INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

DISSERTATION

STATISTICAL ISSUES IN INDIVIDUAL AND POPULATION BIOEQUIVALENCE

Submitted by

Richard J. McNally

Department of Statistics

In partial fulfillment of the requirements

for the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Fall 2002

COLORADO STATE UNIVERSITY

October 18, 2002

WE HEREBY RECOMMEND THAT THE DISSERTATION STATISTICAL ISSUES IN INDIVIDUAL AND POPULATION BIOEQUIVALENCE PREPARED UNDER OUR SUPERVISION BY RICHARD J. MCNALLY BE ACCEPTED AS FULFILLING IN PART REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY.

Committee on Graduate Work

Committee Member

Committee Member

Committee Member

Adviser

Department Head

ABSTRACT

STATISTICAL ISSUES IN INDIVIDUAL AND POPULATION BIOEQUIVALENCE

The US Food and Drug Adminstration (FDA) has proposed new regulations that address the 'prescribability' and 'switchability' of new formulations of already- or nearly-approved drugs. These new criteria are known, respectively, as population and individual bioequivalence. The subject of bioequivalence has been fertile ground for statistical research over the past 3 decades, and the introduction of these new criteria has created new statistical problems to be explored. The focus of this investigation is the development of new tests for individual and population bioequivalence based on the generalized p-value (GPV) methodology of Tsui and Weerahandi (JASA, 1989). We show that the GPV has good statistical properties and avoids a major shortcoming of other testing procedures for individual bioequivalence. In particular, the GPV is shown to be more powerful than any of the known testing procedures for population bioequivalence, including the test recommended by the FDA, for likely bioequivalence study designs. We also examine the asymptotic relative efficiencies of likely candidate designs for both individual and population bioequivalence, and compare these large-sample results with empirical calculations derived from power simulations. We also construct generalized confidence intervals for bioequivalence parameters and examine their properties.

Richard J. McNally
Department of Statistics
Colorado State University
Fort Collins, Colorado 80523
Fall 2002

ACKNOWLEDGEMENTS

I would like to thank Jennifer Hoeting and Jim Loftis for serving on my dissertation committee, and Phil Chapman for being on my committee and supporting me for the last 2 years. I want to thank my advisor, Hari Iyer, both for his guidance and friendship through the dissertation process, and for a few hard (well, for him) runs.

I should thank Thomas Mathew of the University of Maryland, Baltimore County, for introducing us to generalized p-values, for the his ideas on dealing with statistically dependent estimates of variance parameters, and for other insightful comments on Chapter 3.

My career in statistics may have never advanced to this point if not for three people who believed in me enough to give me a chance when I needed it. They are Dr. Stan Altan of Johnson and Johnson Pharmaceutical Research and Development Institute, Raritan N.J, Dr. Ray Buck of Cato Research Ltd., Durham NC, and Professor Emeritus Duane Boes of CSU. I would also like to thank everyone who encouraged me to pursue a Ph.D., including Dr. Dhamika Amaratunga of Johnson and Johnson Pharmaceutical Research and Development, and, of course, my mother, Yone E. McNally. I should also thank Ray Buck for his helpful comments on Chapter 1.

I am also grateful to the Biopharmaceutical Section of the American Statistical, and to the Society of Clinical Trials, who saw fit to honor the work found in Chapter 4 and Chapter 6, respectively, of my dissertation.

Finally, I would to thank my sweetheart and companion, Misty Lees, for all her love and support. Misty's hard work through the last few difficult years made the task of pursuing my dream at CSU much, much easier.

DEDICATION

Thomas Kevin McNally

25 August 1920–27 August 2001

CONTENTS

LIST OF FIGURES

Chapter 1

INTRODUCTION

1.1 What is Bioequivalence?

Before a pharmaceutical product can be marketed in North America and the European Union, it must be studied extensively to demonstrate the effectiveness and safety of its intended use. Part of the evaluation process of a new chemical entity (NCE) is called pharmacokinetics, which is a detailed characterization of how a drug product is absorbed, distributed, metabolized, and excreted in biological systems. In humans, pharmaceutical compounds and metabolites of interest are studied in body fluids that can be collected without the use of overly invasive procedures. Particular attention is paid to the presence of the therapeutic moiety in the circulatory system. Pharmaceutical researchers are interested in the *bioavailability* of a compound and/or its active metabolite(s). Bioavailability characterizes the absorption of an specific dosage form through the intended route of administration into the circulation of an organism. In this investigation, we are concerned with the bioavailability of oral dosage forms, *i.e.*, how well orally administered products are absorbed through the gastrointestinal tract into the systemic circulation. Knowing the pharmacokinetics of a compound is vital in determining the dosing regimen that should be used. It also has implications for the both efficacy and safety of a product.

The development of a pharmaceutical product is a long process, which usually culminates in large-scale clinical studies that provide the bulk of the proof that a product is effective and safe. Because of the considerable expense that pharmaceutical manufacturers incur in bringing new and beneficial products to market, they are afforded a period of exclusivity under US patent law. Once a patent expires for a drug product, other manufacturers are allowed to market the same product. The products produced by the latter

group of manufacturers are referred to as generic drugs. The current authority to approve and to regulate generic drugs in this country was given by the US Congress to the Food and Drug Administration (FDA) through the Drug Price Competition and Patent Term Restoration Act of 1984. Similar agencies in the European Union regulate generic drugs in those jurisdictions. It is the responsibility of these regulatory agencies to determine the conditions under which generic drugs may be marketed.

The chemical structure and properties of the NCE are documented in the New Drug Application (NDA) when it is submitted to the FDA for approval. The NDA also lists the excipients that are used to formulate the drug product. However, exactly how the drug product is formulated is proprietary. Although an innovator must make certain disclosures concerning the manufacturing process to assure the FDA that this process is adequately controlled to gain approval, there are many aspects of the manufacturing process that are not subject to public disclosure. Therefore, it cannot be taken for granted that any generic manufacturer can produce a product that exactly, or even sufficiently, duplicates the action of the innovator's product or another generic product. Therefore, before they can market their formulation of a specific product, the generic manufacturer must provide the FDA with evidence that the generic product can be expected to be as safe and effective as the innovator's product or any other product on the market. There are a few products that may be approved on the basis of *in vitro* dissolution testing. For most products, however, addition evidence of equivalence to the brand name product that must provided. This evidence is derived through comparative bioavailability studies, also known as bioequivalence (BE) studies. Allowing generic products to be approved based on the evidence of bioequivalence studies circumvents the need for expensive and time-consuming clinical studies. This often leads to drugs being available on the market at lower prices. The chief rationale for approvals based only on comparative pharmacokinetic data, without the benefit of direct clinical observation, is the *fundamental bioequivalence assumption.*

> When two drug products are equivalent in the rate and extent to which the active drug ingredient or therapeutic moiety is absorbed and becomes available

> at the site of drug action, it is assumed that they will be therapeutically equivalent (Chow and Liu, 2000, pg. 14).

Therefore, evidence that a generic product has equivalent bioavailability to the innovator product is considered evidence that the generic will have the same therapeutic effect and the same safety profile as the innovator. The evidence of bioequivalence contained in these studies is submitted to the FDA in an Abbreviated New Drug Application (ANDA). It should be noted that the fundamental bioequivalence assumption only applies to products regulated by the FDA Center for Drug Evaluation and Research (CDER). There are currently no generic biologic products (*i.e.*, those products regulated by the Center for Biologic Evaluation and Research (CBER)). For example, a company cannot market its own version of erythropoietin (EPO), a hormone important in the production of red blood cells, based on the evidence of a bioequivalence study.

There are other situations in which bioequivalence studies are important. Before receiving marketing approval, the drug product used in clinical trials is usually made in small batches under conditions quite different from full commercial production. Therefore, the NDA must contain evidence that drug product from large-scale production batches is equivalent to the much smaller batches used in clinical trials. The FDA may also require a manufacturer to demonstrate bioequivalence after a significant change in the manufacturing process or a change in the site of manufacture. Therefore, situations in which a manufacturer, even an innovator, must show its own products are bioequivalent are not uncommon.

1.2 Bioequivalence Basics: Pharmacokinetics

This section is intended to give the basic concepts of bioequivalence and pharmacokinetics necessary for understanding what follows. Pharmacokinetics is the science of how drugs are absorbed, distributed, metabolized, and excreted by biological organisms. It is sometimes characterized as what the body does to the drug, as opposed to *pharmacodynamics*, which is what the drug does to the body. As part of the preclinical development

of a drug product, its pharmacokinetics are studied in one or more mammalian species (*e.g.*, rats, dogs), by administering a radioisotope-enhanced version of the drug to groups of animals. The absorption, distribution, metabolism, and excretion of the drug are assessed by the assaying the quantity of radio-labelled drug in various target organs and body fluids at various times. In human pharmacokinetic studies, we are necessarily limited to assaying the quantity of the drug product in the blood (more properly, either whole blood or blood plasma) and excreta. Of particular interest is how well orally administered products are absorbed from the gastrointestinal tract into the systemic circulation and how quickly they are eliminated. The assumption here is that the pharmacological action of a therapeutic moiety at a given time is directly related to the quantity of it in the blood or plasma.

The most common instrument for assessing bioequivalence in humans is a crossover study with a single dose per treatment period. These studies involve healthy volunteers administered either the test (generic) or reference (innovator) product in each treatment period. They usually enroll 12 to 40 subjects evaluated over 2 to 4 treatment periods. Such studies are generally accepted as the safest and most sensitive method of evaluating bioequivalence.

During each treatment period, the absorption of the drug product is assessed by taking plasma samples from each subject at several times, and then assaying the samples to determine the concentration of the drug or other metabolites of interest in the plasma. Although the choice of sampling times depends on the absorption characteristics of a product, generally blood sampling is more frequent in the first few hours after administration of the drug product, and then less frequent later on. The rate and extent of the absorption of a drug product into the systemic circulation is assessed using the following pharmacokinetic metrics:

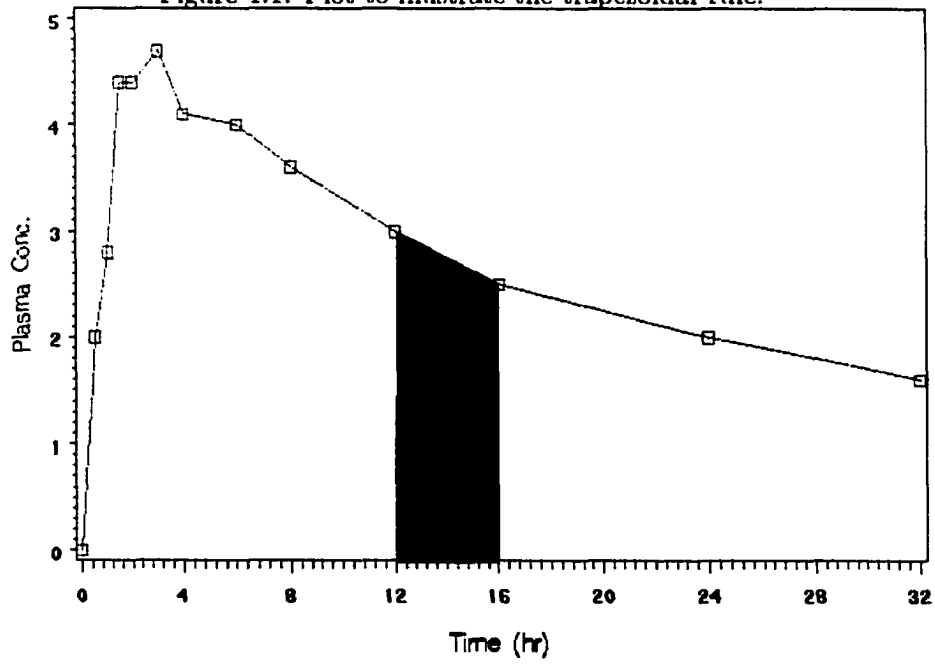
1. AUC–the area under the plasma/blood concentration-time curve-the primary measure of the extent of drug absorption. AUC can be measured from time of administration to the time of the last blood sample (AUC_{0-t}), or to when it is, in

theory, completely eliminated from the systemic circulation ($AUC_{0-\infty}$). The latter parameter requires calculating the elimination rate constant of the drug.

2. Cmax–the maximum plasma concentration-the primary measure of the rate of drug absorption.

3. Tmax–the time at which Cmax occurs. The Tmax, in conjunction with the Cmax, characterizes the rate of drug absorption.

Figure 1.1: Plot to illustrate the trapezoidal rule.



Throughout our discussion of bioequivalence, we will be interested primarily in the parameters AUC and Cmax. AUC is calculated for each subject during each treatment period using the trapezoidal rule for numerical integration. The trapezoidal rule is illustrated in Figure 1.1. The dark shaded area shows the contribution of the time period from Hour 12 to Hour 16 to the total AUC, which is the sum of all of the trapezoids that can be construct from the curve in Figure 1.1. The Cmax is estimated by the maximum

observed plasma concentration for a given subject in a given study period. The increased sampling frequency in the period immediately following product dosing should ensure that the observed Cmax is reasonably close to its true value. The sampling frequency in this period must also be sufficient to ensure that the observed Cmax does not occur at the first sampling time (FDA, 2000). The estimate of the Tmax is then measured on a discrete scale, *i.e.*, it must occur at one of the sampling points. Because of this and the large number of ties in the Tmax data, this parameter is usually evaluated using non-parametric or categorical methods. Although sponsors (*i.e.*, drug manufacturers) are not strictly required to show that their product is equivalent to the reference product with respect to Tmax, it is customary to do some assessment of this metric for bioequivalence.

Statistical approaches to bioequivalence have evolved over the past few decades as generic drugs have become an important public health issue in the United States and in other developed nations. The general idea is that we wish to show that the test formulation T, frequently a generic product, is equivalent to a reference formulation R, most usually a brand name product already on the market. Because of the potentially high variability of the responses between human subjects to the above parameters, bioequivalence of orally administered drugs is assessed almost exclusively using crossover studies.

Since we are trying to demonstrate equivalence, the proper null hypothesis in the bioequivalence problem is one of non-equivalence. But in the early history of bioequivalence, it was not generally recognized that demonstrating equivalence is different from the usual hypothesis testing setting, where one sets out to show that two treatments are different by first assuming that they are the same. Therefore, some early approaches to this problem had to be abandoned when they were subsequently found to have little scientific or statistical bases. One early bioequivalence approach, the 75/75 rule, concluded that two formulations were bioequivalent if the individual treatment ratios, Y_T/Y_R, were between 0.75 and 1.25 for at least 75% of the subjects in a study. This approach was intuitively appealing, but was dropped when it was shown that it had poor statistical properties (see Haynes, 1981) and little scientific basis. Subsequently, an *ad hoc* power approach was adopted, where two formulations were bioequivalent if the standard hypothesis test of no

difference in the means of the formulations (*i.e.*,$H_0 : \mu_T - \mu_R = 0$ where the μ's represent the means of the respective formulations) was insignificant and if one could show that the study had at least 80% power to show at least a 20% difference in formulation means at the $\alpha = 0.05$ level. Schuirmann (1987) showed that this approach had the undesirable property that it was easier it was to show bioequivalence if a product had higher variability. He introduced the two one-sided tests (TOST) approach, under which it is more likely that one would conclude bioequivalence if the mean difference of the formulations has lower variability. The TOST approach, explained in detail below, became the methodology preferred by the FDA, and it is recommended in the guidance document *Statistical Procedures for Bioequivalence Studies Using a Standard Two-Treatment Crossover Design* (1992). Because the TOST approach is equivalent to showing that a $100(1 - 2\alpha)\%$ confidence interval is completely contained in the equivalence region, it is a proper test of the null hypothesis of non-equivalence (*cf.* Hsu *et al.*, 1994).

