 5.2.

2. Numerical method for calculation of $P_{2s}(\theta_2, n_{11}, n_{12}, V, p_0, u)$ and its accuracy

In this section, we will discuss the Simpson's rule to obtain the numerical value for $P_{2s}(\theta_2, n_{11}, n_{12}, V, p_0, u)$ in Equation (5.13) which is reiterated as:

$$P_{2s}(\theta_2, n_{11}, n_{12}, V, p_0, u) = \int_{b}^{\infty}\int_{0}^{a} f(x, y)\,dy\,dx. \qquad (5.19)$$

Here,

$$f(x,y) = \left[\Phi\left(\frac{-u \cdot c_3 \cdot \sqrt{x+y}}{c_4 \cdot V}\right) - \Phi\left(\frac{\theta_1 - \theta_2 + u \cdot c_3 \cdot \sqrt{x+y}}{c_4 \cdot V}\right)\right] \frac{y^{\nu_2/2-1} x^{\nu_1/2-1}}{2^{(\nu_1+\nu_2)/2}\Gamma(\nu_1/2)\Gamma(\nu_2/2) \cdot V^{\nu_1+\nu_2}} e^{-\frac{x+y}{2V^2}}.$$

Since the above integration is an improper integration, then the relative large upper integration limit (denoted by b^*) in the above outer integration can be determined from the inequality $\int_{b^*}^{\infty} \frac{x^{\nu_1/2-1}}{\Gamma(\nu_1/2) \cdot (2V^2)^{\nu_1+\nu_2}} e^{-\frac{x}{2V^2}}\,dx \leq 10^{-6}$. After b^* is determined, we can use the Simpson's rule for obtaining the numerical value for $P_{2s}(\theta_2, n_{11}, n_{12}, V, p_0, u)$ with the finite integration.

Since $\dfrac{\theta_1 - \theta_2 + u \cdot c_3 \cdot \sqrt{x+y}}{c_4 \cdot V} \leq \dfrac{-u \cdot c_3 \cdot \sqrt{x+y}}{c_4 \cdot V} \leq 0$, then

$$0 \leq \Phi\left(\frac{-u \cdot c_3 \cdot \sqrt{x+y}}{c_4 \cdot V}\right) - \Phi\left(\frac{\theta_1 - \theta_2 + u \cdot c_3 \cdot \sqrt{x+y}}{c_4 \cdot V}\right) \leq 0.5.$$

In addition, $\displaystyle\int_{0}^{a} \frac{y^{\nu_2/2-1}}{2^{(\nu_2)/2}\Gamma(\nu_2/2) \cdot V^{\nu_2}} e^{-\frac{y}{2V^2}}\,dy \leq \int_{0}^{\infty} \frac{y^{\nu_2/2-1}}{\Gamma(\nu_2/2) \cdot (2V^2)^{\nu_2/2}} e^{-\frac{y}{2V^2}}\,dy = 1$

$\forall a > 0$. Hence

$$P_{2s}(\theta_2, n_{11}, n_{12}, V, p_0, u) - \int_b^{b^*}\int_0^a f(x,y)\,dy\,dx$$

$$\leq 0.5 \cdot \int_{b^*}^{\infty}\int_0^a \frac{y^{v_2/2-1} x^{v_1/2-1}}{2^{(v_1+v_2)/2}\Gamma(v_1/2)\Gamma(v_2/2)\cdot V^{v_1+v_2}} e^{-\frac{x+y}{2V^2}}\,dy\,dx$$

$$\leq 0.5 \cdot \int_{b^*}^{\infty} \frac{x^{v_1/2-1}}{\Gamma(v_1/2)\cdot(2V^2)^{v_1+v_2}} e^{-\frac{x}{2V^2}}\,dx$$

$$\leq 0.5 \cdot \int_{b^*}^{\infty} \frac{x^{v_1/2-1}}{\Gamma(v_1/2)\cdot(2V^2)^{v_1+v_2}} e^{-\frac{x}{2V^2}}\,dx \leq 5\cdot10^{-7}$$

Let R^{imp} be the upper bound for difference between the improper integration and the finite integration with upper limit b^*. Then R^{imp} is less than or equal to $5\cdot10^{-7}$.

We define $I_1(x) = \int_0^a f(x,y)\,dy$ and $I_2 = \int_b^{b^*} I_1(x)\,dx$. Let $R_S^{inner}(x,f)$ denote the numerical error for the inner integration. $\hat{I}_1(x)$ is the numerical estimate for $I_1(x)$.

In the Simpson's rule, we divide the inner interval $[0, a]$ into the even number (n) steps of length h and obtain $I_1(x) = \int_0^a f(x,y)\,dy = \hat{I}_1(x) + R_S^{inner}(x,f)$, where

$$\hat{I}_1(x) = \frac{h}{3}\left(f(x,y_0) + 4\sum_{m=0}^{n/2-1} f(x,y_{2m+1}) + 2\sum_{m=1}^{n/2-1} f(x,y_{2m}) + f(x,y_n) \right), \quad (5.20)$$

$$R_S^{inner}(x,f) = \frac{a}{180} h^4 \max_{0\leq\xi\leq a}\left|\frac{\partial^4 f(x,\xi)}{\partial y^4}\right|, \qquad (5.21)$$

Here $y_0 = 0$, and $y_i = i*h \ \forall i = 1,2,...,n$.

The numerical approximation error bound, $R_S^{inner}(x, f)$ is an analytical upper bound

since it is calculated from $\max_{0 \leq \xi \leq a} \left| \dfrac{\partial^4 f(x, \xi)}{\partial y^4} \right|$, which is the analytical and uniform

upper bound in y on the bounded interval $[0, a]$ for the fourth derivative with respect

to y of $f(x, y)$. $\max_{0 \leq \xi \leq a} \left| \dfrac{\partial^4 f(x, \xi)}{\partial y^4} \right|$ will be derived in Appendix 5.3.

We approximate $I_3 = \int_b^{b^*} R_S^{inner}(x, f) dx$ by $\hat{I}_3$ using the trapezoidal rule. In the

trapezoidal rule, we divide the outer interval $[b, b^*]$ into the number (n^*) steps of

length h^* and obtain $I_3 = \int_b^{b^*} R_S^{inner}(x, f) dx = \hat{I}_3(x) + R(\hat{I}_3(\zeta))$, where

$$\hat{I}_3 = h^* \left(R_S^{inner}(x_0, f)/2 + R_S^{inner}(x_{n^*}, f)/2 + \sum_{m=1}^{n^*-1} R_S^{inner}(x_m, f) \right), \quad (5.22)$$

$$R(\hat{I}_3(\zeta)) = \frac{b^* - b}{12} h^{*2} \max_{b \leq \varsigma \leq b^*} \left| \frac{\partial^2 R_S^{inner}(\zeta)}{\partial x^2} \right|, \quad (5.23)$$

$x_0 = 0$, $x_i = i * h^* \ \forall i = 1, 2, ..., n^*$, and $\max_{b \leq \varsigma \leq b^*} \left| \dfrac{\partial^2 R_S^{inner}(\zeta)}{\partial x^2} \right|$ is bounded numerical

second-order derivative.

Let $R_S^{outer}(\hat{I}_1(\varsigma))$ denote the numerical error for the outer integration of $\hat{I}_1$. $\hat{I}_2$ is the

numerical estimate for I_2.

141

In the trapezoidal rule, we divide the outer interval $[b, b^*]$ into the number (n^*)

steps of length h^* and obtain $I_2 = \int_b^{b^*} I_1(x)dx = \int_b^{b^*} \hat{I}_1(x)dx + \int_b^{b^*} R_S^{inner}(x, f)dx$, where

$$\int_b^{b^*} \hat{I}_1(x)dx = h^*\left(\hat{I}_1(x_0)/2 + \hat{I}_1(x_{n^*})/2 + \sum_{m=1}^{n^*-1}\hat{I}_1(x_m)\right) + R_S^{outer}(\hat{I}_1(\varsigma)), \qquad (5.24)$$

$$R_S^{outer}(\hat{I}_1(\varsigma)) = \frac{b^* - b}{12}h^{*2}\max_{b \le \varsigma \le b^*}\left|\frac{\partial^2 \hat{I}_1(\varsigma)}{\partial x^2}\right|,$$

(5.25)

$x_0 = 0$, $x_i = i*h^*$ $\forall i = 1, 2, ..., n^*$, and $\max_{b \le \varsigma \le b^*}\left|\dfrac{\partial^2 \hat{I}_1(\varsigma)}{\partial x^2}\right|$ is bounded numerical

second-order derivative.

$R(\hat{I}_3(\varsigma))$ and $R_S^{outer}(\hat{I}_1(\varsigma))$ are approximate bounds depending on the numerical

second-order derivative of the integrand function. $\int_b^{b^*} R_S^{inner}(x, f)dx$ is the numerical

integral of the exact analytical bound for the inner integral.

Let $R_S^{P_{2s}}$ be the total integration error in calculation of $P_{2s}(\theta_2, n_{11}, n_{12}, V, p_0, u)$. Then

$$R_S^{P_{2s}} = R^{imp} + R_S^{outer}(\hat{I}_1(\varsigma)) + \int_b^{b^*} R_S^{inner}(x, f)dx + R(\hat{I}_3(\varsigma)).$$

3. Numerical accuracy for $P_{1s}(\theta_2, n_{11}, n_{12}, V) + P_{2s}(\theta_2, n_{11}, n_{12}, V, p_0, u)$

Let R_S be the total integration error in calculation of $P_{1s}(\theta_2, n_{11}, n_{12}, V)$ and

$P_{2s}(\theta_2, n_{11}, n_{12}, V, p_0, u)$ for fixed V. Then $R_S = R_S^{P_{1s}} + R_S^{P_{2s}}$.

This appendix provides the details for obtaining the upper bound of $\left|\dfrac{\partial^4 f^*(\xi)}{\partial x^4}\right|$. In

order to derive the fourth-order derivative of the integrand $f^*(x)$, we

$$\text{let } g_1(x)=\Phi\left(-\frac{t_\alpha(v_1)}{V}\bullet\sqrt{\frac{x}{v_1}}\right)-\Phi\left(\frac{\theta_1-\theta_2}{\dfrac{V}{2}\sqrt{\dfrac{1}{n_{11}}+\dfrac{1}{n_{12}}}}+\frac{t_\alpha(v_1)}{V}\bullet\sqrt{\frac{x}{v_1}}\right),\text{ and}$$

$$g_2(x)=\frac{x^{v_1/2-1}}{2^{v_1/2}\Gamma(v_1/2)\bullet V^{v_1}}e^{-\frac{x}{2V^2}}.\text{ Then } f^*(x)=g_1(x)\bullet g_2(x).$$

From the chain rule of differentiation, the fourth-order derivative of $f^*(x)$ with

respect to x is written as

$$\frac{d^4 f^*(x)}{dx^4}=\frac{d^4 g_1(x)}{dx^4}g_2(x)+4\frac{d^3 g_1(x)}{dx^3}\frac{dg_2(x)}{dx}+6\frac{d^2 g_1(x)}{dx^2}\frac{d^2 g_2(x)}{dx^2}+4\frac{d^3 g_2(x)}{dx^3}\frac{dg_1(x)}{dx}+\frac{d^4 g_2(x)}{dx^4}g_1(x)$$

.