Three procedures strongly recommended in the 1992 FDA guidance document formed the foundation of the analysis of bioequivalence studies for the decade that followed. The first was the log-transformation of subject responses. The second was Schuirmann's two one-sided tests (TOST) approach. The third was the use of a 2 sequence/2 period crossover design as the standard study design for bioequivalence studies. With the log-transformation, the parameter of interest becomes the difference in treatment means (*i.e.*, $\mu_T - \mu_R$, where μ_k is the mean population response to formulation k), instead of the ratio (μ_T/μ_R). For responses on the log scale, the bioequivalence interval for which a product was set at $(-0.2231, 0.2231)$, which corresponds to an interval of $(0.80, 1.25)$ on the ratio scale. The two one-sided tests approach, with each test at the $\alpha = 0.05$ level, is equivalent to constructing a 90% confidence interval on $\mu_T - \mu_R$, and then determining whether the confidence interval is completely contained in the bioequivalence interval. Since the parameter that is tested in this approach concerns only the treatment means, formulations that pass the two one-sided tests criterion (or any similar test on $\mu_T - \mu_R$) are said to have average bioequivalence. The adoption of the 2×2 crossover as the standard study design for bioequivalence studies made the conduct of these studies relatively easy and straightforward.

1.3 Individual and Population Bioequivalence

By the time the FDA issued the 1992 guidance document on statistical procedures for bioequivalence, the adequacy of average bioequivalence to provide sufficient evidence of the safety and effectiveness of generic drug products was already being questioned. For example, see Hauck and Anderson (1991). It was argued that average bioequivalence could not completely ensure that the same drug products formulated by different manufacturers were interchangeable. With respect to individual patients, this lead to the concept of 'switchability', which is meant to address the question 'Will a patient whose prescription is switched from the brand name to the generic version of a drug product derive the same therapeutic effect from the generic as they would from the innovator product?' Switching between generics produces similar concerns. The mechanics of the innovator-to-generic switch is especially critical for patients who have been titrated to a specific dose of the brand name product and whose control of disease state and toleration of treatment depends on maintaining the concentration of the active moiety in their systemic circulation within a certain range. When the therapeutic window, the range within which plasma levels of a drug or active metabolite must be kept in order for a patient to receive therapeutic benefit without harmful side effects, is relatively small, consistency from one product to another and from one dose to another for a particular patient becomes crucial. The product-to-product variation is called the subject-formulation interaction. The dose-to-dose variation is addressed by considering the differences in the within-subject variance for the different formulations. To assess individual bioequivalence, we must consider both the subject-formulation interaction and the difference in within-subject variability.

The subject-formulation interaction, which will be denoted by σ_D, is used to quantify the variability of the each individual's ratio of response to the test formulation versus to the reference formulation. For example, suppose that the mean difference in individual responses to two formulations (*i.e.*, $\delta = \mu_{ijT} - \mu_{ijR}$, where i is the sequence number and j is the subject within a sequence) is normally distributed with mean 0 and variance σ_D^2. When $\sigma_D = 0.1$, less than 3% of individuals will have a true equivalence ratio outside of

the interval (-0.2231,0.2231). This corresponds, on the original scale, to the ratios being outside of the interval (0.8,1.25), which has been the standard region for bioequivalence since the TOST was adopted. When $\sigma_D = 0.15$, the proportion of patients outside the bioequivalence interval rises to almost 14%. With the potential for so many patients to have a mean response to a generic product outside of the interval where it can be deemed equivalent to the brand name, regulatory agencies become concerned when the point estimate of σ_D reaches this point. On the other hand, the variability of the point estimate of σ_D tends to increase as the variability of the formulations under study increases. Endrenyi, Taback, and Tothfalusi (2000) noted that the variability of the estimate of σ_D increases as the within-subject variance for the reference formulation (σ^2_{WR}) increases. This makes putting a regulatory upper bound on σ_D problematic. As we will see, σ_D is a function of the difference in the between-subject variances and of the correlation of the subject's response to different formulations. A relatively high value of σ_D can signify either a relatively high difference in the between-subject variation for the two formulations, or relatively low correlation of the responses within a subject from dose-to-dose, or both of these things.

In individual bioequivalence, we are also interested in how the within-subject variance differs between the formulations under consideration. It may even be considered beneficial if the new formulation has a smaller within-subject variance. There are two issues are present in the assessment of within-subject variation.

1. We need replication of formulations within a subject to estimate within-subject variance (and the subject-formulation interaction). This leads to bioequivalence studies with higher order crossover designs, such as 3 and 4-period designs.

2. Confidence intervals on the ratios of variances tend to be wide, even in studies with large sample sizes. An approach to equivalence that incorporates the elements of interest into a single function may be more desirable.

Bioequivalence studies tend to be much smaller than typical clinical trials. As stated above the standard study design had been the 2-sequence/2-period crossover design, and

the number of subjects is typically in the range 12 to 40. Because of the increased cost of running longer studies with potentially more subjects, the move to 3 and 4-period designs would be more acceptable if efficient designs are implemented. Mandating the assessment of variance ratios within the equivalence paradigm could curtail or effectively eliminate the development of generic drugs, or could cause the cost of generic drugs to increase. Both of these circumstances could have repercussions throughout the public health system.

Population bioequivalence addresses the issue of whether a patient who is prescribed a drug product *de novo* will derive the same therapeutic benefit as they would had the product that was used in the clinical trials from which the efficacy and safety of the product was concluded. This may be necessary since the clinical trial material was likely produced on a much smaller scale than the marketed product. The underlying idea of population bioequivalence is that differences in the 'ultimate' (*i.e., steady-state*) concentrations of two drug formulations depend on differences in the total variability of these formulations, where the total variability is the sum of the between-subject and within-subject variances. Population bioequivalence does not require within-subject replication, so the standard 2×2 crossover design is sufficient for its assessment.

1.4 The Need for Individual Bioequivalence

Many statisticians in the pharmaceutical industry have questioned the need for mandating the assessment of individual bioequivalence. In 1999, Dr. Jane Henney, the commissioner of the FDA, stated in the Journal of the American Medical Association that 'if the FDA declares [on the basis of average bioequivalence] a generic drug to be therapeutically equivalent to an innovator drug, the two products will provide the same intended clinical effect.' The pharmaceutical industry, through various trade organizations, seems to have taken a strong stance against individual bioequivalence, at least the fashion in which the FDA has chosen to implement it. The purpose of this section is to give a short argument on why individual bioequivalence addresses legitimate public health concerns.

Gould (2000) asserts that documented cases involving the failure of the average bioequivalence criterion are 'very rare'. We have found examples that suggest otherwise. The

first is from a report of study conducted by Meyer *et al* (2000), who examined the individual bioequivalence of brand name and generic formulations of methylphenidate after reports of adverse drug reactions in patients were linked to a certain generic formulation. Meyer *et al* found that although the brand name and generic in question were average bioequivalent (*i.e.*, they were bioequivalent under the guidelines issued by the FDA in 1992) with respect to AUC and Cmax, and the formulations were individual bioequivalent with respect to AUC, they were not individual bioequivalent with respect to Cmax. It appeared that the generic had higher dose-to-dose variation and higher maximum plasma concentrations in most patients. In some patients, the plasma concentrations increased to the point where they experienced reactions such as vomiting and dizziness.

More disturbing is the testimony of Dr. Stephen Schachter on behalf of the Epilepsy Foundation to the opening meeting of the FDA's Pharmaceutical Sciences Advisory Committee (1999), which was convened to discuss the then proposed FDA guidance document on individual and population bioequivalence.

> Let's focus on phenytoin and typical example of this problem [of switching both to and between generic formulations]. Suppose that a patient needs a minimum blood level of 10 to control seizures and cannot have level over 15 in order to avoid disabling side effects. Let's say this patient is on DILANTIN^{TM}, which is the brand name phenytoin. On three DILANTIN a day, the patient has a level of 10. Everything is fine. The patient is then switched to a generic version and her blood level drops to 8, resulting in seizures. This generic was approved by the FDA because it fell within the -20% to $+25\%$ rule of the FDA. The doctor checks a blood test, verifies that the level has dropped, and then increases the daily dose to bring the level back up to 10. Again, everything was fine. Two months later, the pharmacy fills the prescription with another company's generic phenytoin, and now the level goes up to 20. Now the patient complains of side effects, and another blood test, doctor's visit, and possibly an emergency room visit result. This other generic was also approved by the FDA. An additional change in dose is made and the cycle repeats.

Dr. Schachter then cites an actual example where a patient who was switched to a generic phenytoin required hospitalization for their seizures. The hospital bill was nearly $4000, while the cost difference between brand name and generic was $4. Dr. Schachter also cites a survey of epilepsy patients, in which 41% of those who were switched from a brand name to a generic epilepsy medication had worse seizure control on the generic, and 32% had side effects after being switched to the generic.

Both of these examples involve drug products that are used to treat neurological conditions. The epilepsy example might be considered exceptional, because of the well-documented sensitivity of patients to blood levels of the products used to treat their condition. However, it is an example that may be becoming less exceptional, for the FDA, in recent years, has approved products with narrow therapeutic indices (the ratio of the maximum tolerated dose to the minimum effective dose) for conditions such as Alzheimer's disease and schizophrenia. As these products come off patent, they will be subject to generic competition. The potentially serious side effects associated with some of these breakthrough treatments may have prompted the FDA to be proactive in enacting their new bioequivalence regulations, so that the effect on the patient population of generic products entering the market for, say, schizophrenia may be minimized.

1.5 Statistics and Bioequivalence

There are many aspects to the debate on individual and population bioequivalence. Most involve the medical sciences and public health, but some reach into economic and even political issues. The common thread is that these are all fields in which statistical methods have been of great utility. The purpose of the following investigations is to address some of the statistical issues that have arisen in individual and population bioequivalence. We will concentrate on how the statistics relate to pharmacology and bioavailability. Our hope is that the new methods developed here and our investigations of the properties of study designs for bioequivalence will have an impact, not just on the medical and scientific end, but also on public health and economic aspects of bioequivalence.

In the following chapters, we will address the statistical issues of bioequivalence related to the pharmacology of the problem. Chapter 2 sets the statistical foundations for our investigation, and details the current state of the statistical evaluation of individual and population bioequivalence. It also introduces the *generalized p-value* (GPV) within the context of linear mixed models, which will become the basic tools in our evaluation of bioequivalence. Chapter 3 introduces the application of the GPV for assessing individual and population bioequivalence in studies with 4-period crossover designs. Chapter 4 does the same for 3-period crossover, and Chapter 5 looks at the evaluation of population bioequivalence in the standard 2×2 crossover design. Chapter 6 compares the efficiency of candidate crossover designs for bioequivalence studies in which individual and population bioequivalence are evaluated. Chapter 7 will explore the application of *generalized confidence intervals* to individual and population bioequivalence parameters. The major conclusions and summary of our contributions will be contained in Chapter 8, along with suggestions for possible future directions for research.

Chapter 2

STATISTICAL BACKGROUND FOR INDIVIDUAL AND POPULATION BIOEQUIVALENCE

The new FDA guidance document on statistical aspects of bioequivalence testing (US Food and Drug Administration, 2001) is intended to contain most of the necessary information needed to do the proper statistical assessments for bioequivalence studies. It details the criteria, basic study designs, methodology, and other issues that arise in the analysis of bioequivalence. In this chapter, we introduce the FDA's recommended statistical model, criteria, and methodology for individual bioequivalence (IBE) and population bioequivalence (PBE). Then we look at some other testing procedures from the statistical literature for assessing IBE and PBE. Finally, we introduce the generalized p-value (GPV, Tsui and Weerahandi, 1989), which we will use in subsequent chapters to construct tests for the individual and population bioequivalence parameters recommended by the FDA, and the allied concept of the *generalized confidence interval* (GCI, Weerahandi, 1993), which we will use to construct intervals for these bioequivalence parameters.

2.1 Study Design and Statistical Model

The designs that we will consider for the evaluation of individual and population bioequivalence are in Table 1. The first design is a 2-sequence 4-period design (2S4P [we will use this notation, where the first number is the number of sequences and the second is the number of period, throughout]). This is the design preferred by the FDA for the assessment of individual bioequivalence (FDA, 2001). In this design, each subject receives each formulation twice. Variants of this design with 4 and 6 sequences are also used. The second design is a 2-sequence, 3-period design (2S3P) that is cited in the current FDA

statistical guidance document for bioequivalence. We will sometimes call this design the 3-period dual design. A *dual* design is one where each sequence has a counter-sequence with the opposing treatment (formulation) in each period. The 4-period designs that we will study are also dual designs. Under certain conditions, the optimal design for assessment of mean effects is a dual design (Kenward and Jones, 1989), but our experimental situation dictates that we cannot make any of the assumptions that lead to this type of optimality, nor are such assumptions particularly relevant. A property that is apparently relevant to our experimental situation is *uniform within sequence* or *sequence-balanced* (Chinchilli and Esinhart, 1996), which dictates that each sequence have each treatment the same number of times, so that the same information is extracted from each sequence. The third design in Table 1 is a 3-period design that is sequence-balanced, the 3-sequence, 3-period design (3S3P). We will also refer to this design the extra-reference design, since the reference formulation appears twice in each sequence.

The assessment of individual bioequivalence requires study designs with some degree of within subject replication of treatments. All 3 designs discussed in the previous paragraph can be used to assess IBE. The fourth design in Table 1, the 2-sequence, 2-period design (2S2P) cannot be used to assess IBE, since the primary IBE criteria are not estimable in this design. We will still consider this design, however, for the assessment of population bioequivalence.