Clearly we have to obtain all up to fourth-order derivatives of both $g_1(x)$ and $g_2(x)$

with respect to x.

$$\frac{dg_1(x)}{dx}=\phi\left(-\frac{t_\alpha(v_1)}{V}\bullet\sqrt{\frac{x}{v_1}}\right)\frac{-t_\alpha(v_1)}{2V}\bullet\sqrt{\frac{1}{v_1 x}}-\phi\left(\frac{\theta_1-\theta_2}{\dfrac{V}{2}\sqrt{\dfrac{1}{n_{11}}+\dfrac{1}{n_{12}}}}+\frac{t_\alpha(v_1)}{V}\bullet\sqrt{\frac{x}{v_1}}\right)\frac{t_\alpha(v_1)}{2V}\bullet\sqrt{\frac{1}{v_1 x}}$$

.

$$\frac{d^2 g_1(x)}{dx^2} = \phi\left(-\frac{t_\alpha(V_1)}{V} \bullet \sqrt{\frac{x}{V_1}}\right)\left(\frac{t_\alpha(V_1)}{V\sqrt{V_1}}\right)^3 \bullet \frac{1}{4\sqrt{x}} + \phi\left(-\frac{t_\alpha(V_1)}{V} \bullet \sqrt{\frac{x}{V_1}}\right)\frac{t_\alpha(V_1)}{V\sqrt{V_1}} \bullet \frac{1}{4\sqrt{x^3}}$$

$$+ \phi\left(\frac{\theta_1 - \theta_2}{\frac{V}{2}\sqrt{\frac{1}{n_{11}} + \frac{1}{n_{12}}}} + \frac{t_\alpha(V_1)}{V} \bullet \sqrt{\frac{x}{V_1}}\right)\left(\frac{t_\alpha(V_1)}{V\sqrt{V_1}}\right)^3 \bullet \frac{1}{4\sqrt{x}} + \phi\left(\frac{\theta_1 - \theta_2}{\frac{V}{2}\sqrt{\frac{1}{n_{11}} + \frac{1}{n_{12}}}} + \frac{t_\alpha(V_1)}{V} \bullet \sqrt{\frac{x}{V_1}}\right)\frac{t_\alpha(V_1)}{V\sqrt{V_1}} \bullet \frac{1}{4\sqrt{x^3}}$$

.

$$\frac{d^3 g_1(x)}{dx^3} = -\phi\left(-\frac{t_\alpha(V_1)}{V} \bullet \sqrt{\frac{x}{V_1}}\right)\left(\frac{t_\alpha(V_1)}{V\sqrt{V_1}}\right)^5 \bullet \frac{1}{8\sqrt{x}} - \phi\left(-\frac{t_\alpha(V_1)}{V} \bullet \sqrt{\frac{x}{V_1}}\right)\left(\frac{t_\alpha(V_1)}{V\sqrt{V_1}}\right)^3 \bullet \frac{1}{4\sqrt{x^3}}$$

$$- \phi\left(-\frac{t_\alpha(V_1)}{V} \bullet \sqrt{\frac{x}{V_1}}\right)\frac{t_\alpha(V_1)}{V\sqrt{V_1}} \bullet \frac{3}{8\sqrt{x^5}} - \phi\left(\frac{\theta_1 - \theta_2}{\frac{V}{2}\sqrt{\frac{1}{n_{11}} + \frac{1}{n_{12}}}} + \frac{t_\alpha(V_1)}{V} \bullet \sqrt{\frac{x}{V_1}}\right)\left(\frac{t_\alpha(V_1)}{V\sqrt{V_1}}\right)^5 \bullet \frac{1}{8\sqrt{x}}$$

$$- \phi\left(\frac{\theta_1 - \theta_2}{\frac{V}{2}\sqrt{\frac{1}{n_{11}} + \frac{1}{n_{12}}}} + \frac{t_\alpha(V_1)}{V} \bullet \sqrt{\frac{x}{V_1}}\right)\left(\frac{t_\alpha(V_1)}{V\sqrt{V_1}}\right)^3 \bullet \frac{1}{4\sqrt{x^3}} - \phi\left(\frac{\theta_1 - \theta_2}{\frac{V}{2}\sqrt{\frac{1}{n_{11}} + \frac{1}{n_{12}}}} + \frac{t_\alpha(V_1)}{V} \bullet \sqrt{\frac{x}{V_1}}\right)\frac{t_\alpha(V_1)}{V\sqrt{V_1}} \bullet \frac{3}{8\sqrt{x^5}}$$

$$\frac{d^4 g_1(x)}{\partial x^4} = \phi\left(-\frac{t_\alpha(V_1)}{V} \bullet \sqrt{\frac{x}{V_1}}\right)\left(\frac{t_\alpha(V_1)}{V\sqrt{V_1}}\right)^7 \frac{1}{16\sqrt{x}} + \phi\left(-\frac{t_\alpha(V_1)}{V} \bullet \sqrt{\frac{x}{V_1}}\right)\left(\frac{t_\alpha(V_1)}{V\sqrt{V_1}}\right)^5 \frac{3}{16\sqrt{x^3}}$$

$$+ \phi\left(-\frac{t_\alpha(V_1)}{V} \bullet \sqrt{\frac{x}{V_1}}\right)\left(\frac{t_\alpha(V_1)}{V\sqrt{V_1}}\right)^3 \frac{9}{16\sqrt{x^5}} + \phi\left(-\frac{t_\alpha(V_1)}{V} \bullet \sqrt{\frac{x}{V_1}}\right)\left(\frac{t_\alpha(V_1)}{V\sqrt{V_1}}\right)\frac{15}{16\sqrt{x^7}}$$

$$+ \phi\left(\frac{\theta_1 - \theta_2}{\frac{V}{2}\sqrt{\frac{1}{n_{11}} + \frac{1}{n_{12}}}} + \frac{t_\alpha(V_1)}{V} \bullet \sqrt{\frac{x}{V_1}}\right)\left(\frac{t_\alpha(V_1)}{V\sqrt{V_1}}\right)^7 \frac{1}{16\sqrt{x}} + \phi\left(\frac{\theta_1 - \theta_2}{\frac{V}{2}\sqrt{\frac{1}{n_{11}} + \frac{1}{n_{12}}}} + \frac{t_\alpha(V_1)}{V} \bullet \sqrt{\frac{x}{V_1}}\right)\left(\frac{t_\alpha(V_1)}{V\sqrt{V_1}}\right)^5 \frac{3}{16\sqrt{x^3}}$$

$$+ \phi\left(\frac{\theta_1 - \theta_2}{\frac{V}{2}\sqrt{\frac{1}{n_{11}} + \frac{1}{n_{12}}}} + \frac{t_\alpha(V_1)}{V} \bullet \sqrt{\frac{x}{V_1}}\right)\left(\frac{t_\alpha(V_1)}{V\sqrt{V_1}}\right)^3 \frac{9}{16\sqrt{x^5}} + \phi\left(\frac{\theta_1 - \theta_2}{\frac{V}{2}\sqrt{\frac{1}{n_{11}} + \frac{1}{n_{12}}}} + \frac{t_\alpha(V_1)}{V} \bullet \sqrt{\frac{x}{V_1}}\right)\left(\frac{t_\alpha(V_1)}{V\sqrt{V_1}}\right)\frac{15}{16\sqrt{x^7}}$$

.

$$\frac{dg_2(x)}{dx} = \frac{(V_1/2 - 1)x^{V_1/2 - 2}}{2^{V_1/2}\Gamma(V_1/2) \cdot V^{V_1}} e^{-\frac{x}{2V^2}} - \frac{x^{V_1/2 - 1}}{2^{(V_1+2)/2}\Gamma(V_1/2) \cdot V^{V_1+2}} e^{-\frac{x}{2V^2}} .$$

$$\frac{d^2 g_2(x)}{dx^2} = \frac{(v_1/2-1)(v_1/2-2)x^{v_1/2-3}}{2^{v_1/2}\Gamma(v_1/2)\cdot V^{v_1}}e^{-\frac{x}{2V^2}} - \frac{(v_1/2-1)x^{v_1/2-2}}{2^{(v_1+2)/2}\Gamma(v_1/2)\cdot V^{v_1+2}}e^{-\frac{x}{2V^2}}$$
$$+ \frac{x^{v_1/2-1}}{2^{(v_1+4)/2}\Gamma(v_1/2)\cdot V^{v_1+4}}e^{-\frac{x}{2V^2}}.$$

$$\frac{d^3 g_2(x)}{dx^3} = \frac{(v_1/2-1)(v_1/2-2)(v_1/2-3)x^{v_1/2-4}}{2^{(v_1)/2}\Gamma(v_1/2)\cdot V^{v_1}}e^{-\frac{x}{2V^2}} - \frac{3(v_1/2-1)(v_1/2-2)x^{v_1/2-3}}{2^{(v_1+2)/2}\Gamma(v_1/2)\cdot V^{v_1+2}}e^{-\frac{x}{2V^2}}$$
$$+ \frac{3(v_1/2-1)x^{v_1/2-2}}{2^{(v_1+4)/2}\Gamma(v_1/2)\cdot V^{v_1+4}}e^{-\frac{x}{2V^2}} - \frac{x^{v_1/2-1}}{2^{(v_1+6)/2}\Gamma(v_1/2)\cdot V^{v_1+6}}e^{-\frac{x}{2V^2}}.$$

$$\frac{d^4 g_2(x)}{dx^4} = \frac{(v_1/2-1)(v_1/2-2)(v_1/2-3)(v_1/2-4)x^{v_1/2-5}}{2^{(v_1)/2}\Gamma(v_1/2)\cdot V^{v_1}}e^{-\frac{x}{2V^2}}$$
$$- \frac{(v_1/2-1)(v_1/2-2)(v_1/2-3)x^{v_1/2-4}}{2^{(v_1-2)/2}\Gamma(v_1/2)\cdot V^{v_1+2}}e^{-\frac{x}{2V^2}} + \frac{3(v_1/2-1)(v_1/2-2)x^{v_1/2-3}}{2^{(v_1+2)/2}\Gamma(v_1/2)\cdot V^{v_1+4}}e^{-\frac{x}{2V^2}}.$$
$$- \frac{(v_1/2-1)x^{v_1/2-2}}{2^{(v_1+2)/2}\Gamma(v_1/2)\cdot V^{v_1+6}}e^{-\frac{x}{2V^2}} + \frac{x^{v_1/2-1}}{2^{(v_1+8)/2}\Gamma(v_1/2)\cdot V^{v_1+8}}e^{-\frac{x}{2V^2}}$$

Since $\Phi\left(\dfrac{\theta_1-\theta_2}{\dfrac{V}{2}\sqrt{\dfrac{1}{n_{11}}+\dfrac{1}{n_{12}}}}+\dfrac{t_\alpha(v_1)}{V}\bullet\sqrt{\dfrac{x}{v_1}}\right)\geq 0$, then

$$g_1(x) \leq \Phi\left(-\frac{t_\alpha(v_1)}{V}\bullet\sqrt{\frac{x}{v_1}}\right)\leq 0.5 \ \forall x\geq 0.$$

$$\left|\frac{dg_1(x)}{dx}\sqrt{x}\right| = \phi\left(-\frac{t_\alpha(v_1)}{V}\bullet\sqrt{\frac{x}{v_1}}\right)\frac{t_\alpha(v_1)}{2V}\bullet\sqrt{\frac{1}{v_1}} + \phi\left(\frac{\theta_1-\theta_2}{\dfrac{V}{2}\sqrt{\dfrac{1}{n_{11}}+\dfrac{1}{n_{12}}}}+\frac{t_\alpha(v_1)}{V}\bullet\sqrt{\frac{x}{v_1}}\right)\frac{t_\alpha(v_1)}{2V}\bullet\sqrt{\frac{1}{v_1}}$$

$$\leq \frac{t_\alpha(v_1)}{\sqrt{2\pi}V}\bullet\sqrt{\frac{1}{v_1}}$$

, $\forall x\in[0,c_1]$, due to the fact that

$$\phi\left(\frac{\theta_1-\theta_2}{\frac{V}{2}\sqrt{\frac{1}{n_{11}}+\frac{1}{n_{12}}}}+\frac{t_\alpha(v_1)}{V}\bullet\sqrt{\frac{x}{v_1}}\right)\le\phi\left(-\frac{t_\alpha(v_1)}{V}\bullet\sqrt{\frac{x}{v_1}}\right)\text{ when }$$

$$\frac{\theta_1-\theta_2}{\frac{V}{2}\sqrt{\frac{1}{n_{11}}+\frac{1}{n_{12}}}}+\frac{t_\alpha(v_1)}{V}\bullet\sqrt{\frac{x}{v_1}}\le-\frac{t_\alpha(v_1)}{V}\bullet\sqrt{\frac{x}{v_1}}\le0.$$