Table 1. Designs for Individual and Population Bioequivalence

2-sequence/4-period (2S4P)

Sequence 1	T	R	T	R
Sequence 2	R	T	R	T

2-sequence/3-period (2S3P)

Sequence 1	T	R	T
Sequence 2	R	T	R

3-sequence/3-period (3S3P)

Sequence 1	T	R	R
Sequence 2	R	T	R
Sequence 3	R	R	T

2-sequence/2-period (2S2P)

Sequence 1	T	R
Sequence 2	R	T

In all of the designs that we will consider, each subject receives each formulation at least once. This removes from consideration the 4-sequence, 2-period design (4S2P) frequently known as Balaam's design. Balaam's design is a standard 2×2 (2S2P) crossover with uniform sequences of the test (TT) and reference (RR) treatments appended to it. This makes all of the necessary IBE parameters estimable. However, since bioequivalence studies usually have a small number of subjects, there is a possibility of bias due to the presence of variant subgroups in one of the sequences (Chen *et al.*, 2000). Therefore, the FDA strongly advises against the used of Balaam's design for the evaluation of bioequivalence.

The observational model from Chinchilli and Esinhart (1996) is given below.

$$Y_{ijkl} = \mu_k + \gamma_{ikl} + \eta_{ijk} + \epsilon_{ijkl} \quad i = 1, \ldots, s; \quad j = 1, \ldots, n_i; \quad k = T, R; \quad l = 1, 2;$$

Here Y_{ijkl} is the response from subject j in sequence i to the l^{th} application of treatment k, and s is the number of sequences. The parameters μ_T and μ_R are the population mean responses to treatments T and R and γ_{ikl} is the (unknown) fixed effect corresponding to the l^{th} application of treatment k in sequence i. The following estimability conditions are imposed on the γ_{ikl}.

$$\sum_{i=1}^{s} \sum_{l=1}^{2} \gamma_{ikl} = 0 \quad \text{for } k = T, R$$

Furthermore, η_{ijk} is a random effect of subject j in sequence i when receiving treatment k, and ϵ_{ijkl} are within-subject random errors. The random variables ϵ_{ijkl} are jointly independent and distributed normally with mean 0 and variance σ^2_{Wk} $(k = R, T)$, and the vectors $\boldsymbol{\eta}_{ij} = [\eta_{ijT}, \eta_{ijR}]'$ are mutually independent bivariate normal with zero means and a variance matrix given by

$$\boldsymbol{V} = \begin{pmatrix} \sigma^2_{BT} & \rho\sigma_{BT}\sigma_{BR} \\ \rho\sigma_{BT}\sigma_{BR} & \sigma^2_{BR} \end{pmatrix}.$$

Finally, $\{\boldsymbol{\eta}_{ij}\}$ and $\{\epsilon_{ijkl}\}$ are mutually independent. For convenience, we also define $\delta = \mu_T - \mu_R$, $\sigma^2_D = \sigma^2_{BT} + \sigma^2_{BR} - 2\rho\sigma_{BT}\sigma_{BR}$, $\sigma^2_{TT} = \sigma^2_{BT} + \sigma^2_{WT}$, and $\sigma^2_{RR} = \sigma^2_{BR} + \sigma^2_{WR}$.

2.2 Criteria for Individual and Population Bioequivalence

The criteria for individual and population bioequivalence are based on distance metrics proposed by Sheiner (1992) and by Schall and Luus (1993). Consider the following squared distances

$$d(Y_T, Y_R)^2 = E(Y_T - Y_R)^2$$
$$d(Y_R, Y'_R)^2 = E(Y_R - Y'_R)^2.$$

The first expression is the distance between the test formulation T and the reference formulation R. Because of the inherent variability of the process, this distance is compared to the distance between any 2 applications of the reference formulation. Sheiner (1992) proposed comparing a function of the ratio of 2 very similar distance functions–the Sheiner criterion is based on the distance from a constant *target* value T. He referred to as a relative risk ratio

$$\Theta_1 = \frac{E(Y_T - T)^2}{E(Y_R - T)^2}.$$

When we take the target value to be $T = \mu_R$, then Sheiner's criterion can also be expressed as

$$\Theta_1 = \frac{E(Y_T - Y_R)^2}{0.5E(Y_R - Y'_R)^2} - 1.$$

This proposed criterion suffered from a lack of statistical power for IBE: even for a 4-period study with a total of 40 subjects, the test of relative individual risk (RIR) he proposed

had only 16% power for concluding bioequivalence when the regulatory upper bound for the RIR was set at 1.44.

Schall and Luus (1993) proposed examining the difference of the distances, that is

$$\Theta_2 = E(Y_T - Y_R)^2 - E(Y'_R - Y_R)^2.$$

The Schall and Luus criterion appeared to be better with respect to statistical power, since it apparently requires much fewer subjects to be reasonably assured of concluding bioequivalence.

The evaluation of these metrics is based on whether one is interested in individual or population bioequivalence. We will demonstrate by evaluating the Schall and Luus criterion for both IBE and PBE. For IBE, the evaluation of the expectations is conditioned on the individual subjects, yielding

$$\begin{aligned}\Theta_{I2} &= E(Y_T - Y_R)^2 - E(Y'_R - Y_R)^2 \\ &= (\delta^2 + \sigma_D^2 + \sigma_{WT}^2 + \sigma_{WR}^2) - (2\sigma_{WR}^2) \\ &= \delta^2 + \sigma_D^2 + \sigma_{WT}^2 - \sigma_{WR}^2,\end{aligned}$$

where $\delta = \mu_T - \mu_R$ is substituted for the mean difference between the formulations, and $\sigma_D^2 = \sigma_{BT}^2 + \sigma_{BR}^2 - 2\rho\sigma_{BT}\sigma_{BR}$ is the subject-formulation interaction. This form is the manner in which the Schall and Luus IBE criterion is usually presented. For PBE, the distance expression is evaluated by conditioning on the subject population, yielding

$$\begin{aligned}\Theta_{P2} &= E(Y_T - Y_R)^2 - E(Y'_R - Y_R)^2 \\ &= (\delta^2 + \sigma_{TT}^2 + \sigma_{RR}^2) - (2\sigma_{RR}^2) \\ &= \delta^2 + \sigma_{TT}^2 - \sigma_{RR}^2,\end{aligned}$$

where $\sigma_{TT}^2 = \sigma_{BT}^2 + \sigma_{WT}^2$ and $\sigma_{RR}^2 = \sigma_{BR}^2 + \sigma_{WR}^2$ are called the total variances of the respective formulations, since they the sum of the between-subject and within-subject variances. Again, the latter expression is the most common form of expressing the Schall and Luus PBE criterion.

Evaluating the Sheiner criteria in the same manner yields the following expression for IBE and PBE, respectively.

$$\Theta_{I1} = \frac{\delta^2 + \sigma_D^2 + \sigma_{WT}^2}{\sigma_{WR}^2}$$

$$\Theta_{P1} = \frac{\delta^2 + \sigma_{TT}^2}{\sigma_{RR}^2}$$

The distance interpretation of these criteria make them intuitively appealing. Schall and Luus were able to show, by using actual bioequivalence studies as examples, that one could conclude IBE and PBE with a relatively small number of subjects. But it appears that variability of this metric is directly proportional to the components themselves, which makes it problematic to mandate a single upper bound that is widely-applicable over a range of likely values. The FDA has addressed this problem by dividing the Schall and Luus criteria by an appropriate scale factor, so that they become

$$\Theta_{IBE} = \frac{\delta^2 + \sigma_D^2 + \sigma_{WT}^2 - \sigma_{WR}^2}{\sigma_{WR}^2}$$

for IBE, and

$$\Theta_{PBE} = \frac{\delta^2 + \sigma_{TT}^2 - \sigma_{RR}^2}{\sigma_{RR}^2}$$

for PBE. Since the scale factor is based on the variability of the reference product, these criteria are called reference-scaled. Thus the FDA criterion is a hybrid of Θ_1 with its scaling properties, and Θ_2, which is based on an absolute difference. But now there is the potential that these criteria put an unfair burden on test formulations that are compared to reference products with relatively low variation. We illustrate this problem with the following example: suppose that Manufacturer A is comparing their generic to the product of Innovator A, and that the bioequivalence parameters have the values $\delta = 0.05, \sigma_D^2 = 0.01, \sigma_{WT}^2 = 0.03$, and $\sigma_{WR}^2 = 0.02$. We then have $\Theta_{IBE} = 1.125$. These parameters are all well within the allowances (discussed below) set by the FDA to determine the regulatory upper bound. Now consider Manufacturer B, who is comparing their generic with Innovator B, and that the bioequivalence parameters have the values $\delta = 0.05, \sigma_D^2 = 0.01, \sigma_{WT}^2 = 0.25$, and $\sigma_{WR}^2 = 0.2$. The value of $\Theta_{IBE} = 0.5625$ is much lower than for the Generic A comparison, despite that the difference in within subject variances

is much higher than 0.02, the allowance for the difference in these parameters given by the FDA. Therefore, Manufacturer A may have a more difficult time than Manufacturer B showing that their generic is equivalent to its respective innovator product, despite having a formulation that is 'more equivalent'. The solution of the FDA was to put a lower bound on the denominator of the criterion. This results in the FDA mixed-scaled criteria for IBE

$$\Theta_{IBE} = \frac{\delta^2 + \sigma_D^2 + \sigma_{WT}^2 - \sigma_{WR}^2}{max(\sigma_{WR}^2, \sigma_{W0}^2)},$$

and for PBE

$$\Theta_{PBE} = \frac{\delta^2 + \sigma_{TT}^2 - \sigma_{RR}^2}{max(\sigma_{RR}^2, \sigma_{T0}^2)}.$$

The FDA derived the value for σ_{W0}^2 by solving the expression

$$\frac{[log(1.25)]^2}{\sigma_{W0}^2} = 1.25$$

for σ_{W0}^2, which gives $\sigma_{W0}^2 = 0.0398344$, which they rounded to 0.04. A very similar expression for σ_{T0}^2 also yields a value of 0.04. Since the constant scale factor is the same for IBE and PBE, we will often refer to the constant-scale factor merely as σ_0^2.

To derive the regulatory upper bounds for Θ_{IBE} and Θ_{PBE}, the agency plugged-in various allowances for each set of parameters into the appropriate criterion. For IBE, the allowances are $\delta = log(1.25)$, $\sigma_D^2 \in [0.02, 0.03]$, and $\sigma_{WT}^2 - \sigma_{WR}^2 = 0.02$. Setting $\sigma_D^2 = 0.03$ and the denominator equal to $\sigma_{W0} = 0.04$, this process yields

$$\theta_I = \frac{(log(1.25))^2 + 0.03 + 0.02}{0.04} = 2.4948.$$

Similarly, for PBE we have the same allowance for δ, $\sigma_{TT}^2 - \sigma_{RR}^2$, and the same denominator (since $\sigma_{T0}^2 = \sigma_{W0}^2$ to yield

$$\theta_P = \frac{(log(1.25))^2 + 0.02}{0.04} = 1.7448.$$

The above criteria are referred to as aggregate criteria, since the parameter of interest is an aggregation of individual model parameters. Some researchers (*e.g.*, Gould, 2000, Chow and Liu, 2000), have proposed assessing the individual model parameters, such as δ and σ_D^2, separately. This is known as the disaggregate criteria approach. Some of the

problems with disaggregate criteria are that the statistical properties of the estimates of some parameters are not well understood, as for σ_D^2, and that using certain parameters, *e.g.*, $\sigma_{WT}^2/\sigma_{WR}^2$, to assess bioequivalence may lead to dramatic increase in samples sizes, which would have a drastic effect on how bioequivalence studies are conducted. Critics of the aggregate approach, in particular the FDA criteria (*e.g.*, Hsuan, 2000), argue that tradeoffs between parameters in the FDA criteria may lead to, for example, concluding individual bioequivalence for two formulations that are not even close to being average bioequivalent. The FDA has specifically addressed this last criticism by requiring some form of average bioequivalence as a prerequisite to IBE and PBE. The FDA also defends tradeoffs between parameters as, in part, consequences of the estimation process.

2.3 The Statistical Methodology of the FDA Guidance Document

The original draft of the current FDA guidance document on statistical methods for bioequivalence recommended using the non-parametric bootstrap (Efron and Tibshirani, 1992) for assessing Θ_{IBE} and Θ_{PBE}. The statistical methodology of the final FDA guidance document is based on the work of Hyslop, Hsuan, and Holder (2000), which in turn is based on the work of Howe (1974), who was interested in constructing a confidence interval for between-subjects variance component (frequently denoted as σ_a^2) in the one-way random effects model. Howe's idea was that one could construct an approximate $100(1-\alpha)\%$ interval on σ_a^2 by constructing exact intervals on the constituents $n\sigma_a^2 + \sigma_e^2$ and σ_e^2, and then combining the lengths of the respective intervals to get the length of the interval for σ_a^2. Howe showed that his interval had many desirable properties, making it superior to those of Moriguti (1954) and Bulmer (1957). Apparently Howe was not familiar with the work of Tukey (1951) and Williams (1962) on this problem. Howe's approach, however, was more general, and he speculated that his work could be extended to general linear combinations, as was done by Hyslop *et al.* (2000) for the FDA individual bioequivalence criterion Θ_{IBE}. To do so, these authors had to linearize the criterion by multiplying both sides by the appropriate scale factor, either σ_{WR}^2 or σ_{W0}^2. They also reparameterized in terms of $\sigma_I^2 = \sigma_D^2 + 0.5\sigma_{WT}^2 + 0.5\sigma_{WR}^2$ (*n.b.*, $\sigma_I^2 = Var(\hat{\delta})$), and then put all the non-zero

terms in the IBE expression on the same side of the inequality, so that the IBE criterion becomes

$$\delta^2 + \sigma_I^2 + 0.5\sigma_{WT}^2 - 1.5\sigma_{WR}^2 - \theta_I max(\sigma_{WR}^2, \sigma_{W0}^2) \leq 0.$$

The linearized criterion is used to construct a $100(1-\alpha)\%$ lower confidence interval. This is done by constructing exact $100(1-\alpha)\%$ intervals for each of the components, lower confidence intervals (upper bounds) for components with a positive sign, and upper confidence intervals (lower bounds) for components with negative signs. The half-lengths (*i.e.*, the distance between the point estimate and the bound) of each interval are squared, summed, and the square root of the resulting variance estimate added to the point estimate $\hat{\gamma}$ to get an approximate $100(1-\alpha)\%$ interval. This results in the following upper bound for the lower $100(1-\alpha)\%$ intervals:

$$\hat{\gamma}_R + (H_D^2 + H_I^2 + H_{WT}^2 + H_{WRR}^2)^{0.5}$$

for the reference-scaled criterion, and

$$\hat{\gamma}_C + (H_D^2 + H_I^2 + H_{WT}^2 + H_{WRC}^2)^{0.5}$$

for the constant-scaled criterion, where

$$\hat{\gamma}_R = \hat{\delta}^2 + s_I^2 + 0.5s_{WT}^2 - (1.5 + \theta_I)s_{WR}^2$$
$$\hat{\gamma}_C = \hat{\delta}^2 + s_I^2 + 0.5s_{WT}^2 - 1.5s_{WR}^2 - \theta_I\sigma_{W0}^2$$
$$H_D = \left((\overline{Y}_T - \overline{Y}_R) + t_{\alpha,\nu}\sqrt{\frac{s_I^2}{n_1+n_2}} \right)^2 - (\overline{Y}_T - \overline{Y}_R)$$

$$H_I = s_I^2\left(\frac{\nu}{\chi_{\alpha,\nu}^2} - 1\right) \qquad H_{WT} = 0.5s_{WT}^2\left(\frac{\nu}{\chi_{\alpha,\nu}^2} - 1\right)$$
$$H_{WRR} = -(1.5 + \theta_I)s_{WR}^2\left(\frac{\nu}{\chi_{1-\alpha,\nu}^2} - 1\right) \qquad H_{WRC} = -1.5s_{WR}^2\left(\frac{\nu}{\chi_{1-\alpha,\nu}^2} - 1\right)$$

The decision rule is to conclude IBE if the upper bound of this interval is less than 0. A similar formulation can be used to assess PBE, but adapting the procedure in this way will, as we shall see, be problematic. We refer to this procedure for IBE as the Howe-type small sample confidence interval (HSCI) test, since the methodology is based on the work

of Howe. Alternatively, we will also at times refer to this and similar testing procedures as the FDA test, particularly in the context of PBE.

By all accounts, the HSCI test for IBE appears to perform very well. Hyslop *et al* showed that, in most cases, it holds its size at or below the nominal, and that it has better power and smaller sample size requirements than bootstrap procedures for the 4-period design that they studied. HSCI tests have also been independently developed for 3-period crossover designs by Hyslop and Iglewicz (2001), and by Chow, Shao, and Wang (2002). The HSCI test is also relatively straightforward to implement. It does, however, have one drawback: the upper bound that is used as the test criterion is discontinuous when one switches from constant to reference-scaled criterion (FDA, 2001). Of the various solutions proposed for this problem, the FDA guidance document allows the so-called 'play-the-winner' rule, in which one computes upper bounds for both reference and constant-scaled criteria, then chooses the more favorable (to the conclusion of IBE) result. This has the effect of slightly inflating the Type I error of the test. A test procedure that is not discontinuous at the changeover point from constant to reference-scale criterion may be considered a marked improvement.

As well as the HSCI test performs for IBE, it may be surprising that the adaptation of this procedure in the FDA guidance document to PBE is suspect. The title of Howe's paper specifies that he was interested in confidence intervals for linear combinations of independent random variables. But the random variables that are formulated in the assessment of Θ_{PBE} are not all independent. This is why we refer to the testing procedure as the FDA test in this context. We will see that ignoring the dependence structure of the random variable has a detrimental effect on the power of the FDA test, compared to other procedures, such as the bootstrap, which do not ignore this dependence.

2.4 Modified Large Sample Tests for IBE and PBE

Graybill and Wang (1980) developed methods for constructing confidence intervals on linear combinations of variance components with non-negative coefficients which they dubbed as *modified large sample* (MLS) intervals. Over the following decade and a half,

Graybill, Wang, Burdirk, and their associates extended the MLS methodology to a variety of functions of variance components and to a variety of experimental designs. Most of these methods are cataloged in the book by Burdirk and Graybill (1992).

One problem in adapting MLS intervals to bioequivalence is that Θ_{IBE} and Θ_{PBE} are not functions of variance components alone: the mean difference $\delta = \mu_T - \mu_R$ appears as a squared term in both of these parameters. The fact that δ^2 has a non-central χ^2 distribution further complicates the matter. Quiroz *et al.* (2002) address this situation by following Patnaik (1949) in equating the non-central χ^2 with a central χ^2 whose degrees of freedom are estimated by setting the first 2 moments of the non-central χ^2 to the first 2 moments of a central χ^2 distribution. They then use Satterthwaite's method (1946) to combine terms with like signs, and then compute an upper $100(1-\alpha)\%$ bound using the appropriate MLS method. For reference-scaling, this is the method that Lu, Graybill, and Burdirk (1989) developed for rational functions of variance components. The resulting $(1-\alpha)\%$ upper bounds are

$$U_{IBER} = \frac{k_1}{k_2}\left[\frac{2 + k_5/(k_1 k_2) + \sqrt{V_{M4}}}{2(1 - k_4/k_2^2)}\right] - 1.5,$$

where

$$\begin{aligned}
V_{M4} &= (2 + k_5/(k_1 k_2))^2 - 4(1 - k_4/k_2^2)(1 - k_3/k_1^2),\\
k_1 &= S_D^2/(sn) + [1 - 1/(sn)]S_I^2 + S_T^2/2,\\
k_2 &= S_R^2,\\
k_3 &= (F_{\alpha:\infty,n^*} - 1)^2[S_D^2/(sn)]^2 + (F_{\alpha:\infty,n_I} - 1)^2[(1 - 1/(sn))S_I^2]^2\\
&\quad + (F_{\alpha:\infty,n_T} - 1)^2(S_T^2/2)^2,\\
k_4 &= [(1 - F_{1-\alpha:\infty,n_R})S_R^2]^2,\\
k_5 &= S_R^2[k_6 S_D^2/(sn) + k_7(1 - 1/(sn))S_I^2 + k_8 S_T^2/2],\\
k_6 &= \frac{(1 - F_{1-\alpha:n^*,n_R})^2 - (F_{\alpha:\infty,n^*} - 1)^2 F_{1-\alpha:n^*,n_R}^2 - (1 - F_{1-\alpha:\infty,n_R})^2}{F_{1-\alpha:n^*,n_R}},\\
k_7 &= \frac{(1 - F_{1-\alpha:n_I,n_R})^2 - (F_{\alpha:\infty,n_I} - 1)^2 F_{1-\alpha:n_I,n_R}^2 - (1 - F_{1-\alpha:\infty,n_R})^2}{F_{1-\alpha:n_I,n_R}},\\
k_8 &= \frac{(1 - F_{1-\alpha:n_T,n_R})^2 - (F_{\alpha:\infty,n_T} - 1)^2 F_{1-\alpha:n_T,n_R}^2 - (1 - F_{1-\alpha:\infty,n_R})^2}{F_{1-\alpha:n_T,n_R}}.
\end{aligned}$$

For constant-scaling, they use the lower interval (*i.e.* upper bound) of Ting *et al.* (1990) for linear combinations of variance components with coefficients unrestricted in sign. The resulting $(1-\alpha)\%$ upper bounds are

$$U_{IBEC} = \frac{k_1 - 1.5k_2 + \sqrt{V_{M5}}}{\sigma^2_{W0}},$$

where

$$V_{M5} = k_3 + 1.5^2 k_4 + 1.5k_5.$$

Quiroz *et al.* compared test procedures for IBE and PBE based on their MLS intervals to the HSCI test. They found that HSCI test had slightly better power for IBE, while the MLS intervals produced slightly more powerful tests for PBE. The MLS intervals are also on the same scale as the parameters Θ_{IBE} and Θ_{PBE}, not the linearized scale of the HSCI intervals. Quiroz *et al.* claim that this as an advantage of the MLS, although it is not clear, given the novelty of the bioequivalence criteria in question, how useful intervals on this scale are, except for comparison to a pre-specified regulatory bound.

2.5 Wang's Unbiased Test for IBE

The other method in the extant statistical literature for assessing IBE is an unbiased test developed by Wang (1999). His approach closely followed that of Brown, Hwang, and Munk (1997), who developed an unbiased test for average bioequivalence. Wang's approach requires special calculation of tolerance limits for each case. He also developed his test for a certain class of 3-period designs, even though the data example he uses is derived from a 4-period study, not addressing the fact that some of the secondary IBE parameters are non-identifiable in the design that he chooses, and he ignores the use of mixed-scaling, choosing to consider only the reference-scaled criterion. Both of the latter choices appeared to have been made not for scientific reasons, but purely for mathematical convenience. This is probably why his methodology has not gained wide acceptance. Therefore, we do not compare the tests we have derived to this test. It should be noted that the unbiased test of Brown *et al.* had also failed to replace the two one-sided tests approach of Schuirmann (1987) for average bioequivalence, and that the Wang test for IBE has not even been mentioned as an alternative in the FDA guidance document (2001). This appears to be due to the mathmatical complexity of both the Brown *et al.* and the Wang approaches.

2.6 Generalized p-Values

The remainder of this chapter is devoted to the methods that will be the primary tools in our approach to assessing IBE and PBE: generalized p-values and generalized confidence intervals.

Suppose we wish to test $H_0 : \theta \leq \theta_0$ versus $H_a : \theta > \theta_0$. For an observable random vector $\boldsymbol{X}$, Tsui and Weerahandi (1989) defined a generalized test variable (GTV) to be a function $T(\boldsymbol{X}; \boldsymbol{x}, \boldsymbol{\xi})$, where $\boldsymbol{x}$ is the observed value of the random vector $\boldsymbol{X}$, and $\boldsymbol{\xi} = (\theta, \boldsymbol{\zeta})$, with θ being a scalar parameter of interest, and $\boldsymbol{\zeta}$ a vector of one or more nuisance parameters. Tsui and Weerahandi imposed on $T(\boldsymbol{X}; \boldsymbol{x}, \boldsymbol{\xi})$ the following three requirements:

1. $T(\boldsymbol{x}; \boldsymbol{x}, \boldsymbol{\xi}) = t$ is free of $\boldsymbol{\xi}$.

2. For fixed $\boldsymbol{x}$ and $\boldsymbol{\xi}$, the distribution of $T(\boldsymbol{X}; \boldsymbol{x}, \boldsymbol{\xi})$ is free of the nuisance parameter $\boldsymbol{\zeta}$.

3. For fixed $\boldsymbol{x}$ and $\boldsymbol{\zeta}$, $Pr(T(\boldsymbol{X}; \boldsymbol{x}, \boldsymbol{\xi}) \geq t \mid \theta)$ is non-decreasing in θ.

Using the function $T(\boldsymbol{X}; \boldsymbol{x}, \boldsymbol{\xi})$, the generalized p-value for the test of H_0 against H_a is defined as

$$p = \sup_{\theta \leq \theta_0} Pr[T(\boldsymbol{X}; \boldsymbol{x}, \boldsymbol{\xi}) > t|\theta]$$

From property (3) above, it follows that

$$p = \sup_{\theta \leq \theta_0} Pr[T(\boldsymbol{X}; \boldsymbol{x}, \boldsymbol{\xi}) > t|\theta] = Pr[T(\boldsymbol{X}; \boldsymbol{x}, \boldsymbol{\xi}) > t|\theta_0].$$

By properties (1) and (2), the generalized p-value is not dependent on nuisance parameters, and hence can be computed based on data alone. For a nominal α level test of H_0 against H_a one rejects H_0 whenever $p \leq \alpha$.

The above definitions of generalized test variable and generalized p-value are, respectively, analogous to those of test statistic and p-value in classical hypothesis testing.

Weerahandi (1995) refers to the tests derived using the GPV as exact tests in the following sense:

> With respect to a specified probability measure, a sample space, and the parameter of interest fixed at the value specified by the null hypothesis, *p-value is the exact probability of a well-defined extreme region* (a well-defined subset of the sample with the observed sample on its boundary)[.] (Weerahandi, 1995, pg. 112).

It should be noted that for many of the well-known exact tests of classical statistics, *i.e.*, cases where we have a test statistic with a sampling distribution that is both free of nuisance parameters and tabled, we can derive a GPV test that coincides with the classical test. An example of this is the test of $H_0 : \sigma_a^2 = 0$, where σ_a^2 is the between block variance component in the one-way random effect model. Under the null hypothesis, a natural GTV is $F = MSA/MSE$, which is the same as the classical test statistic, where the numerator and denominator are the block and error mean squares, respectively. Therefore, the GPV test corresponding to this GTV and and the classical test are the same. This lends credence to the notion that GPV methods lead to sensible testing procedures. The advantage of the GPV is that we can derive a test for the more general hypothesis $H_0 : \sigma_a^2 < \delta$, for which classical testing procedures do not exist when $\delta > 0$.

2.7 Generalized Confidence Intervals

Closely related to the GPV is the generalized confidence interval (GCI, Weerahandi, 1993). The GCI is based on the generalized pivotal quantity (GPQ), which is the closely relate to the GTV. Thus the GCI is the interval estimation counterpart of the GPV.

Weerahandi (1993) defined $R = r(\boldsymbol{X}; \boldsymbol{x}, \boldsymbol{\xi})$ as a *generalized pivotal quantity* (GPQ) if it satisfies the following two conditions

Property A: R has a probability distribution free of unknown parameters.

Property B: The *observed pivotal*, defined as $r_{obs} = r(\boldsymbol{x}; \boldsymbol{x}, \boldsymbol{\xi})$, does not depend on the nuisance parameter ζ.

Now, we define a set C_γ such that

$$Pr[R \in C_\gamma] = \gamma.$$

In direct analogy to the pivotal quantity method, we obtain a subset $C(\boldsymbol{X})$ of the sample space such that

$$Pr[X \in C(\boldsymbol{x})] = \gamma.$$

We further define a subset Θ_c of the parameter space Θ of θ such that

$$\Theta_c(r) = \{\theta \in \Theta | r(\boldsymbol{x}; \boldsymbol{x}, \boldsymbol{\xi}) \in C_\gamma\}.$$

The interval estimate $\Theta_c(r)$ derived from the generalized pivotal quantity R is a generalized confidence interval of θ. If we are, for example, interested in finding a 95% GCI on R, we would find L and U such that

$$Pr[L \leq R \leq U | R = r(\boldsymbol{x}; \boldsymbol{x}, \boldsymbol{\xi})] = 0.95.$$

If we take L and U such that $F_R(L) = 0.025$ and $F_R(U) = 0.975$, where $F_R(.)$ is the cumulative distribution function (CDF) of R, we get the equal-tails GCI of θ. We can also attempt to find L and U such that the length of the interval is minimized. When evaluating individual and population bioequivalence, we often have $L = -\infty$, so that, by finding U such that $F_R(U) = 0.95$, we get a 95% lower generalized confidence interval on θ. Since Properties A and B of a GPQ correspond to Properties 1 and 2 of a GTV, each GTV has an associated GPQ.

2.8 Comparison of Generalized p-Value and Bayesian Methods

The similarity between the generalized p-value and Bayesian methods has been observed by several authors, including Tsui and Weerahandi (1989) themselves. They noted that, in certain cases, the GPV solution is identical to the Bayesian solution derived using a non-informative prior such as Jeffreys' prior. An important case where this occurs is in the Behrens-Fisher problem, where Tsui and Weerahandi note that GPV solution is the same as the Jeffreys' Bayesian solution (Jeffreys, 1967), and the fiducial solution known to Fisher and his contemporaries. This section examines, in general, some of the connections and differences between the GPV and Bayesian methods.