Similarly,

$$\left|\frac{d^2g_1(x)}{dx^2}\sqrt{x^3}\right|\le\frac{1}{\sqrt{2\pi}}\left(\frac{t_\alpha(v_1)}{V\sqrt{v_1}}\right)^3\bullet\frac{c_1}{2}+\frac{1}{\sqrt{2\pi}}\frac{t_\alpha(v_1)}{V\sqrt{v_1}}\bullet\frac{1}{2},$$

$$\left|\frac{d^3g_1(x)}{dx^3}\sqrt{x^5}\right|\le\frac{1}{\sqrt{2\pi}}\left(\frac{t_\alpha(v_1)}{V\sqrt{v_1}}\right)^5\bullet\frac{c_1^2}{4}+\frac{1}{\sqrt{2\pi}}\left(\frac{t_\alpha(v_1)}{V\sqrt{v_1}}\right)^3\bullet\frac{c_1}{2}+\frac{1}{\sqrt{2\pi}}\frac{t_\alpha(v_1)}{V\sqrt{v_1}}\bullet\frac{3}{4}\text{, and}$$

$$\left|\frac{d^4g_1(x)}{\partial x^4}\sqrt{x^7}\right|\le\frac{1}{\sqrt{2\pi}}\left(\frac{t_\alpha(v_1)}{V\sqrt{v_1}}\right)^7\frac{c_1^3}{8}+\frac{1}{\sqrt{2\pi}}\left(\frac{t_\alpha(v_1)}{V\sqrt{v_1}}\right)^5\frac{3c_1^2}{8}+\frac{1}{\sqrt{2\pi}}\left(\frac{t_\alpha(v_1)}{V\sqrt{v_1}}\right)^3\frac{9c_1}{16}+\frac{1}{\sqrt{2\pi}}\left(\frac{t_\alpha(v_1)}{V\sqrt{v_1}}\right)\frac{15}{8}.$$

Since $g_2(x)\sqrt{x^{-7}}=\dfrac{x^{(v_1-7)/2-1}}{2^{(v_1)/2}\Gamma(v_1/2)\cdot V^{v_1}}e^{-\frac{x}{2V^2}}$ achieves the maximum value at

$x=2V^2((v_1-7)/2-1)$, then $\left|g_2(x)\sqrt{x^{-7}}\right|\le\left|\dfrac{[(v_1-7)/2-1]^{(v_1-7)/2-1}}{2^{9/2}\Gamma(v_1/2)\cdot V^9}e^{-\frac{(v_1-7)/2-1}{2V^2}}\right|.$

Similarly,

$$\left|\frac{dg_2(x)}{dx}\sqrt{x^{-5}}\right|\le\left|\frac{(v_1/2-1)((v_1-5)/2-2)^{(v_1-5)/2-2}}{2^{9/2}\Gamma(v_1/2)\cdot V^9}e^{-\frac{(v_1-5)/2-2}{2V^2}}\right|$$

$$+\left|\frac{((v_1-5)/2-1)^{(v_1-5)/2-1}}{2^{9/2}\Gamma(v_1/2)\cdot V^9}e^{-\frac{(v_1-5)/2-1}{2V^2}}\right|$$

$$\left|\frac{d^2 g_2(x)}{dx^2}\sqrt{x^{-3}}\right| \leq \left|\frac{(\nu_1/2-1)(\nu_1/2-2)[(\nu_1-3)/2-3]^{(\nu_1-3)/2-3}}{2^{9/2}\Gamma(\nu_1/2)\cdot V^9}e^{-\frac{(\nu_1-3)/2-3}{2V^2}}\right|$$

$$+\left|\frac{(\nu_1/2-1)[(\nu_1-3)/2-2]^{(\nu_1-3)/2-2}}{2^{9/2}\Gamma(\nu_1/2)\cdot V^9}e^{-\frac{(\nu_1-3)/2-2}{2V^2}}\right| + \left|\frac{[(\nu_1-3)/2-1]^{(\nu_1-3)/2-1}}{2^{9/2}\Gamma(\nu_1/2)\cdot V^9}e^{-\frac{(\nu_1-3)/2-1}{2V^2}}\right| \quad,$$

$$\left|\frac{d^3 g_2(x)}{dx^3}\sqrt{x^{-1}}\right| \leq \left|\frac{(\nu_1/2-1)(\nu_1/2-2)(\nu_1/2-3)((\nu_1-1)/2-4)^{(\nu_1-1)/2-4}}{2^{9/2}\Gamma(\nu_1/2)\cdot V^9}e^{-\frac{(\nu_1-1)/2-4}{2V^2}}\right|$$

$$+\left|\frac{3(\nu_1/2-1)(\nu_1/2-2)((\nu_1-1)/2-3)^{(\nu_1-1)/2-3}}{2^{9/2}\Gamma(\nu_1/2)\cdot V^9}e^{-\frac{(\nu_1-1)/2-3}{2V^2}}\right| \quad, \text{ and}$$

$$+\left|\frac{3(\nu_1/2-1)[(\nu_1-1)/2-2]^{(\nu_1-1)/2-2}}{2^{9/2}\Gamma(\nu_1/2)\cdot V^9}e^{-\frac{(\nu_1-1)/2-2}{2V^2}}\right| + \left|\frac{[(\nu_1-1)/2-1]^{[\nu_1-1]/2-1}}{2^{9/2}\Gamma(\nu_1/2)\cdot V^9}e^{-\frac{(\nu_1-1)/2-1}{2V^2}}\right|$$

$$\left|\frac{d^4 g_2(x)}{dx^4}\right| \leq \left|\frac{(\nu_1/2-1)(\nu_1/2-2)(\nu_1/2-3)(\nu_1/2-4)(\nu_1/2-5)^{\nu_1/2-5}}{2^5\Gamma(\nu_1/2)\cdot V^{10}}e^{-\frac{\nu_1/2-5}{2V^2}}\right|$$

$$+\left|\frac{(\nu_1/2-1)(\nu_1/2-2)(\nu_1/2-3)(\nu_1/2-4)^{\nu_1/2-4}}{2^3\Gamma(\nu_1/2)\cdot V^{10}}e^{-\frac{(\nu_1/2-4)}{2V^2}}\right| + \left|\frac{3(\nu_1/2-1)(\nu_1/2-2)(\nu_1/2-3)^{\nu_1/2-3}}{2^4\Gamma(\nu_1/2)\cdot V^{10}}e^{-\frac{(\nu_1/2-3)}{2V^2}}\right|$$

$$+\left|\frac{(\nu_1/2-1)(\nu_1/2-2)^{\nu_1/2-2}}{2^3\Gamma(\nu_1/2)\cdot V^{10}}e^{-\frac{(\nu_1/2-2)}{2V^2}}\right| + \left|\frac{(\nu_1/2-1)^{\nu_1/2-1}}{2^5\Gamma(\nu_1/2)\cdot V^{10}}e^{-\frac{(\nu_1/2-1)}{2V^2}}\right|$$

Hence, we have derived the upper bound for $\left|\dfrac{d^4 f^*(x)}{dx^4}\right|$ as:

$$\left|\frac{d^4 f^*(x)}{dx^4}\right| \leq \left|\frac{d^4 g_1(x)}{dx^4}\sqrt{x^7}\right|\left|g_2(x,y)\sqrt{x^{-7}}\right| + 4\left|\frac{d^3 g_1(x)}{dx^3}\sqrt{x^5}\right|\left|\frac{dg_2(x)}{dx}\sqrt{x^{-5}}\right| + 6\left|\frac{d^2 g_1(x)}{dx^2}\sqrt{x^3}\right|\left|\frac{d^2 g_2(x)}{dx^2}\sqrt{x^{-3}}\right|,$$

$$+4\left|\frac{dg_1(x)}{dx}\sqrt{x}\right|\left|\frac{d^3 g_2(x)}{dx^3}\sqrt{x^{-1}}\right| + \left|\frac{d^4 g_2(x)}{\partial x^4}\right|\left|g_1(x)\right|$$

$$\forall x \in [0, c_1].$$

Appendix 5.3 Derivation of the analytical upper bound for $\dfrac{\partial^4}{\partial x^4}\left(f(x,\xi)\right)$

This appendix provides the details for obtaining the upper bound of $\left|\dfrac{\partial^4 f(x,\xi)}{\partial y^4}\right|$. In

order to derive the fourth-order derivative of the integrand $f(x,y)$, we

let $g_1(x,y)=\Phi\!\left(\dfrac{-u\cdot c_3\cdot\sqrt{x+y}}{c_4\cdot V}\right)-\Phi\!\left(\dfrac{\theta_1-\theta_2+u\cdot c_3\cdot\sqrt{x+y}}{c_4\cdot V}\right)$, and

$g_2(x,y)=\dfrac{y^{\nu_2/2-1}x^{\nu_1/2-1}}{2^{(\nu_1+\nu_2)/2}\Gamma(\nu_1/2)\Gamma(\nu_2/2)\cdot V^{\nu_1+\nu_2}}e^{-\frac{x+y}{2V^2}}$. Then $f(x,y)=g_1(x,y)\cdot g_2(x,y)$.

From the chain rule of differentiation, the partial fourth-order derivative of

$f(x,y)$ with respect to y is written as

$$
\frac{\partial^4 f(x,y)}{\partial y^4}=\frac{\partial^4 g_1(x,y)}{\partial y^4}g_2(x,y)+4\frac{\partial^3 g_1(x,y)}{\partial y^3}\frac{\partial g_2(x,y)}{\partial y}+6\frac{\partial^2 g_1(x,y)}{\partial y^2}\frac{\partial^2 g_2(x,y)}{\partial y^2}
$$
$$
+4\frac{\partial^3 g_2(x,y)}{\partial y^3}\frac{\partial g_1(x,y)}{\partial y}+\frac{\partial^4 g_2(x,y)}{\partial y^4}g_1(x,y)
$$
.

Clearly we have to obtain all up to fourth-order partial derivatives of

both $g_1(x,y)$ and $g_2(x,y)$ with respect to y.

$$
\frac{\partial g_1(x,y)}{\partial y}=\phi\!\left(\frac{-u\cdot c_3\cdot\sqrt{x+y}}{c_4\cdot V}\right)\frac{-u\cdot c_3}{2c_4\cdot V\sqrt{x+y}}-\phi\!\left(\frac{\theta_1-\theta_2+u\cdot c_3\cdot\sqrt{x+y}}{c_4\cdot V}\right)\frac{u\cdot c_3}{2c_4\cdot V\sqrt{x+y}}
$$
.

$$
\frac{\partial^2 g_1(x,y)}{\partial y^2}=\phi\!\left(\frac{-u\cdot c_3\cdot\sqrt{x+y}}{c_4\cdot V}\right)\!\left(\frac{u\cdot c_3}{c_4\cdot V}\right)^3\frac{1}{4\sqrt{x+y}}+\phi\!\left(\frac{-u\cdot c_3\cdot\sqrt{x+y}}{c_4\cdot V}\right)\frac{u\cdot c_3}{c_4\cdot V}\frac{1}{4\sqrt{(x+y)^3}}
$$
$$
+\phi\!\left(\frac{\theta_1-\theta_2+u\cdot c_3\cdot\sqrt{x+y}}{c_4\cdot V}\right)\!\left(\frac{u\cdot c_3}{c_4\cdot V}\right)^3\frac{1}{4\sqrt{x+y}}+\phi\!\left(\frac{\theta_1-\theta_2+u\cdot c_3\cdot\sqrt{x+y}}{c_4\cdot V}\right)\frac{u\cdot c_3}{c_4\cdot V}\frac{1}{4\sqrt{(x+y)^3}}
$$
.