We start with the definition of the generalized p-value. For testing the null hypothesis $H_0 : \theta \leq \theta_0$ versus the alternative $H_a : \theta > \theta_0$, if $T = T(\boldsymbol{X}, \boldsymbol{x}, \boldsymbol{\xi})$ is the generalized test variable which is stochastically increasing in θ (the parameter of interest), then the GPV is

$$p = Pr(T \geq t_{obs} | \theta = \theta_0)$$

where $t_{obs} = T(\boldsymbol{x}, \boldsymbol{x}, \boldsymbol{\xi}_0)$ and $\boldsymbol{\xi}_0 = (\theta_0, \boldsymbol{\zeta})$. This probability is taken over the joint distribution of $\boldsymbol{X}$ given H_0 regarding $\boldsymbol{x}$ as a known fixed constant vector. To facilitate the comparison to Bayesian methods, we need to provide some justification for making 'frequency' type statements within a Bayesian framework. This justification was argued for within the context of applied statistics by Rubin (1984), whose ideas were formalized by Meng (1994) into what the latter called the *posterior predictive p-value* (PPV).

The posterior predictive p-value is defined as follows: suppose that we have a test variable $T(X)$ that we will use to test the same null hypothesis that is in the previous paragraph. To compute a classical p-value, $T(X)$ needs to be a pivotal quantity, *i.e.*, free of unknown nuisance parameters. To compute a Bayesian p-value, we consider a 'future observation' x^{rep}, which is called a *replication* of the fixed observed value x of the random variable X. We then define a *discrepancy variable* $D(x, \psi)$ that is a function of the data and the parameters, the latter feature distinguishing it from a classical test statistic $T(x)$. The posterior predictive p-value is then defined as

$$p_B = Pr\{D(x^{rep}, \psi) \geq D(x, \psi) | x, H_0\},$$

with this probability taken over the joint posterior distribution of x^{rep} and ψ given H_0 and $\boldsymbol{x}$.

Although the interpretations are different, the similarities between the GPV and PPV are now apparent. Both are based on a test variable that is a function of both the data and the parameters. They differ in that, for the GPV, the probability calculation does not depend on the nuisance parameters, whereas the PPV calculation handles the nuisance parameters through the specification of the prior distribution. In both approaches, the selection of a test variable is critical. Because removing the dependence on nuisance

parameters is necessary in order to compute a GPV, Weerahandi (1993,1995) often appeals to invariance arguments to justify the choice of a generalized test variable. On the other hand, Meng (1994) seems to favor basing his discrepancy variables on likelihood ratios. The respective derivations of $T(\boldsymbol{X}, \boldsymbol{x}, \boldsymbol{\xi})$ and $D(x, \psi)$ may not always be straightforward. Although Meng does not discuss desirable properties for a discrepancy variable, it appears that choosing $D(x, \psi)$ so that it has some sort of monotonicity properties like the GTV would be advantageous.

It is noteworthy that Meng derived a slightly different solution to the Behren-Fisher problem from that of Jeffreys' (1967) and Tsui and Weerahandi (1989), but Meng then notes that the Jeffreys'/GPV solution can be obtained using the PPV if one ignores the information $\mu_1 = \mu_2$ in the null hypothesis.

In general, the GPV will not always correspond to the Bayesian solution derived using Jeffreys' prior. This was shown by Chiang (2001) for the related problem of constructing confidence intervals for general functions of variance components (Chiang's *method of surrogate variables* (SV) is the same as Weerahandi's generalized confidence interval methodology). Chiang observed that the fiducial distribution associated with the SV/GCI differs from the posterior distribution derived using Jeffrey's prior in that the use of Jeffreys' prior restricts the parameter space of the variance components to positive values. Chiang notes that this restriction of the parameter space can lead to some peculiar results. For example, Chiang's results seem to indicate that in the one-way random effects model, Bayesian intervals computed using the methodology of Wolfinger and Kass (2000) had poor coverage for the within-block variance component σ_a^2, compared to SV/GCI intervals, when σ_a^2 was close to zero. However, these results, as mentioned above, are based on a particular prior distribution. There seems to be no reason why one cannot choose a prior distribution so that the resulting posterior distribution matches the fiducial distribution given by Chiang. We could specify a prior distribution which allows the circumstance $\sigma_a^2 = 0$ to have a probability $p > 0$, making the Bayesian interval more general. Therefore, the question becomes not whether the sampling distribution associated with the GPV for a particular problem matches the posterior distribution derived using Jeffreys' prior for

that problem, but whether (and when) it is possible to find a prior distribution so that the Bayesian and GPV solutions will match each other.

The advantage of equating the GPV with established statistical methodologies is that we can then adopt the properties of these methodologies as properties of the GPV. The apparent relationship between the GPV and fiducial inference is useful in some circumstances, particularly with respect to the Behren-Fisher problem, but the lack of general results concerning fiducial distributions severely limits the applicability of this relationship. Establishing a strong link between the GPV and Bayesian methods, especially the posterior predictive p-value, would imbue the GPV with strong general properties such as average risk optimality. Weerahandi (1995) himself has attempted to exploit such links to claim that the results of Meng (1994) concerning the PPV also apply to the GPV. Unfortunately, at present this link between Bayesian methods and GPV only exists for those problems, such as the Behren-Fisher problem, where it has been shown that the numerical calculation of the GPV can be duplicated by some Bayesian computation. To make the desired link, we would need a general result concerning the existence of prior and posterior distributions that always generate 'p-values' numerically equivalent to the GPV. But a general proof is likely to be very difficult, especially for multi-parameter problems. Even a weaker result, such as one where additional properties are imposed upon the generalized test variable or the discrepancy variable (or, most likely both) would be extremely valuable.

2.9 Chapter Summary

This chapter was intended to lay the statistical foundation for developing methods to assess individual and population bioequivalence. We have examined the bioequivalence criteria and the statistical model recommended by the FDA. We have also presented a short survey of methodological approaches to this problem in the extant statistical literature. With this information and the scientific background given by the previous chapter, it is hoped that the reader will be sufficiently prepared to understand the content of the chapters that follow.

Chapter 3

TESTS FOR INDIVIDUAL AND POPULATION BIOEQUIVALENCE BASED ON GENERALIZED *P*-VALUES FOR 4-PERIOD STUDIES

SUMMARY

The US Food and Drug Administration (FDA) has proposed new regulations that address the 'prescribability' and 'switchability' of new formulations of already-approved drugs. These new criteria are known, respectively, as population and individual bioequivalence. Two methods have been proposed in the bioequivalence literature for assessing population and individual bioequivalence that calculate an upper 95% confidence bound for the bioequivalence criterion in question, and then test bioequivalence by comparing this bound to the limit established by the FDA. In this paper we propose applying the generalized test variable (GTV) methodology of Tsui and Weerahandi (JASA, 1989, pp 602–607) to this problem to produce tests based on a generalized p-value (GPV). This methodology allows us to construct hypothesis tests in the presence of nuisance parameters. Using simulation we show that these tests perform well in comparison to the confidence interval methods, and have superior power for assessing population bioequivalence.

3.1 Introduction

In this chapter we examine the generalized p-value approach for making inferences concerning the FDA-recommended PBE and IBE criteria. The underlying statistical model is described in Section 3.2. The Generalized p-value approach is described, along with the notation we use, in Section 3.3. Sections 3.4 and 3.5 present the generalized test variables for IBE and PBE, respectively. Examples are also given for illustrating the application of the proposed methods, with direct comparisons to the HSCI and MLS methods. Section 3.6 discusses the results of a simulation study we conducted for evaluating the performances of the GPV methods. Comparisons to the HSCI and the MLS methods are included. Section 3.7 discusses some of the advantages of the GPV approach over the existing methods.

Hyslop, Hsuan, and Holder (HHH, 2000) only applied the Howe (1974) methodology to IBE. The extension of this method to PBE is given in the FDA guidance document (2001), using, for the most part, the notation of HHH. Quiroz *et al.* (QTWB, 2002) have addressed both IBE and PBE, using their own notation. It is assumed that the reader is familiar with or can refer to the methods in the FDA guidance document (which are largely the methods of HHH), and that the reader is also familiar with MLS methods for constructing confidence intervals for variance components. For information on the latter, one can refer to the book by Burdick and Graybill (1992). It is further assumed that the reader is familar with the concepts relating to the evaluation of pharmacokinetics and bioequivalence.

3.2 Study Design and Statistical Model

Bioequivalence studies almost always have some type of crossover design, where each subject is administered each formulation of the study drug at least once. The crossover designs that we will consider in this chapter are the 4-period designs given in Table 1.

Table 1a

FDA-recommended 2-sequence 4-period design

Sequence 1	T	R	T	R
Sequence 2	R	T	R	T

Table 1b

FDA-recommended 4-sequence 4-period design

Sequence 1	T	R	R	T
Sequence 2	R	T	T	R
Sequence 3	T	T	R	R
Sequence 4	R	R	T	T

The observational model for 4-period crossover designs is given below.

$$Y_{ijkl} = \mu_k + \gamma_{ikl} + \eta_{ijk} + \epsilon_{ijkl} \quad i = 1, \ldots, s; \quad j = 1, \ldots, n_i; \quad k = T, R; \quad l = 1, 2;$$

where Y_{ijkl} is the response from subject j in sequence i to the l^{th} application of treatment k, and s is the number of sequences. The parameters μ_T and μ_R are the population mean responses to treatments T and R and γ_{ikl} is the (unknown) fixed effect corresponding to the l^{th} application of treatment k in sequence i. The following estimability conditions are imposed on the γ_{ikl} (see Chinchilli and Esinhart, 1996).

$$\sum_{i=1}^{s}\sum_{l=1}^{2}\gamma_{ikl} = 0 \qquad \text{for } k = T, R$$

Furthermore, η_{ijk} is a random effect of subject j in sequence i when receiving treatment k, and ϵ_{ijkl} are within-subject random errors. The random variables ϵ_{ijkl} are jointly independent and distributed normally with mean 0 and variance σ^2_{Wk} $(k = R, T)$, and the vectors $\boldsymbol{\eta}_{ij} = [\eta_{ijT}, \eta_{ijR}]'$ are mutually independent bivariate normal with zero means and a variance matrix given by

$$\boldsymbol{V} = \begin{pmatrix} \sigma^2_{BT} & \rho\sigma_{BT}\sigma_{BR} \\ \rho\sigma_{BT}\sigma_{BR} & \sigma^2_{BR} \end{pmatrix}.$$

Finally, $\{\boldsymbol{\eta}_{ij}\}$ and $\{\epsilon_{ijkl}\}$ are mutually independent. For convenience, we also define $\delta = \mu_T - \mu_R$, $\sigma^2_D = \sigma^2_{BT} + \sigma^2_{BR} - 2\rho\sigma_{BT}\sigma_{BR}$, $\sigma^2_{TT} = \sigma^2_{BT} + \sigma^2_{WT}$, and $\sigma^2_{RR} = \sigma^2_{BR} + \sigma^2_{WR}$.

The methods under consideration by the FDA are based on the following parameters, one for PBE and one for IBE. The parameters of interest, Θ_{PBE} and Θ_{IBE}, are given in Chapter 2. For PBE, we use the following reparameterization of the Θ_{PBE}

$$\Theta_{PBE} = \begin{cases} \dfrac{\delta^2 + \sigma_T^2 + 0.5\sigma_{WT}^2 - \sigma_R^2 - 0.5\sigma_{WR}^2}{\sigma_R^2 + 0.5\sigma_{WR}^2} & \text{when } \sigma_R^2 + 0.5\sigma_{WR}^2 > \sigma_{T0}^2 \\ & \text{(reference-scaled criterion)} \\ \dfrac{\delta^2 + \sigma_T^2 + 0.5\sigma_{WT}^2 - \sigma_R^2 - 0.5\sigma_{WR}^2}{\sigma_{T0}^2} & \text{when } \sigma_R^2 + 0.5\sigma_{WR}^2 \leq \sigma_{T0}^2 \\ & \text{(constant-scaled criterion)} \end{cases}$$

where $\sigma_k^2 = \sigma_{Bk}^2 + 0.5\sigma_{Wk}, k = T, R$. The reference-scaled criterion is to be applied when the total variance $\sigma_R^2 + 0.5\sigma_{WR}^2$ of the reference formulation is greater than or equal to a FDA specified threshold value σ_{T0}^2. Otherwise, the constant-scaled criterion is to be used. At present, the FDA recommended value for σ_{T0}^2 is 0.04. Population bioequivalence is concluded if it can be demonstrated that $\Theta_{PBE} < \Theta_P$ where the FDA recommended value for Θ_P is 1.7448.

Likewise, for IBE, the reparameterization we will use is

$$\Theta_{IBE} = \begin{cases} \dfrac{\delta^2 + \sigma_I^2 + 0.5\sigma_{WT}^2 - 1.5\sigma_{WR}^2}{\sigma_{WR}^2} & \text{when } \sigma_{WR}^2 > \sigma_{W0}^2 \\ & \text{(reference-scaled criterion)} \\ \dfrac{\delta^2 + \sigma_I^2 + 0.5\sigma_{WT}^2 - 1.5\sigma_{WR}^2}{\sigma_{W0}^2} & \text{when } \sigma_{WR}^2 \leq \sigma_{W0}^2 \\ & \text{(constant-scaled criterion)} \end{cases}$$

where $\sigma_I^2 = \sigma_D^2 + 0.5\sigma_{WT}^2 + 0.5\sigma_{WR}^2$. The reference-scaled criterion for IBE is to be applied when σ_{WR}^2 is greater than or equal to a FDA specified threshold value σ_{W0}^2. Otherwise, the constant-scaled criterion is to be used. At present, the FDA recommended value for σ_{W0}^2 is 0.04. Individual bioequivalence is concluded if it can be demonstrated that $\Theta_{IBE} < \Theta_I$. Currently, the FDA recommends a value of 2.4948 for Θ_I.

The quantity σ_D^2, defined by Ekbohm and Melander (1989) to assess the subject-formulation interaction, is key to determining IBE. If we examine the identity

$$\sigma_D^2 = (\sigma_{BT} - \sigma_{BR})^2 + 2(1 - \rho)\sigma_{BT}\sigma_{BR}$$

we see that σ_D^2 is small when (1) σ_{BT} is close to σ_{BR}, and (2) ρ is close to 1. Thus the role of the correlation coefficient ρ in determining IBE is apparent. Although its role is

less apparent in determining PBE, we will see later that ρ is very important in this case as well.

3.3 The Generalized p-Value

The generalized test variable is the key to computing a generalized p-value which provides a measure of discrepancy between data and the null hypothesis. Suppose we wish to test $H_0 : \theta \geq \theta_0$ versus $H_a : \theta < \theta_0$. For an observable random vector $\boldsymbol{X}$, Tsui and Weerahandi (1989) defined a generalized test variable (GTV) to be a function $T(\boldsymbol{X}; \boldsymbol{x}, \boldsymbol{\xi})$, where $\boldsymbol{x}$ is the observed value of the random vector $\boldsymbol{X}$, and $\boldsymbol{\xi} = (\theta, \boldsymbol{\zeta})$, with θ being a scalar parameter of interest, and $\boldsymbol{\zeta}$ a vector of one or more nuisance parameters. Tsui and Weerahandi imposed on $T(\boldsymbol{X}; \boldsymbol{x}, \boldsymbol{\xi})$ the following three requirements:

1. $T(\boldsymbol{x}; \boldsymbol{x}, \boldsymbol{\xi}) = t$ is free of $\boldsymbol{\xi}$.