$$\frac{\partial^3 g_1(x,y)}{\partial y^3} = -\phi\left(\frac{-u\cdot c_3 \cdot \sqrt{x+y}}{c_4 \cdot V}\right)\left(\frac{u\cdot c_3}{c_4 \cdot V}\right)^5 \frac{1}{8\sqrt{x+y}} - \phi\left(\frac{-u\cdot c_3 \cdot \sqrt{x+y}}{c_4 \cdot V}\right)\left(\frac{u\cdot c_3}{c_4 \cdot V}\right)^3 \frac{1}{4\sqrt{(x+y)^3}}$$
$$-\phi\left(\frac{-u\cdot c_3 \cdot \sqrt{x+y}}{c_4 \cdot V}\right)\frac{3u\cdot c_3}{c_4 \cdot V}\frac{1}{8\sqrt{(x+y)^5}} - \phi\left(\frac{\theta_1 - \theta_2 + u\cdot c_3 \cdot \sqrt{x+y}}{c_4 \cdot V}\right)\left(\frac{u\cdot c_3}{c_4 \cdot V}\right)^5 \frac{1}{8\sqrt{x+y}}$$
$$-\phi\left(\frac{\theta_1 - \theta_2 + u\cdot c_3 \cdot \sqrt{x+y}}{c_4 \cdot V}\right)\left(\frac{u\cdot c_3}{c_4 \cdot V}\right)^3 \frac{1}{4\sqrt{(x+y)^3}} - \phi\left(\frac{\theta_1 - \theta_2 + u\cdot c_3 \cdot \sqrt{x+y}}{c_4 \cdot V}\right)\frac{3u\cdot c_3}{c_4 \cdot V}\frac{1}{8\sqrt{(x+y)^5}}$$

$$\frac{\partial^4 g_1(x,y)}{\partial y^4} = \phi\left(\frac{-u\cdot c_3 \cdot \sqrt{x+y}}{c_4 \cdot V}\right)\left(\frac{u\cdot c_3}{c_4 \cdot V}\right)^7 \frac{1}{16\sqrt{x+y}} + \phi\left(\frac{-u\cdot c_3 \cdot \sqrt{x+y}}{c_4 \cdot V}\right)\left(\frac{u\cdot c_3}{c_4 \cdot V}\right)^5 \frac{3}{16\sqrt{(x+y)^3}}$$
$$+\phi\left(\frac{-u\cdot c_3 \cdot \sqrt{x+y}}{c_4 \cdot V}\right)\left(\frac{u\cdot c_3}{c_4 \cdot V}\right)^3 \frac{9}{16\sqrt{(x+y)^5}} + \phi\left(\frac{-u\cdot c_3 \cdot \sqrt{x+y}}{c_4 \cdot V}\right)\left(\frac{u\cdot c_3}{c_4 \cdot V}\right)\frac{15}{16\sqrt{(x+y)^7}}$$
$$+\phi\left(\frac{\theta_1 - \theta_2 + u\cdot c_3 \cdot \sqrt{x+y}}{c_4 \cdot V}\right)\left(\frac{u\cdot c_3}{c_4 \cdot V}\right)^7 \frac{1}{16\sqrt{x+y}} + \phi\left(\frac{\theta_1 - \theta_2 + u\cdot c_3 \cdot \sqrt{x+y}}{c_4 \cdot V}\right)\left(\frac{u\cdot c_3}{c_4 \cdot V}\right)^5 \frac{3}{16\sqrt{(x+y)^3}}$$
$$+\phi\left(\frac{\theta_1 - \theta_2 + u\cdot c_3 \cdot \sqrt{x+y}}{c_4 \cdot V}\right)\left(\frac{u\cdot c_3}{c_4 \cdot V}\right)^3 \frac{9}{16\sqrt{(x+y)^5}} + \phi\left(\frac{\theta_1 - \theta_2 + u\cdot c_3 \cdot \sqrt{x+y}}{c_4 \cdot V}\right)\left(\frac{u\cdot c_3}{c_4 \cdot V}\right)\frac{15}{16\sqrt{(x+y)^7}}$$

$$\frac{\partial g_2(x,y)}{\partial y} = \frac{(\nu_2/2-1)y^{\nu_2/2-2}x^{\nu_1/2-1}}{2^{(\nu_1+\nu_2)/2}\Gamma(\nu_1/2)\Gamma(\nu_2/2)\cdot V^{\nu_1+\nu_2}}e^{-\frac{x+y}{2V^2}} - \frac{y^{\nu_2/2-1}x^{\nu_1/2-1}}{2^{(\nu_1+\nu_2+2)/2}\Gamma(\nu_1/2)\Gamma(\nu_2/2)\cdot V^{\nu_1+\nu_2+2}}e^{-\frac{x+y}{2V^2}}.$$

$$\frac{\partial^2 g_2(x,y)}{\partial y^2} = \frac{(\nu_2/2-1)(\nu_2/2-2)y^{\nu_2/2-3}x^{\nu_1/2-1}}{2^{(\nu_1+\nu_2)/2}\Gamma(\nu_1/2)\Gamma(\nu_2/2)\cdot V^{\nu_1+\nu_2}}e^{-\frac{x+y}{2V^2}} - \frac{(\nu_2/2-1)y^{\nu_2/2-2}x^{\nu_1/2-1}}{2^{(\nu_1+\nu_2)/2}\Gamma(\nu_1/2)\Gamma(\nu_2/2)\cdot V^{\nu_1+\nu_2+2}}e^{-\frac{x+y}{2V^2}}$$
$$+\frac{y^{\nu_2/2-1}x^{\nu_1/2-1}}{2^{(\nu_1+\nu_2+4)/2}\Gamma(\nu_1/2)\Gamma(\nu_2/2)\cdot V^{\nu_1+\nu_2+4}}e^{-\frac{x+y}{2V^2}}.$$

$$\frac{\partial^3 g_2(x,y)}{\partial y^3}$$
$$= \frac{(\nu_2/2-1)(\nu_2/2-2)(\nu_2/2-3)y^{\nu_2/2-4}x^{\nu_1/2-1}}{2^{(\nu_1+\nu_2)/2}\Gamma(\nu_1/2)\Gamma(\nu_2/2)\cdot V^{\nu_1+\nu_2}}e^{-\frac{x+y}{2V^2}} - \frac{3(\nu_2/2-1)(\nu_2/2-2)y^{\nu_2/2-3}x^{\nu_1/2-1}}{2^{(\nu_1+\nu_2+2)/2}\Gamma(\nu_1/2)\Gamma(\nu_2/2)\cdot V^{\nu_1+\nu_2+2}}e^{-\frac{x+y}{2V^2}}.$$
$$+\frac{3(\nu_2/2-1)y^{\nu_2/2-2}x^{\nu_1/2-1}}{2^{(\nu_1+\nu_2+4)/2}\Gamma(\nu_1/2)\Gamma(\nu_2/2)\cdot V^{\nu_1+\nu_2+4}}e^{-\frac{x+y}{2V^2}} - \frac{y^{\nu_2/2-1}x^{\nu_1/2-1}}{2^{(\nu_1+\nu_2+6)/2}\Gamma(\nu_1/2)\Gamma(\nu_2/2)\cdot V^{\nu_1+\nu_2+6}}e^{-\frac{x+y}{2V^2}}$$

149

$$\frac{\partial^4 g_2(x,y)}{\partial y^4} = \frac{(\nu_2/2-1)(\nu_2/2-2)(\nu_2/2-3)(\nu_2/2-4)y^{\nu_2/2-5}x^{\nu_1/2-1}}{2^{(\nu_1+\nu_2)/2}\Gamma(\nu_1/2)\Gamma(\nu_2/2)\cdot V^{\nu_1+\nu_2}}e^{-\frac{x+y}{2V^2}}$$

$$-\frac{(\nu_2/2-1)(\nu_2/2-2)(\nu_2/2-3)y^{\nu_2/2-4}x^{\nu_1/2-1}}{2^{(\nu_1+\nu_2-2)/2}\Gamma(\nu_1/2)\Gamma(\nu_2/2)\cdot V^{\nu_1+\nu_2+2}}e^{-\frac{x+y}{2V^2}}+\frac{3(\nu_2/2-1)(\nu_2/2-2)y^{\nu_2/2-3}x^{\nu_1/2-1}}{2^{(\nu_1+\nu_2+2)/2}\Gamma(\nu_1/2)\Gamma(\nu_2/2)\cdot V^{\nu_1+\nu_2+4}}e^{-\frac{x+y}{2V^2}}\cdot$$

$$-\frac{(\nu_2/2-1)y^{\nu_2/2-2}x^{\nu_1/2-1}}{2^{(\nu_1+\nu_2+2)/2}\Gamma(\nu_1/2)\Gamma(\nu_2/2)\cdot V^{\nu_1+\nu_2+6}}e^{-\frac{x+y}{2V^2}}+\frac{y^{\nu_2/2-1}x^{\nu_1/2-1}}{2^{(\nu_1+\nu_2+8)/2}\Gamma(\nu_1/2)\Gamma(\nu_2/2)\cdot V^{\nu_1+\nu_2+8}}e^{-\frac{x+y}{2V^2}}$$

Since $\Phi\left(\dfrac{\theta_1-\theta_2+u\cdot c_3\cdot\sqrt{x+y}}{c_4\cdot V}\right)\geq 0$,

then $g_1(x,y)\leq\Phi\left(\dfrac{-u\cdot c_3\cdot\sqrt{x+y}}{c_4\cdot V}\right)\leq\Phi\left(\dfrac{-u\cdot c_3\cdot\sqrt{x}}{c_4\cdot V}\right)$, $\forall y\geq 0$.

$$\left|\frac{\partial g_1(x,y)}{\partial y}\right|=\phi\left(\frac{-u\cdot c_3\cdot\sqrt{x+y}}{c_4\cdot V}\right)\frac{u\cdot c_3}{2c_4\cdot V\sqrt{x+y}}+\phi\left(\frac{\theta_1-\theta_2+u\cdot c_3\cdot\sqrt{x+y}}{c_4\cdot V}\right)\frac{u\cdot c_3}{2c_4\cdot V\sqrt{x+y}},$$

$$\leq\phi\left(\frac{-u\cdot c_3\cdot\sqrt{x}}{c_4\cdot V}\right)\frac{u\cdot c_3}{c_4\cdot V\sqrt{x}}$$

$\forall y\geq 0$, due to the fact that $\phi\left(\dfrac{\theta_1-\theta_2+u\cdot c_3\cdot\sqrt{x+y}}{c_4\cdot V}\right)\leq\phi\left(\dfrac{-u\cdot c_3\cdot\sqrt{x+y}}{c_4\cdot V}\right)$ when

$$\frac{\theta_1-\theta_2+u\cdot c_3\cdot\sqrt{x+y}}{c_4\cdot V}\leq\frac{-u\cdot c_3\cdot\sqrt{x+y}}{c_4\cdot V}\leq 0.$$

Similarly,

$$\left|\frac{\partial^2 g_1(x,y)}{\partial y^2}\right|\leq\phi\left(\frac{-u\cdot c_3\cdot\sqrt{x}}{c_4\cdot V}\right)\left[\left(\frac{u\cdot c_3}{c_4\cdot V}\right)^3\frac{1}{2\sqrt{x}}+\frac{u\cdot c_3}{c_4\cdot V}\frac{1}{2\sqrt{x^3}}\right],$$

$$\left|\frac{\partial^3 g_1(x,y)}{\partial y^3}\right|\leq\phi\left(\frac{-u\cdot c_3\cdot\sqrt{x}}{c_4\cdot V}\right)\left[\left(\frac{u\cdot c_3}{c_4\cdot V}\right)^5\frac{1}{4\sqrt{x}}+\left(\frac{u\cdot c_3}{c_4\cdot V}\right)^3\frac{1}{2\sqrt{x^3}}+\frac{3u\cdot c_3}{c_4\cdot V}\frac{1}{4\sqrt{x^5}}\right],\text{ and}$$

$$\left|\frac{\partial^4 g_1(x,y)}{\partial y^4}\right| \le \phi\left(\frac{-u\cdot c_3\cdot\sqrt{x}}{c_4\cdot V}\right)\left[\left(\frac{u\cdot c_3}{c_4\cdot V}\right)^7\frac{1}{8\sqrt{x}}+\left(\frac{u\cdot c_3}{c_4\cdot V}\right)^5\frac{3}{8\sqrt{x^3}}+\left(\frac{u\cdot c_3}{c_4\cdot V}\right)^3\frac{9}{8\sqrt{x^5}}+\left(\frac{u\cdot c_3}{c_4\cdot V}\right)\frac{15}{8\sqrt{x^7}}\right]$$

.

Since $g_2(x,y)=\dfrac{y^{\nu_2/2-1}x^{\nu_1/2-1}}{2^{(\nu_1+\nu_2)/2}\Gamma(\nu_1/2)\Gamma(\nu_2/2)\cdot V^{\nu_1+\nu_2}}e^{-\frac{x+y}{2V^2}}$ achieves the maximum value at

$y=2V^2(\nu_2/2-1)$ for an given x, then $g_2(x,y)\le\left|\dfrac{(\nu_2/2-1)^{\nu_2/2-1}x^{\nu_1/2-1}e^{-\left(\frac{x}{2V^2}+\nu_2/2-1\right)}}{2^{(\nu_1+2)/2}\Gamma(\nu_1/2)\Gamma(\nu_2/2)\cdot V^{\nu_1+2}}\right|$.