2. For fixed $\boldsymbol{x}$ and $\boldsymbol{\xi}$, the distribution of $T(\boldsymbol{X}; \boldsymbol{x}, \boldsymbol{\xi})$ is free of the nuisance parameter $\boldsymbol{\zeta}$.

3. For fixed $\boldsymbol{x}$ and $\boldsymbol{\zeta}$, $Pr(T(\boldsymbol{X}; \boldsymbol{x}, \boldsymbol{\xi}) \geq t \mid \theta)$ is non-decreasing in θ.

Using the function $T(\boldsymbol{X}; \boldsymbol{x}, \boldsymbol{\xi})$, the generalized p-value for the test of H_0 against H_a is defined as

$$p = \sup_{\theta \geq \theta_0} Pr[T(\boldsymbol{X}; \boldsymbol{x}, \boldsymbol{\xi}) \leq t|\theta].$$

Note that Tsui and Weerahandi (1989) define the generalized p-value as

$$p = \sup_{\theta \leq \theta_0} Pr[T(\boldsymbol{X}; \boldsymbol{x}, \boldsymbol{\xi}) \geq t|\theta]$$

which is slightly different from the definition given above. This is because their null hypothesis and alternative hypothesis are $H_0 : \theta \leq \theta_0$ and $H_a : \theta > \theta_0$, respectively, rather than $H_0 : \theta \geq \theta_0$ and $H_a : \theta < \theta_0$, as in our case.

From property (3) above, it follows that

$$p = \sup_{\theta \geq \theta_0} Pr[T(\boldsymbol{X}; \boldsymbol{x}, \boldsymbol{\xi}) \leq t|\theta] = Pr[T(\boldsymbol{X}; \boldsymbol{x}, \boldsymbol{\xi}) \leq t|\theta_0].$$

By properties (1) and (2), the generalized p-value is not dependent on nuisance parameters, and hence can be computed based on data alone. For a nominal α level test of H_0 against H_a one rejects H_0 whenever $p \leq \alpha$.

The above definitions of generalized test variable and generalized p-value are, respectively, analogous to those of test statistic and p-value in classical hypothesis testing.

3.3.1 Summary Statistics

Consider a bioequivalence study with a 4-period, s-sequence, crossover design, where subjects are treated with both the test and reference formulation twice. As before, let Y_{ijkl} denote the response of subject j in sequence i to the l^{th} application of treatment k ($k = T, R$). In the statistical model of Section 2.1, the parameters that are needed to compute the IBE and the PBE criteria are μ_T, μ_R, σ^2_{BT}, σ^2_{BR}, σ^2_{WT}, σ^2_{WR}, and ρ. We now

define the following quantities.

$$\overline{Y}_{ijk} = \frac{1}{2}(Y_{ijk1} + Y_{ijk2}) \qquad E_{ijk} = \frac{1}{\sqrt{2}}(Y_{ijk1} - Y_{ijk2})$$

$$D_{ij} = \overline{Y}_{ijT} - \overline{Y}_{ijR} \qquad \overline{D}_i = \frac{1}{n_i}\sum_j D_{ij}$$

$$D = \frac{1}{s}\sum_i \overline{D}_i \qquad SS_I = \sum_{i=1}^{s}\sum_{j=1}^{n_i}(D_{ij} - \overline{D}_i)^2$$

$$\overline{E}_{i\bullet k} = \frac{1}{n_i}\sum_{j=1}^{n_i} E_{ijk} \qquad \overline{Y}_{i\bullet k} = \frac{1}{n_i}\sum_{j=1}^{n_i} \overline{Y}_{ijk}$$

$$SS_{WT} = \sum_{i=1}^{s}\sum_{j=1}^{n_i}(E_{ijT} - \overline{E}_{i\bullet T})^2 \qquad SS_{WR} = \sum_{i=1}^{s}\sum_{j=1}^{n_i}(E_{ijR} - \overline{E}_{i\bullet R})^2$$

$$SS_T = \sum_{i=1}^{s}\sum_{j=1}^{n_i}(\overline{Y}_{ijT} - \overline{Y}_{i\bullet T})^2 \qquad SS_R = \sum_{i=1}^{s}\sum_{j=1}^{n_i}(\overline{Y}_{ijR} - \overline{Y}_{i\bullet R})^2$$

$$SS_{TR} = \sum_{i=1}^{s}\sum_{j=1}^{n_i}(\overline{Y}_{ijT} - \overline{Y}_{i\bullet T})(\overline{Y}_{ijR} - \overline{Y}_{i\bullet R})$$

$$SS_{R|T} = SS_R - \frac{SS_{TR}^2}{SS_T} \qquad B = \frac{SS_{TR}}{SS_T}$$

$$\delta = \mu_T - \mu_R \qquad \sigma_D^2 = \sigma_{BT}^2 + \sigma_{BR}^2 - 2\rho\sigma_{BT}\sigma_{BR}$$

$$\sigma_T^2 = \sigma_{BT}^2 + \frac{1}{2}\sigma_{WT}^2 \qquad \sigma_R^2 = \sigma_{BR}^2 + \frac{1}{2}\sigma_{WR}^2$$

$$\sigma_{TR} = \rho\sigma_{BT}\sigma_{BR} \qquad \beta = \frac{\sigma_{TR}}{\sigma_T^2}$$

$$\sigma_{R|T}^2 = \sigma_R^2 - \frac{\sigma_{TR}^2}{\sigma_T^2} \qquad c = \sqrt{\frac{1}{s^2}\sum_i \frac{1}{n_i}}$$

$$\sigma_I^2 = \sigma_T^2 + \sigma_R^2 - 2\sigma_{TR} = \sigma_D^2 + \frac{1}{2}(\sigma_{WT}^2 + \sigma_{WR}^2) \qquad \nu = (\Sigma_{i=1}^{s} n_i) - s\ .$$

It is easily seen that the statistics $D, SS_I, SS_{WT}, SS_{WR}, SS_T, SS_{R|T}$, and B are associated with the following pivotal quantities with known distributions as indicated.

$$Z_D = \frac{D - \delta}{c\sigma_I} \sim N(0,1) \qquad U_I = \frac{SS_I}{\sigma_I^2} \sim \chi_\nu^2 \qquad U_{WT} = \frac{SS_{WT}}{\sigma_{WT}^2} \sim \chi_\nu^2$$

$$U_{WR} = \frac{SS_{WR}}{\sigma_{WR}^2} \sim \chi_\nu^2 \qquad U_T = \frac{SS_T}{\sigma_T^2} \sim \chi_\nu^2 \qquad U_{R|T} = \frac{SS_{R|T}}{\sigma_{R|T}^2} \sim \chi_{\nu-1}^2$$

$$Z_B = (B - \beta)\sqrt{\frac{SS_T}{\sigma_{R|T}^2}} \sim N(0,1)$$

The collection of random variables Z_D, U_I, U_{WT}, U_{WR} are mutually independent. Also, the collection of random variables $Z_D, U_{WT}, U_{WR}, U_T, U_{R|T}, Z_B$ are mutually independent.

Note that, the quantities B and $SS_{R|T}$ are, respectively, the linear regression coefficient and the residual sum of squares from the regression of $\overline{Y}_{ijR}$ on $\overline{Y}_{ijT}$.

3.4 Generalized p-Value for the Test of Individual Bioequivalence

The model parameters that appear in the definition of individual bioequivalence are $|\delta|, \sigma_I^2, \sigma_{WT}^2$, and σ_{WR}^2. Equivalently, we can use the parameters $\boldsymbol{\xi} = (\Theta_{IBE}, \sigma_I^2, \sigma_{WT}^2, \sigma_{WR}^2)$ as there is clearly a one to one correspondence between the two parameterizations. Using the latter parameterization we have $\theta = \Theta_{IBE}$ as the parameter of interest and $\boldsymbol{\zeta} = (\sigma_I^2, \sigma_{WT}^2, \sigma_{WR}^2)$ as the vector of nuisance parameters. A generalized test variable for IBE is given by

$$
\begin{aligned}
T_I(\boldsymbol{X}; \boldsymbol{x}, \xi) &= \frac{\Theta_{IBE}\, max\left(\sigma_{W0}^2, \frac{ss_{WR}}{SS_{WR}}\sigma_{WR}^2\right) + 1.5\frac{ss_{WR}}{SS_{WR}}\sigma_{WR}^2}{\left(d - \frac{(D-\delta)}{\sqrt{c^2\sigma_I^2}}\sqrt{c^2\sigma_I^2\frac{ss_I}{SS_I}}\right)^2 + \frac{ss_I}{SS_I}\sigma_I^2 + 0.5\frac{ss_{WT}}{SS_{WT}}\sigma_{WT}^2} \\
&= \frac{\Theta_{IBE}\, max\left(\sigma_{W0}^2, \frac{ss_{WR}}{U_{WR}}\right) + 1.5\frac{ss_{WR}}{U_{WR}}}{\left(d - Z_D\sqrt{c^2\frac{ss_I}{U_I}}\right)^2 + \frac{ss_I}{U_I} + 0.5\frac{ss_{WT}}{U_{WT}}}
\end{aligned}
$$

Here d, ss_I, ss_{WT}, and ss_{WR} are the observed values that correspond to D, SS_I, SS_{WT}, and SS_{WR}, respectively, and are regarded as known constants. Now, the distribution of $T_I(\boldsymbol{X}; \boldsymbol{x}, \xi)$ is free of all nuisance parameters. As a result of the way that the GTV is defined, $T_I(\boldsymbol{x}; \boldsymbol{x}, \xi) = 1$ when D, SS_{WR}, SS_I, and SS_{WT} are set equal to their respective observed values. Lastly, the GTV increases as Θ_{IBE} increases. Therefore, the test function $T_I(\boldsymbol{X}; \boldsymbol{x}, \xi)$ we have defined has the required properties set forth by Tsui and Weerahandi (1989).

Once the GTV is defined, the testing procedure requires computing the generalized p-value $p = Pr(T_I(\boldsymbol{X}; \boldsymbol{x}, \xi) < 1 \mid \Theta_{IBE} = \Theta_I)$. In some cases, *e.g.*, the Behrens-Fisher setup for testing the equality of two means with different variances, the GPV can be evaluated using numerical integration (Tsui and Weerahandi, 1989). In more complex problems, the required probability may be estimated using Monte Carlo methods (Zhou and Mathew, 1994). In this case there is error associated with the point estimate of the

GPV, but this can be minimized by using a large number of simulations. For instance, it is quite feasible to estimate the generalized p-value using 100,000 or even 1,000,000 simulations. Fortunately, doing so is relatively straightforward. We have written code in SAS® that estimates the GPV using the Monte Carlo method with the number of Monte Carlo evaluations equal to a user specified number. Such a procedure estimates the GPV with a standard error less than or equal to 0.2% when the number of Monte Carlo evaluations is 50000 and with a standard error less than or equal to 0.05% when the number of Monte Carlo evaluations is 1000000.

3.4.1 Example 1: Verapamil

Hyslop *et al.* (2000) applied their method to a four sequence/four-period study to assess the bioequivalence of two formulations of Verapamil in 23 healthy male subjects (Chinchilli and Esinhart, 1994) . Using SAS, we computed the following estimates for the area under the plasma concentration/time curve (AUC) for the log-transformed data:

$$\begin{array}{lll} \hat{\mu}_T = 5.6360 & \hat{\mu}_R = 5.6556 & \hat{\delta} = d = -0.01961 \\ \hat{\sigma}^2_{WR} = 0.07450 & \hat{\sigma}^2_{WT} = 0.1271 & \hat{\sigma}^2_I = 0.05924 \end{array}$$

These estimates match those given by Chinchilli and Esinhart (1996). It follows that,

$$ss_{WR} = 1.4155 \quad ss_{WT} = 2.4156 \quad ss_I = 1.1256.$$

Using HSCI. we obtain a 95% upper bound of -0.04006 for the linearized IBE criterion, from which we conclude IBE of the two formulations since the bound is less than 0. The MLS method gives an 95% upper bound of 1.5903, less than $\Theta_I = 2.4948$, which also indicates IBE. The generalized p-value, estimated using 50,000 simulated random variates, is 0.0181 (with 95% confidence interval (0.0169,0.0193)). Here we also conclude IBE, since the GPV is less than 0.05.

3.4.2 Example 2: Methylphenidate

Meyer *et al.* (2000) conducted a study to characterize the intrasubject variability of methylphenidate. In this study 20 healthy male subjects were randomized to one of

four 4-period sequences. Using SAS we computed the following estimates for AUC for the log-transformed data

$$\begin{array}{lll} \hat{\mu}_T = 3.5032 & \hat{\mu}_R = 3.4252 & \hat{\delta} = d = 0.07796 \\ \hat{\sigma}^2_{WT} = 0.05187 & \hat{\sigma}^2_{WR} = 0.03404 & \hat{\sigma}^2_I = 0.02636. \end{array}$$

It follows that,

$$ss_{WR} = 0.54461 \quad ss_{WT} = 0.82991 \quad ss_I = 0.42181.$$

Using the HSCI method we computed a 95% upper confidence bound of -0.0478, matching the value given by Meyer for the constant-scaled criterion. The MLS method gives a 95% upper bound of 1.338. The GPV is $p = 0.0031$ (with 95% confidence interval (0.0026,0.0036)). Based on any of these methods one would conclude IBE for this parameter.

We also computed estimates for maximum plasma concentration (Cmax)

$$\begin{array}{lll} \hat{\mu}_T = 1.9564 & \hat{\mu}_R = 1.8652 & \hat{\delta} = d = 0.09115 \\ \hat{\sigma}^2_{WR} = 0.03057 & \hat{\sigma}^2_{WT} = 0.06702 & \hat{\sigma}^2_I = 0.06922 \\ ss_{WR} = 0.48907 & ss_{WT} = 1.07237 & ss_I = 1.10746. \end{array}$$

Now the HSCI 95% upper confidence bound is 0.0502, again matching the value given by Meyer for the constant-scaled criterion. The MLS upper bound is 5.7381. The GPV is $p = 0.2441$ (with 95% confidence interval (0.2403,0.2479)). Hence, IBE of the formulations is not demonstrated with respect to Cmax.