Similarly,

$$\left|\frac{\partial g_2(x,y)}{\partial y}\right|\le\frac{x^{\nu_1/2-1}e^{-\frac{x}{2V^2}}\left[\left|(\nu_2/2-1)(\nu_2/2-2)^{\nu_2/2-2}e^{-(\nu_2/2-2)}\right|+\left|(\nu_2/2-1)^{\nu_2/2-1}e^{-(\nu_2/2-1)}\right|\right]}{2^{(\nu_1+4)/2}\Gamma(\nu_1/2)\Gamma(\nu_2/2)\cdot V^{\nu_1+4}},$$

$$\left|\frac{\partial^2 g_2(x,y)}{\partial y^2}\right|\le\frac{x^{\nu_1/2-1}e^{-\frac{x}{2V^2}}}{2^{(\nu_1+6)/2}\Gamma(\nu_1/2)\Gamma(\nu_2/2)\cdot V^{\nu_1+6}}$$
$$\left[\left|(\nu_2/2-1)(\nu_2/2-2)(\nu_2/2-3)^{\nu_2/2-3}e^{-(\nu_2/2-3)}\right|+\left|2(\nu_2/2-1)(\nu_2/2-2)^{\nu_2/2-2}e^{-(\nu_2/2-2)}\right|+\left|(\nu_2/2-1)^{\nu_2/2-1}e^{-(\nu_2/2-1)}\right|\right]$$

,

$$\left|\frac{\partial^3 g_2(x,y)}{\partial y^3}\right|$$
$$\le\left|\frac{(\nu_2/2-1)(\nu_2/2-2)(\nu_2/2-3)(\nu_2/2-4)^{\nu_2/2-4}x^{\nu_1/2-1}}{2^{(\nu_1+8)/2}\Gamma(\nu_1/2)\Gamma(\nu_2/2)\cdot V^{\nu_1+8}}e^{-\left(\frac{x}{2}+\frac{\nu_2}{2}-4\right)}\right|$$
$$+\left|\frac{3(\nu_2/2-1)(\nu_2/2-2)(\nu_2/2-3)^{\nu_2/2-3}x^{\nu_1/2-1}}{2^{(\nu_1+8)/2}\Gamma(\nu_1/2)\Gamma(\nu_2/2)\cdot V^{\nu_1+8}}e^{-\left(\frac{x}{2}+\frac{\nu_2}{2}-3\right)}\right|$$
$$+\left|\frac{3(\nu_2/2-1)(\nu_2/2-2)^{\nu_2/2-2}x^{\nu_1/2-1}}{2^{(\nu_1+8)/2}\Gamma(\nu_1/2)\Gamma(\nu_2/2)\cdot V^{\nu_1+8}}e^{-\left(\frac{x+\nu_2}{2}-2\right)}\right|+\left|\frac{(\nu_2/2-1)^{\nu_2/2-1}x^{\nu_1/2-1}}{2^{(\nu_1+8)/2}\Gamma(\nu_1/2)\Gamma(\nu_2/2)\cdot V^{\nu_1+8}}e^{-\left(\frac{x+\nu_2}{2}-1\right)}\right|$$

,

and

$$\left|\frac{\partial^4 g_2(x,y)}{\partial y^4}\right|$$

$$\leq \left|\frac{(\nu_2/2-1)(\nu_2/2-2)(\nu_2/2-3)(\nu_2/2-4)(\nu_2/2-5)^{\nu_2/2-5}x^{\nu_1/2-1}}{2^{(\nu_1+10)/2}\Gamma(\nu_1/2)\Gamma(\nu_2/2)\cdot V^{\nu_1+10}}e^{-\left(\frac{x}{2}+\nu_2/2-5\right)}\right|$$

$$+\left|\frac{4(\nu_2/2-1)(\nu_2/2-2)(\nu_2/2-3)(\nu_2/2-4)^{\nu_2/2-4}x^{\nu_1/2-1}}{2^{(\nu_1+10)/2}\Gamma(\nu_1/2)\Gamma(\nu_2/2)\cdot V^{\nu_1+10}}e^{-\left(\frac{x}{2}+\nu_2/2-4\right)}\right|$$

$$+\left|\frac{6(\nu_2/2-1)(\nu_2/2-2)(\nu_2/2-3)^{\nu_2/2-3}x^{\nu_1/2-1}}{2^{(\nu_1+10)/2}\Gamma(\nu_1/2)\Gamma(\nu_2/2)\cdot V^{\nu_1+10}}e^{-\left(\frac{x+\nu_2}{2}-3\right)}\right|$$

$$+\left|\frac{4(\nu_2/2-1)(\nu_2/2-2)^{\nu_2/2-2}x^{\nu_1/2-1}}{2^{(\nu_1+10)/2}\Gamma(\nu_1/2)\Gamma(\nu_2/2)\cdot V^{\nu_1+10}}e^{-\left(\frac{x+\nu_2}{2}-2\right)}\right|$$

$$+\left|\frac{(\nu_2/2-1)^{\nu_2/2-1}(\nu_1/2-1)^{\nu_1/2-1}}{2^{(\nu_1+10)/2}\Gamma(\nu_1/2)\Gamma(\nu_2/2)\cdot V^{\nu_1+10}}e^{-\left(\frac{x+\nu_2}{2}-1\right)}\right|$$

Hence, we have derived the upper bound for $\left|\frac{\partial^4 f(x,y)}{\partial y^4}\right|$ as:

$$\left|\frac{\partial^4 f(x,y)}{\partial y^4}\right|\leq\left|\frac{\partial^4 g_1(x,y)}{\partial y^4}g_2(x,y)\right|+4\left|\frac{\partial^3 g_1(x,y)}{\partial y^3}\frac{\partial g_2(x,y)}{\partial y}\right|+6\left|\frac{\partial^2 g_1(x,y)}{\partial y^2}\frac{\partial^2 g_2(x,y)}{\partial y^2}\right|$$

$$+4\left|\frac{\partial^3 g_2(x,y)}{\partial y^3}\frac{\partial g_1(x,y)}{\partial y}\right|+\left|\frac{\partial^4 g_2(x,y)}{\partial y^4}g_1(x,y)\right|$$

$\forall y\in[0,a]$.

152

Chapter 6   Sensitivity analyses

6.1 Introduction

In the evaluation of a bioequivalence study, it is a common practice to assume that both log(AUC) and log($Cmax$) are normal variables [1]. However the normality assumption of the log($Cmax$) in a bioequivalence study cannot always be assumed due to the skewness based on the limited simulation results in Chapter 2. Our limited simulation scenarios in Chapter 2 show that the distribution of the log(AUC) can be heavy tailed.

In this chapter, we compare the type I error rate, sample size, and power of two one-sided tests (TOST) we describe in Chapters 4 and 5 if the log($Cmax$) (or log(AUC)) follows one of three non-normal distributions: heavy tailed t distribution, skew-normal distribution, or a mixture of two normal distributions, respectively, with the type I error rate, sample size, and power if the log($Cmax$) (or log(AUC)) follows a normal distribution.

In Section 6.2, we will briefly discuss three non-normal distributions: the skew-normal distribution, t-distribution, and a mixture of two normal distributions. These three distributions closely mimic the empirical distributions from simulations based on the pharmacokinetic compartment models in Chapter 2. However, Chapter 2 only looked at a single time period and failed to account for the crossover design used in the bioequivalence setting. In Section 6.3, we simulate the response variable from the linear mixed effect model in Section 3.2. From the simulations, we look into the impact of non-normality of the response variable on the type I error rate, sample size,

and power of TOST for the one stage study discussed in Chapter 4. In Section 6.4, we simulate the response variable from the linear mixed effect model in Section 5.2. Based on the simulations, we also look into the impact of non-normality of the response variable on the type I error rate and sample size for a two-stage study discussed in Chapter 5.

6.2 Non-normal distributions

In this section, we focus on three non-normal distributions: the skew-normal distribution, *t*-distribution, and a mixture of two normal distributions. The motivation for these three non-normal models is discussed in Chapter 2. Recall that Y_{ijk} denotes the response (log(AUC) or log($Cmax$)) of the kth subject in the jth period of the ith sequence in the 2-period 2-treatment crossover study $\forall$ i=1, 2, j=1, 2 and k=1,.., n_i. Y_{ijk} is modeled as $Y_{ijk} = \gamma + S_{ik} + P_j + F_{ij} + \varepsilon_{ijk}$.

From Chapter 3 to Chapter 5, we assume that ε_{ijk} denotes a random variable following a normal distribution with $E(\varepsilon_{ijk}) = 0$, and $V(\varepsilon_{ijk}) = \sigma_T^2$ if $i \neq j, \forall i, j = 1,2$ for the test formulation and $V(\varepsilon_{ijk}) = \sigma_R^2$ if $i = j, \forall i, j = 1,2$ for the reference formulation. We assume $\sigma_T = \sigma_R = \sigma$ in two following sections. Let $N(0,1)$ stand for the standard normal variable, then $\varepsilon_{ijk} = \sigma \cdot N(0,1)$.

Now let us discuss three non-normal distributions. The first non-normal distribution is the *t*-distribution with ν degrees of freedom denoted by $t(\nu)$. The second non-normal distribution is the skew-normal distribution with the location parameter being 0, the scale parameter being 1, and the shape parameter being ς,

which is denoted by $SN(0,1,\varsigma)$. The expectation of $SN(0,1,\varsigma)$ is $\sqrt{2/\pi}\varsigma/\sqrt{1+\varsigma^2}$ and

the variance of $SN(0,1,\varsigma)$ is $1-2\varsigma^2/\left[\left(1+\varsigma^2\right)\pi\right]$. We use the R package [62] for

simulating skew-normal errors. In the following subsections, we respectively

compare the type I error rate, power, and sample size if ε_{ijk} follows $\sigma\cdot t(v)$

or $\sigma\cdot SN(0,1,\varsigma)$ with the type I error rate, power, and sample size if ε_{ijk} follows

$\sigma\cdot N(0,1)$.

The third distribution in our simulation study is a mixture of two normal

distributions. Assume that U is uniformly distributed in [0, 1], we write the mixture of

2 normal variables as $I(U \geq \delta_0)\cdot N\left(\mu_1,\sigma_1^2\right)+I(U \leq \delta_0)\, N\left(\mu_2,\sigma_2^2\right)$. Here

$I(\cdot)$ is the indicator function, δ_0 is a constant in [0, 1], and $N(\mu_i,\sigma_i^2)$ denotes the

normal variable with mean μ_i and variance σ_i^2.

We use this mixture model to investigate the impact of bimodality of the response

distribution, in which a small fraction of the population (a subgroup) responds to a

test formulation with a mean response μ_1 unequal to the mean response μ_2 of the

majority of the population, on the type I error rate, power, and sample size.

The subgroup's variability σ_1 is also different from the variability of the majority

of population, σ_2.

6.3 Impact of non-normality of the response variable on the type I error rate and power for the single stage study

In this section, we simulate the within-subject difference in the response variable between the test and the reference from the linear mixed effect model in Section 3.2 with non-normal distributed errors. Then we compare the type I error rate and power for the one stage study under non-normality than those under normality.

6.3.1 Simulation scheme

We use the following simulation scheme for evaluating the power under θ^* with the assumption of a true constant period effect which was assigned the fixed value of 0.05 in order to conduct the simulations. In the context of a balanced cross over design, the value is arbitrary and will not impact the results of interest. To avoid specifying another parameter for random subject effect, we simulate the difference in the response variable for each subject in each sequence. Let d_{ik} denote the difference in the response variable for Subject k between the period 2 and period 1 in Sequence i.

So $d_{ik} = (-1)^{i+1} \theta^* + P_2 - P_1 + \varepsilon_{i2k} - \varepsilon_{i1k}$, $k=1,2\ldots,n_i$.

1) Simulate ε_{ijk} from the non-normal distribution, e.g., $\sigma \cdot t(v)$, $\forall\, i=1, 2, j=1, 2$ and $k=1,..,n_i$.

2) Obtain the data d_{11}, $d_{12},\ldots,d_{1n_1}$ and d_{21}, $d_{22},\ldots,d_{2n_2}$

3) Compute $\hat{D} = \dfrac{\sum\limits_{i=1}^{2}\left(\sum\limits_{k=1}^{n_i} d_{ik}\Big/n_i\right)}{2}$, $S^2 = \dfrac{1}{n_1+n_2-2}\left[\sum\limits_{i=1}^{2}\sum\limits_{k=1}^{n_i}(d_{ik}-\sum\limits_{k=1}^{n_i} d_{ik}\Big/n_i)^2\right]$,

$$T_1 = \dfrac{\hat{D}-\theta_1}{\dfrac{S}{2}\bullet\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}\text{ , and } T_2 = \dfrac{\hat{D}-\theta_2}{\dfrac{S}{2}\bullet\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}\text{ .}$$

4) If $T_1 \geq t_\alpha(n_1+n_2-2)$ and $T_2 \leq t_\alpha(n_1+n_2-2)$ then pass=1. Otherwise pass=0.

5) Repeat Steps 1 to 5 for 100,000 times.

6) Calculate the percentage of rejecting H_0 under H_0 in (3.1).

6.3.2 Type I error rate and power in a one stage study

In this subsection, we compare the simulated type I error rate and power for non-normal distributions with those for normal distributions for the one stage study.

The simulated type I error rate under $\theta^* = \log(1.25)$ and the simulated power under $\theta^* = 0.05$ in Tables 6.1 and 6.2, respectively if $\varepsilon_{ijk} \sim \sigma \cdot N(0,1)$, $\sigma \cdot t(v)$, and $\sigma \cdot SN(0,1,\varsigma)$. The simulation standard errors are less than 0.0005 when 10^6 Monte Carlo replications are used in Tables 6.1 to 6.4. Table 6.1 shows that the type I error rates under any of above three distributions for various sample size (n) are comparable to those under normality. From Table 6.2, we can see that the power values under $\sigma \cdot t(5)$ or $\sigma \cdot t(10)$ for various sample size are smaller than those under normality. But there is not much difference in power between the skew-normal and normal distributions.

Table 6.1 Comparison of the simulated type I error rates (%) for various n values under $\theta^* = \log(1.25)$, $P_2 - P_1 = 0.05$, and $\sigma = 0.2$ among $\varepsilon_{ijk} \sim \sigma N(0,1)$, $\sigma t(v)$, $\sigma\sqrt{(v-2)/v}\,t(v)$, and $\sigma SN(0,1,\varsigma)$.

ε_{ijk} n	$\sigma \cdot N(0,1)$	$\sigma \cdot t(v)$		$\sigma\sqrt{(v-2)/v} \cdot t(v)$	$\sigma \cdot SN(0,1,\varsigma)$		
		$v=5$	$v=10$	$v=5$	$\varsigma=-1$	$\varsigma=-0.5$	$\varsigma=0.5$
20	4.99	4.87	4.96	4.94	4.97	5.01	5.01
24	5.01	4.98	5.01	4.98	5.01	4.96	4.98
28	5.00	4.99	5.00	4.99	4.98	5.02	4.97
32	5.00	5.00	5.02	4.98	4.97	5.01	4.98
36	4.93	4.98	4.98	4.99	5.02	5.01	5.01
40	5.03	4.96	4.99	5.00	4.98	5.00	5.00

Table 6.2 Comparison of the estimated power (%) from simulations for various n values under $\theta^* = \log(1.25)$, $P_2 - P_1 = 0.05$, and $\sigma = 0.2$ among $\varepsilon_{ijk} \sim \sigma N(0,1)$, $\sigma t(v)$, $\sigma\sqrt{(v-2)/v}\,t(v)$, and $\sigma SN(0,1,\varsigma)$.

ε_{ijk} n	$\sigma \cdot N(0,1)$	$\sigma \cdot t(v)$		$\sigma\sqrt{\frac{v-2}{v}} \cdot t(v)$	$\sigma \cdot SN(0,1,\varsigma)$		
		$v=5$	$v=10$	$v=5$	$\varsigma=-1$	$\varsigma=-0.5$	$\varsigma=0.5$
20	83.23	60.70	74.20	83.09	93.80	87.63	87.67
24	89.45	70.40	82.12	89.00	96.94	92.80	92.76
28	93.34	77.51	87.61	92.84	98.50	95.84	95.83
32	95.86	82.73	91.41	95.32	99.30	97.59	97.60
36	97.46	86.78	94.05	96.99	99.67	98.63	98.63
40	98.46	89.87	95.90	98.03	99.85	99.22	99.23

The simulated type I error rate under $\theta^* = \log(1.25)$ and the simulated power

under $\theta^* = 0.05$ are listed in Tables 6.3 and 6.4, respectively if

$$\varepsilon_{ijk} \sim I(U \geq 0.7) \cdot N(0.07, \sigma_1) + I(U \leq 0.7) \, N(-0.03, \sigma_2).$$

In Table 6.3, the subpopulation has a mean shift from the majority of the

population, but the overall mean is $\log(1.25)$. Table 6.3 shows that the type I error

rate is about 5.5% when the within-subject difference in the response variable for the

subpopulation is more variable than that for the majority of the population; the type I

error rate is below 5% when the within-subject difference in the response variable for

the subpopulation is less variable than for the majority of the population; the type I

error rate jumps around 5% when the within-subject difference in the response

variable for the subpopulation has the same variance as the majority of the

population. Table 6.4 shows that the power is much smaller when one of two

subpopulations has a larger variance than that under normality. This is because the

overall variance for the mixture distribution with a more variable major population is

much larger than the variance of the normal distribution. The power when two

subpopulations have equal variance is similar to that under normality.

Table 6.3 Comparison of the simulated type I error rates (%) for various n values under $\theta^* = \log(1.25)$, $P_2 - P_1 = 0.05$, and $\sigma = 0.2$ when $\varepsilon_{ijk} \sim$ $I(U \geq 0.7) \cdot N(0.07, \sigma_1) + I(U \leq 0.7)\, N(-0.03, \sigma_2)$.

n	$\sigma_1 = 2\sigma, \sigma_2 = \sigma$	$\sigma_1 = \sigma_2 = \sigma$	$\sigma_1 = \sigma, \sigma_2 = 2\sigma$
20	5.44	4.98	3.42
24	5.51	5.04	4.09
28	5.50	5.06	4.49
32	5.47	5.02	4.66
36	5.47	5.01	4.76
40	5.45	4.99	4.77

Table 6.4 Comparison of the simulated power (%) for various n values under $\theta^* = \log(1.25)$, $P_2 - P_1 = 0.05$, and $\sigma = 0.2$ when $\varepsilon_{ijk} \sim I(U \geq 0.7) \cdot N(0.07, \sigma_1) + I(U \leq 0.7)\, N(-0.03, \sigma_2)$

n	$\sigma_1 = 2\sigma, \sigma_2 = \sigma$	$\sigma_1 = \sigma_2 = \sigma$	$\sigma_1 = \sigma, \sigma_2 = 2\sigma$
20	53.10	82.30	22.30
24	63.15	88.71	33.17
28	70.81	92.82	43.37
32	76.81	95.47	51.91
36	81.34	97.15	59.16
40	85.02	98.24	65.37

6.4 Impact of non-normality of the response variable on the type I error rate and expected sample size for the two-stage study described in Chapter 5.

In this section, we simulate the within-subject difference in the response variable between the test and the reference from the linear mixed effect model in Section 5.2 with non-normal distributed errors for a sample size of $n_{1\bullet}$. From this, we obtain the sample variance S_1^2. Using S_1^2 as the variance V, we estimate the sample size ($n_{2\bullet}$)

for the second stage. Then we simulate within-subject differences in the response variable for $n_{2\bullet}$ subjects from the linear mixed effect model in Section 5.2 with non-normal distributed errors. The critical value ($u_{0.05}$) is 1.715 when $n_{1\bullet}$=20. Last we compare the type I error rate and expected sample size for the two-stage stage study under non-normality to the corresponding type I error rate and expected sample size under normality.

6.4.1 Simulation scheme

Again we simulate the within-subject difference in the response variable between Period 2 and Period 1 for each subject in each sequence. Let d_{lik} denote the within-subject difference in the response variable for Subject k between the period 2 and period 1 in Sequence i at Stage l. So $d_{1ik} = (-1)^{i+1}\theta^* + P_2 - P_1 + \varepsilon_{i2k} - \varepsilon_{i1k}$, k=1,2...,n_{1i}.

$d_{2ik} = (-1)^{i+1}\theta^* + P_2 - P_1 + \varepsilon_{i2k} - \varepsilon_{i1k}$, k=n_{1i}+1,...,n_{1i}+n_{2i}.

The simulation scheme is described as follows.

1) Simulate ε_{ijk} from the non-normal distribution, e.g., $t(5)$, $\forall$ i=1, 2, j=1, 2 and k=1,.., n_{1i}.

2) Obtain the data d_{111}, d_{112},...,d_{11n_1} and d_{121}, d_{122},...,d_{12n_2}

3) Compute $S_1^2 = \dfrac{1}{n_{11}+n_{12}-2}\left[\displaystyle\sum_{i=1}^{2}\sum_{k=1}^{n_{1i}}(d_{1ik} - \sum_{k=1}^{n_{1i}}d_{1ik}\Big/n_{1i})^2\right]$.

4) Calculate the new total sample size ($n_1 + n_2$) from the power function (5.2) assuming that $p_0 = 0.9$, $\theta^* = 0.05$ and $V^2 = S_1^2$ if $S_1^2 > S_0^2$. If $n_2 = 0$, then stop here. Otherwise continue to next step.

5) Simulate ε_{ijk} from the non-normal distribution, e.g., $t(5)$, $\forall i=1, 2, j=1, 2$ and $k=n_{1i}+1,\ldots,n_{1i}+n_{2i}$. Note that $n_{21} = (n_1 + n_2)/2 - n_{11}$ and $n_{22} = (n_1 + n_2)/2 - n_{12}$.

6) Obtain the data $d_{21,n_{11}+1}$, $d_{21,n_{11}+2}, \ldots, d_{21,n_{11}+n_{21}}$ and $d_{22,n_{12}+1}$, $d_{22n_{12}+2}, \ldots, d_{22n_{12}+n_{22}}$

7) Compute $S_2^2 = \dfrac{1}{n_{21} + n_{22} - 2}\left[\displaystyle\sum_{i=1}^{2}\sum_{k=n_{1i}+1}^{n_{1i}+n_{2i}}\left(d_{2ik} - \sum_{k=n_{1i}+1}^{n_{1i}+n_{2i}}d_{2ik}\Big/n_{1i}\right)^2\right]$

8) Compute $\left(S^*\right)^2 = \dfrac{(n_1 - 2)S_1^2 + (n_2 - 2)S_2^2}{n_1 + n_2 - 4}$

9) Compute $\hat{D}_1 = \dfrac{\displaystyle\sum_{i=1}^{2}\left(\sum_{k=1}^{n_{1i}}d_{1ik}\Big/n_{1i}\right)}{2}$ and $\hat{D}_2 = \dfrac{\displaystyle\sum_{i=1}^{2}\left(\sum_{k=n_{1i}+1}^{n_{1i}+n_{2i}}d_{2ik}\Big/n_{2i}\right)}{2}$, and

$\hat{D} = \left(n_1\hat{D}_1 + n_2\hat{D}_2\right)/(n_1 + n_2)$.

10) Calculate $T_{21}^* = \left(\hat{D} - \theta_1\right)\Big/\left(\dfrac{S^*}{2}\cdot\sqrt{\dfrac{1}{n_{11} + n_{21}} + \dfrac{1}{n_{12} + n_{22}}}\right)$ and

$T_{22}^* = \left(\hat{D} - \theta_2\right)\Big/\left(\dfrac{S^*}{2}\cdot\sqrt{\dfrac{1}{n_{11} + n_{21}} + \dfrac{1}{n_{12} + n_{22}}}\right)$

11) If $T_{21}^* \geq u_\alpha$ and $T_{22}^* \geq -u_\alpha$ then pass=1. Otherwise pass=0.

12) Repeat Steps 1 to 11 many times, for example 50,000 times, to calculate the percentage of rejecting the H_0 under the H_0 in (5.1) if $\theta^* = \log(1.25)$ or $\theta^* = 0.05$.

Below we investigate the impact of non-normality of the response variable on the type I error rate and expected sample size.

6.4.2 Impact of non-normality of the response variable on the type I error rate and sample size for the two-stage study in Chapter 5

1. Type I error rate

We used 10^5 simulation replicates in this section, so the simulation standard error is 0.0007. From Table 6.5, we can see that the type I error rates using our two stage approach when ε_{ijk} follows $\sigma \cdot t(5)$ or $\sqrt{0.6}\sigma t(5)$ are similar to those under normality. It also shows that the type I error rates when ε_{ijk} follows $\sigma \cdot SN(0,1,\varsigma)$, and

$I(U \geq 0.7) \cdot N(0.07,\sigma_1) + I(U \leq 0.7)\, N(-0.03,\sigma_2)$ are smaller compared to those under normality. The type I error rates for statistical analyses of the two-stage data when ε_{ijk} follows the mixture of 2 normal variables are above 0.05 and larger than those under normality for $\sigma = 0.25$. Clearly it is clear that our proposed critical values for statistical analyses of the two-stage data are robust to non-normal distribution for most scenarios if the overall mean is not mis-specified.