3.5 Generalized Test Variable for PBE

The model parameters that appear in the definition of population bioequivalence are $|\delta|, \sigma^2_T, \sigma^2_{R|T}$, β, σ^2_{WT} and σ^2_{WR}. Equivalently, we can use the parameters $\boldsymbol{\xi} = (\Theta_{PBE}, \sigma^2_T, \sigma^2_{R|T}, \beta, \sigma^2_{WT}, \sigma^2_{WR})$ as there is clearly a one to one correspondence between the two parameterizations. Using the latter parameterization we have $\theta = \Theta_{PBE}$ as the parameter of interest and $\boldsymbol{\zeta} = (\sigma^2_T, \sigma^2_{R|T}, \beta, \sigma^2_{WT}, \sigma^2_{WR})$ as the vector of nuisance parameters. We first define the following quantities which will be used in the definition of a GTV

for population bioequivalence.

$$
\begin{aligned}
T_1(\boldsymbol{X};\boldsymbol{x},\xi) &= \left(\frac{ss_T\sigma_T^2}{SS_T}\right)\left(b-1-\frac{(B-\beta)}{\sqrt{\frac{\sigma_{R|T}^2}{SS_T}}}\sqrt{\frac{\sigma_{R|T}^2 ss_{R|T}}{ss_T SS_{R|T}}}\right)^2+\frac{ss_{R|T}\sigma_{R|T}^2}{SS_{R|T}} \\
&= \left(\frac{ss_T}{U_1}\right)\left(b-1-Z_B\sqrt{\frac{ss_{R|T}}{ss_T U_{R|T}}}\right)^2+\frac{ss_{R|T}}{U_{R|T}} \\
T_2(\boldsymbol{X};\boldsymbol{x},\xi) &= d-\sqrt{T_1(\boldsymbol{X};\boldsymbol{x},\xi)}\left(\frac{D-\delta}{\sigma_I}\right) \\
&= d-cZ_D\sqrt{T_1(\boldsymbol{X};\boldsymbol{x},\xi)} \\
T_3(\boldsymbol{X};\boldsymbol{x},\xi) &= \frac{ss_{R|T}\sigma_{R|T}^2}{SS_{R|T}}+\left(b-\frac{(B-\beta)}{\sqrt{\frac{\sigma_{R|T}^2}{SS_T}}}\sqrt{\frac{\sigma_{R|T}^2 ss_{R|T}}{ss_T SS_{R|T}}}\right)^2\frac{ss_T\sigma_T^2}{SS_T}+\frac{ss_{WR}\sigma_{WR}^2}{2SS_{WR}} \\
&= \frac{ss_{R|T}}{U_{R|T}}+\left(b-Z_B\sqrt{\frac{ss_{R|T}}{ss_T U_{R|T}}}\right)^2\frac{ss_T}{U_T}+\frac{ss_{WR}}{2U_{WR}}
\end{aligned}
$$

Here $d, ss_T, ss_{R|T}, ss_{WT}, ss_{WR}, b$ are the observed values of $D, SS_T, SS_{R|T}$, SS_{WT}, SS_{WR}, B, respectively, and are regarded as fixed constants. Note that each $T_i(\boldsymbol{X};\boldsymbol{x},\xi)$ is distributed free of nuisance parameters, and that $T_1(\boldsymbol{x};\boldsymbol{x},\xi) = \sigma_T^2(\beta-1)^2+\sigma_{R|T}^2 = \sigma_I^2$, $T_2(\boldsymbol{x};\boldsymbol{x},\xi) = \delta$, and $T_3(\boldsymbol{x};\boldsymbol{x},\xi) = \sigma_{TR}^2$. A generalized test variable for PBE is then given by

$$
\begin{aligned}
T_P(\boldsymbol{X};\boldsymbol{x},\xi) &= \frac{\Theta_{PBE}\,max\left\{\sigma_{T0}^2, T_3(\boldsymbol{X};\boldsymbol{x},\xi)\right\}+T_3(\boldsymbol{X};\boldsymbol{x},\xi)}{[T_2(\boldsymbol{X};\boldsymbol{x},\xi)]^2+\frac{ss_T\sigma_T^2}{SS_T}+\frac{ss_{WT}\sigma_{WT}^2}{2SS_{WT}}} \\
&= \frac{\Theta_{PBE}\,max\left\{\sigma_{T0}^2, T_3(\boldsymbol{X};\boldsymbol{x},\xi)\right\}+T_3(\boldsymbol{X};\boldsymbol{x},\xi)}{[T_2(\boldsymbol{X};\boldsymbol{x},\xi)]^2+\frac{ss_T}{U_T}+\frac{ss_{WT}}{2U_{WT}}}
\end{aligned}
$$

It follows that $T_P(\boldsymbol{x};\boldsymbol{x},\xi) = 1$. Now, the distribution of $T_P(\boldsymbol{X};\boldsymbol{x},\xi)$ is free of all nuisance parameters. Also, the GTV increases as Θ_{PBE} increases. Therefore, the test function $T_P(\boldsymbol{X};\boldsymbol{x},\xi)$ we have defined has the required properties set forth by Tsui and Weerahandi (1989). The generalized p-value for testing

$$H_0 : \Theta_{PBE} \geq \Theta_P \quad \text{versus} \quad H_a : \Theta_{PBE} < \Theta_P$$

is given by $Pr(T_P(\boldsymbol{X};\boldsymbol{x},\xi) < 1 \,|\Theta_{PBE} = \Theta_P)$. This probability can be easily estimated using a Monte Carlo approach.

The two datasets discussed in Section 4 are used again to illustrate and compare the three methods for PBE.

3.5.1 Example 1: Verapamil

For verapamil, using $\sigma_T^2 = \sigma_{BT}^2 + 0.5\sigma_{WT}^2$ and $\sigma_T^2 = \sigma_{BR}^2 + 0.5\sigma_{WR}^2$ to denote the variances of $\overline{Y}_{ijT}$ and $\overline{Y}_{ijR}$, respectively, we have the following additional estimates.

$$\begin{array}{lll} \hat{\sigma}_T^2 = 0.27140 & \hat{\sigma}_R^2 = 0.31633 & \hat{\beta} = b = 0.97363 \\ ss_T = 5.1566 & ss_{R|T} = 1.1220. & \end{array}$$

Using HSCI, we obtain a 95% upper bound of -0.22959 for Θ_{PBE}, from which we conclude PBE of the two formulations since the bound is less than 0. The MLS upper 95% bound is 0.80933, also indicating PBE since it is less than $\Theta_P = 1.7448$. The generalized p-value, estimated using 100,000 simulated random variates, is 0.00001. Here we also conclude PBE, since the GPV is less than 0.05.

3.5.2 Example 2: Methylphenidate

The additional estimates for the methylphenidate AUC data are

$$\begin{array}{lll} \hat{\sigma}_T^2 = 0.12991 & \hat{\sigma}_R^2 = 0.09305 & \hat{\beta} = b = 0.75669 \\ ss_T = 2.07850 & ss_{R|T} = 0.29876. & \end{array}$$

Using the HSCI method we computed a 95% upper confidence bound of 0.02836; the MLS upper bound is 1.93625. Both methods lead to a decision of nonequivalence with respect to PBE for the two formulations. However, the GPV is $p = 0.0038$ (with 95% confidence interval (0.0021,0.0055)). Based on the GPV one would conclude PBE for this parameter.

A similar situation occurs for maximum plasma concentration (Cmax). The additional estimates are

$$\begin{array}{lll} \hat{\sigma}_T^2 = 0.11220 & \hat{\sigma}_R^2 = 0.08850 & \hat{\beta} = b = 0.58594 \\ ss_T = 1.7951 & ss_{R|T} = 0.79969. & \end{array}$$

Now the HSCI 95% upper confidence bound is 0.0245, and for QTWB it is 1.8635, both outside of their respective ranges for equivalence. But the GPV is $p = 0.0250$ (with 95%

confidence interval (0.0207,0.0293)). This is sufficient to conclude equivalence for PBE with respect to Cmax.

The disparity between the GPV method and the other two methods can be explained based on the simulation results which suggest that the GPV method appears to have substantially greater power for a range of parameter value combinations. Moreover, this increased power is attained without significantly exceeding the nominal type-I error rate. When it comes to establishing population bioequivalence, the conservative nature of HSCI and MLS methods results in a loss of power, thus increasing the burden on the manufacturer of a generic drug product.

It is worth noting that, the GPV procedure for PBE takes into account the statistical dependence that exists between SS_T and SS_R whereas the HSCI and the MLS methods appear to have been derived assuming independence between these statistics. This may explain, at least partly, why they have decreased power for testing PBE. In an attempt to further investigate this, we also analyzed the methylphenidate data using a bootstrap approach since it properly accounts for the dependence between SS_T and SS_R. More specifically, we used the nonparametric bias-corrected accelerated bootstrap (Efron and Tibshirani, 1992) to compute 95% upper bounds for both AUC and Cmax. We chose this method over the unadjusted nonparametric bootstrap because of the assertion by Efron and Tibshirani that intervals from the former have higher coverage then those from the latter (pg. 174*ff*). The upper bounds we computed were 1.0965 for AUC and 1.5937 for Cmax, both well below $\Theta_P = 1.7448$. For completeness, we also computed the upper bounds using the ordinary bootstrap percentile method. The bounds were 1.0404 for AUC and 1.4967 for Cmax. Thus, the bootstrap results lead to the same conclusion of PBE that we drew from the GPV analysis. This result is consistent with our thinking that the reduction in power of the HSCI and the MLS methods may be due to not accounting for the dependence structure among the sufficient statistics.

3.6 Simulation Study

A simulation study was conducted to compare the three methods – HSCI, MLS, and GPV – for population bioequivalence as well as for individual bioequivalence. The

methods were examined for their ability to maintain the nominal type-I error rate (false rejection rate) and for their power to reject the null hypothesis of bioinequivalence when the alternative hypothesis of bioequivalence is actually true. Each simulation assumed a two-sequence four-period study design with n subjects in each sequence. Simulation parameters were chosen based on the power tables given in the appendix of the FDA guidance document (2001). For IBE, we used $\sigma_D = 0.01, 0.1, 0.15$, and, for assessing the sizes of the three test methods, chose the other parameter values so that $\Theta_I = 2.4948$ (the boundary of the null hypothesis). Five thousand simulated data sets were generated for each parameter combination. GPVs were estimated based on 10,000 Monte Carlo runs. All tests were done at the nominal $\alpha = 0.05$ level. All simulations were done using SAS® Version 8.1 for Windows. Because the test is conducted using an estimate of the generalized p-value based on 10,000 Monte Carlo evaluations rather than an exact calculation, strictly speaking, the simulation results are applicable only when the generalized p-value is estimated using 10,000 Monte Carlo runs. In any real application the number of Monte Carlo evaluations is likely to be much larger than 10000. We conducted a limited number of additional simulations the results of which indicate that our general conclusions regarding relative performances of the three methods do not change when more than 10000 Monte Carlo evaluations are used to estimate the generalized p-value. Our simulation results are summarized below, first for IBE, then for PBE. Since the empirical size estimates are based on 5000 simulated data sets, the standard errors of the estimates are all approximately equal to 0.34% (corresponding to an approximately 5% false rejection rate). For empirical power estimates, the maximum standard error, which would correspond to an empirical power of 50%, is calculated to be 0.8%. These estimation errors should be kept in mind when making comparisons between the reported empirical power estimates of the different methods.

3.6.1 Individual Bioequivalence

The examination of the size of the three procedures for IBE was done at a number of points on the boundary surface of the null hypothesis, i.e., where $\Theta_I = 2.4948$ (the value

proposed by the FDA). We also used the FDA-specified value $\sigma^2_{W0} = 0.04$ for constant-scaling. The estimated sizes of the procedures are summarized in Tables 2a-c.

Table 2a. Estimated sizes of the test procedures HSCI, MLS, and GPV for testing IBE using 2 sequence/4 period designs

	$\sigma_D = 0.01$								
	$\delta = 0.3157$						$\delta = 0.3631$		
	$\sigma_{WT} = \sigma_{WR} = 0.15$			$\sigma_{WT} = \sigma_{WR} = 0.2$			$\sigma_{WT} = \sigma_{WR} = 0.23$		
n	HSCI	MLS	GPV	HSCI	MLS	GPV	HSCI	MLS	GPV
6	0.0412	0.0378	0.0426	0.0502	0.0486	0.0570	0.0422	0.0426	0.0374
12	0.0452	0.0376	0.0444	0.0522	0.0490	0.0660	0.0446	0.0430	0.0414
18	0.0500	0.0420	0.0480	0.0562	0.0516	0.0692	0.0500	0.0458	0.0446
24	0.0502	0.0438	0.0458	0.0554	0.0518	0.0692	0.0482	0.0454	0.0442
	$\delta = 0.4737$			$\delta = 0.7834$					
	$\sigma_{WT} = \sigma_{WR} = 0.3$			$\sigma_{WT} = \sigma_{WR} = 0.5$					
n	HSCI	MLS	GPV	HSCI	MLS	GPV			
6	0.0414	0.0424	0.0336	0.0426	0.0436	0.0344			
12	0.0490	0.0466	0.0430	0.0468	0.0448	0.0398			
18	0.0450	0.0422	0.0394	0.0480	0.0460	0.0422			
24	0.0482	0.0466	0.0448	0.0498	0.0462	0.0446			

	$\sigma_D = 0.10$								
	$\delta = 0.2997$						$\delta = 0.3493$		
	$\sigma_{WT} = \sigma_{WR} = 0.15$			$\sigma_{WT} = \sigma_{WR} = 0.2$			$\sigma_{WT} = \sigma_{WR} = 0.23$		
n	HSCI	MLS	GPV	HSCI	MLS	GPV	HSCI	MLS	GPV
6	0.0390	0.0318	0.0374	0.0510	0.0488	0.0570	0.0438	0.0448	0.0388
12	0.0478	0.0374	0.0456	0.0578	0.0530	0.0686	0.0480	0.0454	0.0436
18	0.0524	0.0412	0.0486	0.0570	0.0512	0.0718	0.0574	0.0544	0.0536
24	0.0568	0.0450	0.0526	0.0596	0.0538	0.0724	0.0444	0.0408	0.0406
	$\delta = 0.4737$			$\delta = 0.7834$					
	$\sigma_{WT} = \sigma_{WR} = 0.3$			$\sigma_{WT} = \sigma_{WR} = 0.5$					
n	HSCI	MLS	GPV	HSCI	MLS	GPV			
6	0.0496	0.0504	0.0402	0.0434	0.0452	0.0356			
12	0.0480	0.0462	0.0414	0.0486	0.0448	0.0406			
18	0.0492	0.0474	0.0448	0.0492	0.0462	0.0432			
24	0.0534	0.0502	0.0484	0.0488	0.0452	0.0428			