Table 6.5 Comparison of the estimated type I error rates (%) from simulations for various σ values with $u_{0.05} = 1.715$ for $n_{11} = n_{12} = 20$ if ε_{ijk} follows $\sigma \cdot t(5)$, $\sqrt{0.6}\sigma \cdot t(5)$, $\sigma \cdot SN(0,1,\varsigma)$, and $I(U \ge 0.7) \cdot N(0.07, \sigma_1) + I(U \le 0.7) \, N(-0.03, \sigma_2)$

σ	ε_{ijk}						
	$N(0,\sigma)$	$\sigma \cdot t(5)$	$\sqrt{0.6}\sigma t(5)$	$\sigma \cdot SN(0,1,\varsigma)$		$I(U \ge 0.7) \cdot N(0.07,\sigma_1)$ $+ I(U \le 0.7)\, N(-0.03,\sigma_2)$	
				$\varsigma = -0.5$	$\varsigma = 0.5$	$\sigma_1 = 2\sigma_2 = 2\sigma$	$\sigma_1 = \sigma_2 = \sigma$
0.25	4.99	4.77	4.79	5.03	4.97	5.34	5.09
0.3	4.92	4.86	4.64	4.49	4.61	5.05	4.99
0.35	4.70	4.64	4.68	4.66	4.74	4.93	4.74
0.4	4.54	4.56	4.58	4.71	4.88	4.75	4.44
0.45	4.65	4.44	4.41	4.78	4.67	4.97	4.56

2. Sample size

Table 6.6 shows that the expected sample sizes when ε_{ijk} follows $\sigma \cdot t(5)$,

$\sqrt{0.6}\sigma t(5)$, or $\sigma \cdot SN(0,1,\varsigma)$ are much larger than those under normality assumption.

The expected sample sizes when ε_{ijk} follows $\sqrt{0.6}\sigma t(5)$ are slightly larger than those

under normality assumption. It also shows that the expected sample sizes when ε_{ijk}

follows a mixture of 2 normal variables, in which the variability of a subpopulation is

twice as large as that of the majority of the population, are larger than those under

normality. The expected sample sizes are similar when ε_{ijk} follows a mixture of 2

normal variables with equal variance to those under normality.

Table 6.6 Comparison of the expected sample sizes from simulations for various σ values with with $u_{0.05} = 1.715$ for $n_{11} = n_{12} = 20$ if ε_{ijk} follows $\sigma \cdot t(5)$, $\sqrt{0.6}\sigma \cdot t(5)$, $\sigma \cdot SN(0,1,\varsigma)$, and $I(U \geq 0.7) \cdot N(0.07, \sigma_1) + I(U \leq 0.7) \, N(-0.03, \sigma_2)$

σ	ε_{ijk}						
	$N(0,\sigma)$	$\sigma \cdot t(5)$	$\sqrt{0.6}\sigma t(5)$	$\sigma \cdot SN(0,1,\varsigma)$		$I(U \geq 0.7) \cdot N(0.07, \sigma_1)$ $+ I(U \leq 0.7) \, N(-0.03, \sigma_2)$	
				$\varsigma = -0.5$	$\varsigma = 0.5$	$\sigma_1 = 2\sigma_2 = 2\sigma$	$\sigma_1 = \sigma_2 = \sigma$
0.25	40	62.2	43.6	40.8	40.8	40.0	40.0
0.3	40.1	88.4	54.7	48.3	48.3	41.9	40.1
0.35	42.0	119.1	43.2	63.4	63.3	49.6	42.3
0.4	49.0	154.8	49.9	82.1	82.1	62.5	49.6
0.45	60.3	195.8	60.7	103.4	103.6	78.1	60.8

6.5 Discussion and conclusions

For TOST in the one stage design, our sensitivity analyses show that the type I error rates are not inflated when errors are distributed as the $\sigma \cdot t(5)$ or $\sigma \cdot t(10)$, or $\sigma \cdot SN(0,1,\varsigma)$ if the true mean difference is log(1.25) but the powers under the $\sigma \cdot t(5)$ or $\sigma \cdot t(10)$ for $\sigma = 0.25$ are smaller compared to those under normality. However the type I error rates under the mixture of two normal variables can be above 5% or below 5% or equal to 5% depending on the magnitude of the variance of subpopulation relative to that of the majority of the population. The resulting power under the mixture of two normal variables is much smaller when the variance of subpopulation is not equal to the variance of the majority of the population than those when the variance of subpopulation is equal to the variance of the majority of the population.

For the two-stage study, the type I error rates are robust if when errors are random variables distributed as $\sigma \cdot t(5)$, $\sqrt{0.6}\sigma t(5)$, and $\sigma \cdot SN(0,1,\varsigma)$ under various σ values shown in Tables 6.5 and 6.6 if the true mean difference is log(1.25). The type I error rates for two-stage study are 5.34% for a mixture of two normal variables such as $I(U \geq 0.7) \cdot N(0.07,\sigma_1) + I(U \leq 0.7) N(-0.03,\sigma_2)$, $\sigma_1 = 2\sigma_2 = 2\sigma$, if the overall mean is not mis-specified and $\sigma = 0.25$ The expected sample size for all non-normal distributions investigated are larger than those under normality.

Chapter 7 Final conclusions and recommendation

This dissertation attacked three problems related to the statistical hypothesis testing in bioequivalence studies. First we investigated normality assumption for the response variable ($\log(AUC)$ or $\log(Cmax)$) from the simulated concentration-time profiles for a large number of subjects. Second, we presented the exact power formula and the sample size calculation for planning a fixed sample size bioequivalence study. Third, we developed simpler test statistics and obtained the exact critical values, which these simpler test statistics are compared to, for the two-stage study in the new unblinded sample size re-estimation procedure.

7.1 Summary and conclusions

In Chapter 2, we simulated the concentration-time profiles from the two-stage one (or two) compartment models and multiplicative measurement errors. Comparing the histogram of the standardized response variable ($\log(AUC)$ or $\log(Cmax)$) for a large sample with the standard normal density curve under many different parameter combinations revealed that the sampling distribution of the standardized $\log(AUC)$ often had heavy tails. The sampling distribution of $\log(Cmax)$ was skewed either to the left or to the right and was not robust to many perturbations studied in Chapter 2. For TOST in the one stage design, our sensitivity analyses in Chapter 6 showed that the type I error rates were robust when errors were distributed as $\sigma \cdot t(5)$, $\sigma\sqrt{0.6} \cdot t(5)$, $\sigma \cdot t(10)$, or $\sigma \cdot SN(0,1,0.5)$, which closely mimicked the empirical distributions from simulations in Chapter 2, if the true mean difference is $\log(1.25)$. But the powers under $\sigma \cdot t(5)$ or $\sigma \cdot t(10)$ for $\sigma = 0.2$ were smaller compared to those

under normality. Type I error rates were inflated if there was a subgroup responding differently to the test product from the majority of the population.

In Chapter 3, the two explicit power functions for the power approach and two one-sided tests procedure provided the fundamental understanding of the difference between the two one-sided tests procedure and the power approach. We concluded that the power approach in practice usually consisted of testing the hypothesis of no difference at level 0.05 and a lack of significance was often used to incorrectly infer equivalence. In Chapter 4, we derived the exact power formula for two one-sided tests. From the exact power function, we calculated the exact powers and sample sizes in all bioequivalence study settings, including those for unequal variance of the test and reference products. The comparison between the exact power and a previous approximate power approach showed this approximate method could significant underestimate the power for many practical settings. Accurate numerical calculation of exact power could be readily obtained with our R code for the widely used free R software package.

In Chapter 5, our simulation showed that not much information was lost if the simpler test statistics (T_{21}^*, T_{22}^*) was used instead of (T_{21}, T_{22}). In our unblinded sample size re-estimation procedure, we calculated the new sample size using the exact power function derived in Chapter 4 in which the sample variance from Stage 1 data replaced the true variance. To assure the experimentwise type I error at the nominal level α, we presented the exact power function for one stage study in which we compared (T_{11}, T_{12}) with the $(t_\alpha(n_{1\bullet} - 2), -t_\alpha(n_{1\bullet} - 2))$ and the exact power function for

two-stage study in which we compared simpler test statistics $\left(T_{21}^*, T_{22}^*\right)$ with the exact critical values $\left(u_\alpha, -u_\alpha\right)$. The positive exact critical value, u_α, was derived as the largest value of u for which the following condition holds: that supremum of the probability of rejecting the null under the null in the composite hypothesis testing (5.1) is exactly α. We proved that the maximum of this probability occurs within the carefully chosen range for V and u. We developed a numerical analysis for calculating the exact type I error rates and bounded the numerical and computation errors for approximating the exact type I error rates at grid points and quantified the maximum difference for any subinterval between two adjacent grid points in calculating exact type I error rate. To make sure the size of the sample size re-estimation procedure at the nominal level α-10^{-4}, we chose the sufficient number of grid points in the carefully-chosen interval so that the sum of all levels of errors is about 10^{-4}.

For the sample size re-estimation study design, the type I error rates and expected sample sizes are robust when errors are random variables distributed as $\sigma \cdot t(5)$, $\sqrt{0.6}\sigma t(5)$, and $\sigma \cdot SN(0,1,0.5)$ under various σ values. Type I error rates are inflated if errors are distributed as the mixture of 2 normal distributions.

7.2 Recommendation for future work

This work mainly focused on statistical methods in bioequivalence studies under normality. Our conclusions about the distributional assumption for log(AUC) or log($Cmax$) should be scrutinized further with many real cases. If normality of log(AUC) or log($Cmax$) cannot be assumed, statistical methods using robust statistics should be developed. Since we did not analytically prove that the numerical

169

maximum was the global one, then there could be more work for theoretical

evaluation of the sample size re-estimation procedure. Adaptive design combining the

use of the group sequential method and the sample size re-estimation method should

be a potential research topic.

Bibliography

1. Guidance for Industry: Statistical Approaches to Establishing Bioequivalence, U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), January 2001.

2. Chow SC and Liu JP. Design and Analysis of Bioavailability and Bioequivalence Studies. New York: Chapman and Hall/CRC, 2008, third Edition, ISBN-13: 000-1584886684.

3. Jones B and Kenward MG. Design and Analysis of Crossover Trials. New York: Chapman & Hall/CRC), 2003, ISBN-13: 978-0412300004.

4. Lacey LF, Keene ON, Pritchard JF, Bye A. Common noncompartmental pharmacokinetic variables: are they normally or log-normally distributed? Journal of Biopharmaceutical Statistics, 1997; 7(1): 171-178.

5. Shapiro, SS and Wilk, MB. An analysis of variance test for normality (complete samples). Biometrika 1965; 62: 591-611.

6. Liu JP and Weng CS. Evaluation of log-transformation in assessing bioequivalence, Communications in Statistics – Theory and Methods 1994; 23(2): 421-434.

7. Guidance published in 1992: Statistical procedures for bioequivalence studies using a standard two-treatment crossover design, U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), July 1992.

8. Westlake WJ. The design and analysis of comparative blood level trials. In: Swarbrick J. ed. Current concepts in the pharmaceutical sciences, dosage form design and bioavailability, Philadelphia: Lea and Febiger, 1973; 149-179.

9. Westlake WJ. Bioavailability and bioequivalence of pharmaceutical formulations. In: Peace KE. ed. Biopharmaceutical statistics for drug development. New York: Marcel Dekker, Inc., 1988; 329-352.

10. Schuirmann D. A Comparison of the Two One-Sided Tests Procedure and the Power Approach for Assessing the Equivalence of Average Bioavailability, Journal of Pharmacokinetics and Biopharmaceutics, 1987; 15(6).

11. Westlake WJ. Use of confidence intervals in analysis of comparative bioavailability trials, Journal of Pharmaceutical Sciences 1972; 61: 1340-1341.

12. Hauck WW and Anderson S. A New Statistical Procedure for Testing Equivalence in Two-Group Comparative Bioavailability Trials, Journal of Pharmacokinetics and Biopharmaceutics 1984; 12(1).

13. Gibaldi M and Perrier D. Pharmacokinetics. New York: Marcel Dekker, Inc., 1982.

14. Beal S, Sheiner LB, Boeckmann A, and Bauer RJ. NONMEM User's Guides. Ellicott City, MD: Icon Development Solutions, 2009.

15. Russek-Cohen E, Martinez MN, Nevius AB, A SAS/IML program for simulating pharmacokinetic data, Computer Methods and Programs in Biomedicine 2005; 78(1):39-60.

16. Comets E, Brendel K, Mentré F. Model evaluation in nonlinear mixed effect models, with applications to pharmacokinetics, Journal de la Société Francaise de Statistiques 2010; 151: 106-128.

17. Sheiner LB and Beal SL. Evaluation of methods for estimating population pharmacokinetic parameters. II. Biexponential model and experimental pharmacokinetic data, Journal of Pharmacokinetics and Biopharmaceutics 1981; 9(5): 635-51.

18. Davidian M and David MG. Nonlinear Models for Repeated Measurement Data. Bury St Edmunds, Suffolk: St Edmundsbury Press, 1995, ISBN 0 412 98431 9.

19. Uhlenbeck GE and Ornstein LS. On the theory of Brownian Motion. Physical Review 1930; 36: 823–841.

20. Kaye CM and Nicholls B. Clinical Pharmacokinetics of Ropinirole, Clinical Pharmacokinetics 2000; 39 (4): 243-254.

21. Shapiro SS, Wilk MB and Chen HJ. A Comparative Study of Various Tests for Normality. Journal of the American Statistical Association 1968; 63 (324):1343-1372.

22. SAS Institute Inc. SAS Procedure Guide, Version 9.2. Cary, NC: SAS Institute Inc., 2014.

23. Royston JP. Approximating the Shapiro-Wilk's W-Test for Non-normality. Statistics and Computing 1992; 2: 117–119.

24. Snedecor GW and Cochran WG. Statistical methods. Ames, IA: Blackwell publishing, 1989, eighth editions.

25. Chakravarti MI, Laha RG, and Roy J. Handbook of Methods of Applied Statistics, Volume I. New York: 1967, John Wiley and Sons.

26. Anderson TW and Darling DA. A test of goodness of fit. Journal of the American Statistical Association 1954; 49: 765-69.

27. Cramer H. Mathematical Methods of Statistics. Princeton, NJ: 1946, Princeton University Press.

28. Lilliefors HW. On the Kolmogorov-Smirnov Tests for Normality with Mean and Variance Unknown, Journal of the American Statistical Association 1967; 62: 399-402.

29. Lilliefors HW. On the Kolmogorov-Smirnov Tests for the Ex- ponential Distribution with Mean Unknown, Journal of the American Statistical Association 1969; 64: 387-89.

30. Stephens MA. EDF Statistics for Goodness of Fit and Some Comparisons, Journal of the American Statistical Association 1974; 69: 730-737.

31. D'Agostino RB and Stephens MA. Goodness-of-Fit Techniques, New York: 1986, Marcel Dekker, Inc.

32. Sheather SJ and Jones MC. A reliable data-based bandwidth selection method for kernel density estimation. Journal of the Royal Statistical Society, Series B 1991; 683–690.

33. Kramer et al. Pharmacokinetics of Digoxin: Comparison of a Two- and a Three-Compartment Model in Man, Journal of Pharmacokinetics and Biopharmaceutics, 1974; 2(.4).

34. Berger RL and Hsu J. Bioequivalence Trials, Intersection-Union Tests and Equivalence Confidence Sets, Statistical Science 1996; 11 (4): 283-319.

35. Owen DB. A special case of a bivariate non-central t-distribution. Biometrika 1965; 52: 437-446.

36. Phillips KF. Power of the two one-sided tests procedure in bioequivalence. Journal of Pharmacokinetics and Biopharmaceutics. 1990; 18: 137-143.

37. Diletti E, Hauschke D, and Steinijans VW. Sample size determination for bioequivalence assessment by means of confidence intervals. International Journal of Clinical Pharmacology, Therapy, and Toxicology1991; 29: 1–8.

38. Liu JP and Chow SC. Sample size determination for the two one-sided tests procedure in bioequivalence. Journal of Pharmacokinetics and Biopharmaceutics 1992; 20: 101-104.

39. Hauschke D, Steinijans VW, Diletti E, and Burke M. Sample size determination for bioequivalence assessment using a multiplicative

model. Journal of Pharmacokinetics and Biopharmaceutics 1992; 20: 557–561.

40. Kieser M and Hauschke D. Approximate sample sizes for testing hypotheses about the ratio and difference of two means. Journal of Biopharmaceutical Statistics 1999; 9: 641–650.

41. Kieser M and Hauschke D. Statistical methods for demonstrating equivalence in crossover trials based on the ratio of two location parameters. Drug Information Journal 2000; 34: 563–568.

42. Chow SC; Wang H. On Sample Size Calculation in Bioequivalence Trials. Journal of Pharmacokinetics and Pharmacodynamics 2001; 28, 155–169.

43. Hauschke D, Steinijans V and Pigeot I. Bioequivalence Studies in Drug Development: Methods and Applications. West Sussex, England: 2007, John Wiley & Sons, Ltd., ISBN: 978-0-470-09475-4.

44. Phillips KF. Power for Testing Multiple Instances of the Two One-Sided Tests Procedure. The International Journal of Biostatistics 2009; 5(1), Article 15.

45. Kelley K and Lai K. MBESS, http://nd.edu/~kkelley/site/MBESS.html, accessed December 2014.

46. Potvin D, DiLiberti CE, Hauck WW, Parr AF, Schuirmann DJ, Smith RA. Sequential design approaches for bioequivalence studies with crossover designs. Pharmaceutical Statistics 2008; 7: 245–262.

47. Golkowski D, Friede T, and Kieser M. Blinded sample size re-estimation in crossover bioequivalence trials, Pharmaceutical Statistics 2014; 13(3): 157-162.

48. Shen M, Russek-Cohen E, and Slud EV. Exact calculation of power and sample size in bioequivalence studies using two one-sided tests, Pharmaceutical Statistics 2015; 14 (2): 95-101.

49. Liu F and Li Q. Exact sequential test of equivalence hypothesis based on bivariate non-central t-statistics, Computational Statistics and Data Analysis 2014; 77:14–24.

50. Casella G and Berger RL. Statistical Inference. Pacific grove, CA: Cengage Learning, 2002, second edition, ISBN: 0534243126.

51. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria: 2013, URL http://www.R-project.org/.

52. Blume H and Midha KK. Report of Consensus Meeting: Bio-international '92. Conference on Bioavailability, Bioequivalence and Pharmacokinetic Studies. Bad Homburg, Germany, 20-22 May, 1992. European Journal of Pharmaceutical Sciences 1993; 1: 165-171.

53. Blume H, McGilveray I, Midha KK. Report of Consensus Meeting: Bio-international '94, Conf. on Bioavailability, Bioequivalence and Pharmacokinetic Studies, Munich. 14-18 June, 1994. European Journal of Pharmaceutical Sciences 1995; 3: 113-124.

54. Davit B, Nwakama P, Buehler G, Conner D, Haidar S, Patel D, Yang Y, Lawrence Y, Woodcock J. Comparing Generic and Innovator Drugs: A Review of 12 Years of Bioequivalence Data from the United States Food and Drug Administration, Annals of Pharmacotherapy 2009; 43: 1583-97.

55. Stein C. A two-sample test for a linear hypothesis whose power is independent of the variance, Annals of Mathematical Statistics 1945, 16: 43-58.

56. Wittes J, Schabenberger O, Zucker D, Brittain E and Proschan M. Internal pilot studies I: type I error rate of the naive t-test', Statistics in Medicine 1999, 18: 348-3491

57. Friede T and Kieser M. Blinded sample size reassessment in noninferiority and equivalence trials. Statistics in Medicine 2003; 22: 995–1007.

58. Glimm E and Läuter J. Some notes on blinded sample size re-estimation. arXiv: 1301.4167v1 [stat.ME].

59. Gould A L and Shih W. Sample size re-estimation without unblinding for normally distributed outcomes with unknown variance, Communications in Statistics – Theory and Methods 1992, 21: 2833-2853.

60. Apostol TM. Calculus, Vol. 1: One-Variable Calculus, with an Introduction to Linear Algebra. New York: 1991, Wiley, second edition, ISBN-10: 0471000051.

61. Atkinson K. An Introduction to Numerical Analysis. New York: 1989, Wiley, second edition, ISBN-10: 0471624896.

62. Azzalini A. The R 'sn' package: The skew-normal and skew-t

distributions (version 1.1-0). 2014, URL http://azzalini.stat.unipd.it/SN

Index of Notation

AUC: Area under the concentration-time curve

C_{ij}: Concentration at the jth time (t_{ij}) for the ith subject

Cmax: Maximum concentration in the concentration profile

CV: Coefficient variation for untransformed concentration data

D: Dose amount, mg

$\hat{D}$: Average of the averages of the intra-subject differences in *Y* between the test and the reference for the two sequences

d_{lik} : Difference between Period 2 and Period 1 for Subject *k* in Sequence *i* at Stage *l*, *l*=1,2

$\bar{d}_h$: Mean of the period differences of two sequences at Stage *l*, *l*=1,2

e_{ij}: Normal random measurement error at the jth time (t_{ij}) for the ith subject

F : Bioavailability fraction

$F(x,\boldsymbol{\theta})$: Distribution function

f(t$_{ij}$, $\boldsymbol{\beta_i}$): A nonlinear function

F_R: μ_R

F_T: μ_T

ka: Apparent first-order absorption rate constant

ke: Apparent first-order elimination rate constant

n_1: Number of subjects in Sequence 1 for one-stage study

n_2: Number of subjects in Sequence 2 for one-stage study

n_{li} : Sample size of the ith sequence at the lth stage $\forall i, l = 1,2$

$n_{l\bullet}$: Total sample size $\left(n_{l1} + n_{l2}\right)$ at Stage l, l=1,2

$P_1\!\left(\theta^*, \sigma\right)$: Probability of rejecting the null hypothesis in a two one-sided hypothesis tests procedure for a single endpoint

$P_1^{Power}\left(\theta^*, \sigma\right)$: Probability of rejecting the null hypothesis in power approach for a single endpoint

$P_{1s}\!\left(\theta_2, n_{11}, n_{12}, V\right)$: Joint probability of rejecting H_0 in Equation (5.1) under H_0 for one-stage study and $S_1 \leq S_0$

$P_{2s}\!\left(\theta_2, n_{11}, n_{12}, V, p_0, u\right)$: Joint probability of rejecting H_0 in Equation (5.1) under H_0 for two-stage study ($n_{2\bullet} > 0$) and $S_1 > S_0$

S^2: Pooled estimate of the variance of an intra-subject difference in Y

S_0^2: Initial variance value for $\sigma_T^2 + \sigma_R^2$ from the historical data

S_l^2: Sample variance for $\sigma_T^2 + \sigma_R^2$ from Stage l, l=1,2

S^*: Pooled variance of Stage 1 and Stage 2, $\sqrt{\dfrac{\left(n_{1\bullet} - 2\right)S_1^2 + \left(n_{2\bullet} - 2\right)S_2^2}{n_{1\bullet} + n_{2\bullet} - 4}}$

t_{ij}: The jth sampling time for the ith subject

Tmax: Time to reach maximum concentration

T_{21}^*: Test statistics for combined data of Stage 1 and Stage 2

T_{22}^*: Test statistics for combined data of Stage 1 and Stage 2

V: $\sqrt{\sigma_T^2 + \sigma_R^2}$

V_a: Apparent volume of distribution

W: Shapiro and Wilk's normality testing statistics

$Y:$ log*(AUC)* or log*(Cmax)*

μ_{C_t}: True concentration of a drug

μ_T: Population mean of Y for the test product

μ_R: Population mean of Y for the reference product

ν: Degrees of freedom

$\theta^*:$ $\mu_T - \mu_R$

θ_1: Lower limit of bioequivalence margin

θ_2: Upper limit of bioequivalence margin

σ_{ij}^2: Variance of $\log(C_{ij})$ at the jth time (t_{ij}) for the ith subject

σ_R^2: Variance of Y for reference product

σ_T^2: Variance of Y for the test product

$\mathbf{b_i}$: Normal random vector of inter-subject variation

$\mathbf{e_i}$: Errors of the ith subject

$\mathbf{g_i(\beta_i)}$: Functions of the ith subject

$\mathbf{\beta_i}$: ($p \times 1$) vector of pharmacokinetic parameters

$\mathbf{R_i}$: Variance-covariance matrix of log-transformed data within the ith subject

$\mathbf{\gamma}$: Positive vector of population pharmacokinetic parameters

$\mathbf{\Sigma}$: Covariance matrix