	$\sigma_D = 0.15$								
	$\delta = 0.2780$						$\delta = 0.3309$		
	$\sigma_{WT} = \sigma_{WR} = 0.15$			$\sigma_{WT} = \sigma_{WR} = 0.2$			$\sigma_{WT} = \sigma_{WR} = 0.23$		
n	HSCI	MLS	GPV	HSCI	MLS	GPV	HSCI	MLS	GPV
6	0.0458	0.0366	0.0468	0.0494	0.0452	0.0534	0.0424	0.0414	0.0370
12	0.0498	0.0346	0.0464	0.0504	0.0436	0.0586	0.0514	0.0472	0.0468
18	0.0544	0.0386	0.0498	0.0552	0.0478	0.0676	0.0512	0.0470	0.0492
24	0.0520	0.0400	0.0472	0.0504	0.0440	0.0662	0.0494	0.0448	0.0458
	$\delta = 0.4495$			$\delta = 0.7754$					
	$\sigma_{WT} = \sigma_{WR} = 0.3$			$\sigma_{WT} = \sigma_{WR} = 0.5$					
n	HSCI	MLS	GPV	HSCI	MLS	GPV			
6	0.0404	0.0416	0.0340	0.0398	0.0406	0.0316			
12	0.0486	0.0460	0.0430	0.0450	0.0418	0.0400			
18	0.0546	0.0510	0.0500	0.0516	0.0486	0.0460			
24	0.0464	0.0434	0.0420	0.0476	0.0448	0.0436			

Table 2b. Estimated sizes of the test procedures HSCI, MLS, and GPV for testing IBE using 2 sequence/4 period designs

	$\delta = 0.05$								
	$\sigma_D = 0.01$								
	$\sigma_{WT} = 0.3460$ $\sigma_{WR} = 0.15$			$\sigma_{WT} = 0.3704$ $\sigma_{WR} = 0.2$			$\sigma_{WT} = 0.4269$ $\sigma_{WR} = 0.23$		
n	HSCI	MLS	GPV	HSCI	MLS	GPV	HSCI	MLS	GPV
6	0.0150	0.0150	0.0130	0.0292	0.0386	0.0290	0.0218	0.0308	0.0178
12	0.0286	0.0248	0.0268	0.0434	0.0448	0.0432	0.0326	0.0366	0.0270
18	0.0276	0.0250	0.0244	0.0432	0.0430	0.0466	0.0328	0.0328	0.0268
24	0.0318	0.0254	0.0276	0.0468	0.0450	0.0506	0.0410	0.0422	0.0358
	$\sigma_{WT} = 0.5585$			$\sigma_{WT} = 0.9333$					
	$\sigma_{WR} = 0.3$			$\sigma_{WR} = 0.5$					
n	HSCI	MLS	GPV	HSCI	MLS	GPV			
6	0.0304	0.0382	0.0234	0.0226	0.0298	0.0160			
12	0.0372	0.0424	0.0292	0.0328	0.0370	0.0268			
18	0.0368	0.0394	0.0300	0.0406	0.0424	0.0326			
24	0.0398	0.0404	0.0352	0.0368	0.0394	0.0296			

	$\sigma_D = 0.10$								
	$\sigma_{WT} = 0.3313$ $\sigma_{WR} = 0.15$			$\sigma_{WT} = 0.3568$ $\sigma_{WR} = 0.2$			$\sigma_{WT} = 0.4152$ $\sigma_{WR} = 0.23$		
n	HSCI	MLS	GPV	HSCI	MLS	GPV	HSCI	MLS	GPV
6	0.0158	0.0180	0.0136	0.0248	0.0338	0.0248	0.0262	0.0364	0.0210
12	0.0246	0.0228	0.0230	0.0346	0.0384	0.0372	0.0312	0.0356	0.0266
18	0.0324	0.0278	0.0294	0.0366	0.0354	0.0426	0.0374	0.0374	0.0320
24	0.0348	0.0290	0.0316	0.0468	0.0466	0.0538	0.0342	0.0350	0.0296
	$\sigma_{WT} = 0.5496$			$\sigma_{WT} = 0.9280$					
	$\sigma_{WR} = 0.3$			$\sigma_{WR} = 0.5$					
n	HSCI	MLS	GPV	HSCI	MLS	GPV			
6	0.0260	0.0334	0.0186	0.0246	0.0344	0.0172			
12	0.0294	0.0358	0.0230	0.0340	0.0368	0.0252			
18	0.0362	0.0382	0.0302	0.0316	0.0338	0.0238			
24	0.0344	0.0356	0.0280	0.0378	0.0406	0.0318			

	$\sigma_D = 0.15$								
	$\sigma_{WT} = 0.3119$ $\sigma_{WR} = 0.15$			$\sigma_{WT} = 0.3388$ $\sigma_{WR} = 0.2$			$\sigma_{WT} = 0.3999$ $\sigma_{WR} = 0.23$		
n	HSCI	MLS	GPV	HSCI	MLS	GPV	HSCI	MLS	GPV
6	0.0166	0.0190	0.0144	0.0264	0.0350	0.0256	0.0258	0.0332	0.0228
12	0.0270	0.0232	0.0226	0.0408	0.0426	0.0448	0.0302	0.0330	0.0240
18	0.0298	0.0226	0.0262	0.0432	0.0438	0.0496	0.0392	0.0392	0.0326
24	0.0316	0.0276	0.0278	0.0412	0.0416	0.0470	0.0366	0.0364	0.0320
	$\sigma_{WT} = 0.5381$			$\sigma_{WT} = 0.9213$					
	$\sigma_{WR} = 0.3$			$\sigma_{WR} = 0.5$					
n	HSCI	MLS	GPV	HSCI	MLS	GPV			
6	0.0250	0.0336	0.0202	0.0262	0.0362	0.0210			
12	0.0342	0.0372	0.0268	0.0342	0.0388	0.0276			
18	0.0310	0.0344	0.0248	0.0356	0.0364	0.0284			
24	0.0362	0.0366	0.0306	0.0384	0.0398	0.0306			

Table 2c. Estimated sizes of the test procedures HSCI, MLS, and GPV for testing IBE using 2 sequence/4 period designs

	$\delta = 0.15$								
	$\sigma_D = 0.01$								
	$\sigma_{WT} = 0.3157$ $\sigma_{WR} = 0.15$			$\sigma_{WT} = 0.3423$ $\sigma_{WR} = 0.2$			$\sigma_{WT} = 0.4269$ $\sigma_{WR} = 0.23$		
n	HSCI	MLS	GPV	HSCI	MLS	GPV	HSCI	MLS	GPV
6	0.0180	0.0136	0.0170	0.0286	0.0316	0.0304	0.0264	0.0294	0.0230
12	0.0278	0.0184	0.0244	0.0446	0.0392	0.0476	0.0386	0.0366	0.0340
18	0.0334	0.0204	0.0310	0.0432	0.0368	0.0524	0.0376	0.0340	0.0324
24	0.0336	0.0234	0.0310	0.0478	0.0416	0.0580	0.0428	0.0404	0.0390
	$\sigma_{WT} = 0.5403$			$\sigma_{WT} = 0.9225$					
	$\sigma_{WR} = 0.3$			$\sigma_{WR} = 0.5$					
n	HSCI	MLS	GPV	HSCI	MLS	GPV			
6	0.0248	0.0284	0.0172	0.0252	0.0342	0.0210			
12	0.0322	0.0318	0.0256	0.0314	0.0340	0.0232			
18	0.0386	0.0354	0.0324	0.0362	0.0360	0.0292			
24	0.0412	0.0352	0.0322	0.0390	0.0364	0.0320			

	$\sigma_D = 0.10$								
	$\sigma_{WT} = 0.2997$ $\sigma_{WR} = 0.15$			$\sigma_{WT} = 0.3276$ $\sigma_{WR} = 0.2$			$\sigma_{WT} = 0.3903$ $\sigma_{WR} = 0.23$		
n	HSCI	MLS	GPV	HSCI	MLS	GPV	HSCI	MLS	GPV
6	0.0168	0.0136	0.0144	0.0326	0.0380	0.0338	0.0242	0.0280	0.0204
12	0.0328	0.0198	0.0306	0.0372	0.0312	0.0410	0.0366	0.0322	0.0318
18	0.0356	0.0222	0.0326	0.0396	0.0334	0.0460	0.0370	0.0318	0.0312
24	0.0302	0.0194	0.0266	0.0460	0.0382	0.0544	0.0444	0.0402	0.0400
	$\sigma_{WT} = 0.5311$			$\sigma_{WT} = 0.9172$					
	$\sigma_{WR} = 0.3$			$\sigma_{WR} = 0.5$					
n	HSCI	MLS	GPV	HSCI	MLS	GPV			
6	0.0280	0.0326	0.0208	0.0226	0.0328	0.0164			
12	0.0376	0.0356	0.0296	0.0356	0.0356	0.0278			
18	0.0374	0.0342	0.0308	0.0392	0.0376	0.0318			
24	0.0408	0.0358	0.0334	0.0348	0.0344	0.0286			

	$\sigma_D = 0.15$								
	$\sigma_{WT} = 0.2780$ $\sigma_{WR} = 0.15$			$\sigma_{WT} = 0.3079$ $\sigma_{WR} = 0.2$			$\sigma_{WT} = 0.3740$ $\sigma_{WR} = 0.23$		
n	HSCI	MLS	GPV	HSCI	MLS	GPV	HSCI	MLS	GPV
6	0.0234	0.0190	0.0224	0.0356	0.0370	0.0362	0.0280	0.0320	0.0236
12	0.0276	0.0160	0.0256	0.0404	0.0330	0.0442	0.0370	0.0328	0.0332
18	0.0368	0.0204	0.0326	0.0484	0.0372	0.0560	0.0386	0.0356	0.0352
24	0.0316	0.0160	0.0284	0.0470	0.0370	0.0542	0.0450	0.0376	0.0376
	$\sigma_{WT} = 0.5192$			$\sigma_{WT} = 0.9103$					
	$\sigma_{WR} = 0.3$			$\sigma_{WR} = 0.5$					
n	HSCI	MLS	GPV	HSCI	MLS	GPV			
6	0.0238	0.0300	0.0168	0.0212	0.0298	0.0156			
12	0.0328	0.0320	0.0274	0.0372	0.0382	0.0300			
18	0.0334	0.0304	0.0276	0.0342	0.0330	0.0286			
24	0.0440	0.0396	0.0376	0.0396	0.0382	0.0340			

In all, 180 different parameter combinations under H_0 were examined and empirical false rejection rates computed based on 5000 data sets. An empirical false rejection rate exceeding 0.055 would indicate (at the 95% confidence level) that the actual false rejection rate exceeds the nominal value of 0.05. This rate for the MLS method was below the threshold value of 0.055 in all cases examined whereas the HSCI method exceeded the threshold in 8 of the 180 cases and the GPV method in 13 of the 180 cases. Interestingly, however, in 145 of the 180 cases examined, the false rejection rate for HSCI was greater than or equal to that for GPV. Further details about these false rejection rates are provided in the discussion section.

Power of the three procedures was examined for $\delta = 0.05, 0.1$, $\sigma_D = 0.01, 0.1, 0.15$, and $\sigma_{WT} = \sigma_{WR} = 0.15, 0.2, 0.23, 0.3, 0.5$. We have summarized these results by reporting the estimated minimum sample size per group needed to achieve 80% power, and the estimate of the actual power at the given sample size for the 3 methods. These results are summarized in Table 3.

Table 3. Estimated power for testing IBE of the test procedures HSCI, MLS, and GPV, for 2 sequence/4 period designs

		$\delta = 0.05$									
		$\sigma_{WT} = 0.15$ $\sigma_{WR} = 0.15$		$\sigma_{WT} = 0.2$ $\sigma_{WR} = 0.2$		$\sigma_{WT} = 0.23$ $\sigma_{WR} = 0.23$		$\sigma_{WT} = 0.3$ $\sigma_{WR} = 0.3$		$\sigma_{WT} = 0.5$ $\sigma_{WR} = 0.5$	
Method	σ_D	n	Power	n	Power	n	Power	n	Power	n	Power
HSCI	0.01	5	0.866	8	0.811	11	0.820	13	0.806	13	0.821
	0.10	7	0.834	11	0.810	15	0.820	15	0.802	14	0.821
	0.15	10	0.803	17	0.815	21	0.811	18	0.811	15	0.820
MLS	0.01	5	0.820	9	0.812	12	0.804	14	0.815	13	0.813
	0.10	8	0.831	13	0.820	16	0.811	16	0.815	13	0.800
	0.15	13	0.824	20	0.811	23	0.804	19	0.818	15	0.817
GPV	0.01	5	0.834	8	0.822	11	0.849	14	0.810	14	0.818
	0.10	7	0.807	11	0.821	14	0.819	16	0.811	15	0.811
	0.15	11	0.816	17	0.827	20	0.818	19	0.813	15	0.800

		$\delta = 0.1$									
		$\sigma_{WT} = 0.15$ $\sigma_{WR} = 0.15$		$\sigma_{WT} = 0.2$ $\sigma_{WR} = 0.2$		$\sigma_{WT} = 0.23$ $\sigma_{WR} = 0.23$		$\sigma_{WT} = 0.3$ $\sigma_{WR} = 0.3$		$\sigma_{WT} = 0.5$ $\sigma_{WR} = 0.5$	
Method	σ_D	n	Power	n	Power	n	Power	n	Power	n	Power
HSCI	0.01	6	0.896	10	0.838	13	0.810	15	0.838	14	0.828
	0.10	8	0.836	14	0.836	17	0.806	17	0.810	14	0.801
	0.15	13	0.814	22	0.826	26	0.812	20	0.800	15	0.812
MLS	0.01	6	0.802	12	0.825	15	0.814	15	0.810	14	0.809
	0.10	10	0.823	16	0.805	20	0.805	18	0.813	15	0.815
	0.15	17	0.833	25	0.810	29	0.815	22	0.815	16	0.807
GPV	0.01	6	0.878	10	0.855	12	0.801	15	0.821	14	0.800
	0.10	8	0.813	13	0.814	16	0.813	18	0.825	15	0.806
	0.15	14	0.841	21	0.815	24	0.809	21	0.815	16	0.807

From the power comparisons it appears that HSCI and GPV are very similar with respect to power and sample size requirements, with GPV doing somewhat better for values of σ_{WR} close to $\sigma_{W0} = 0.2$ (however, GPV is also slightly anti-conservative in this case), and HSCI doing slightly better when the σ_{WR} is relatively small or large. The MLS method appeared to have somewhat lower power when compared to the other two methods. This may be, at least in part, due to our previous observation that the MLS method tends to have lower false rejection rates compared to the other two methods.

3.6.2 Population Bioequivalence

Tables 4a and 4b show simulation results comparing the respective sizes of the 3 tests. All of the parameter combinations in these tables correspond to $\Theta_P = 1.7448$, with σ^2_{T0}

taken to be 0.04. Both values are as recommended in the FDA guidance document. Here, correlations of $\rho = 0.8$ and $\rho = 0.9$ are examined. The reasoning for examining cases with such seemingly high correlation is that, the responses within a subject to different formulations of the same drug are still likely to have a strong degree of correlation even if the formulations are only borderline bioequivalent.

Table 4a. Estimated sizes of the test procedures HSCI, MLS, and GPV for testing PBE using 2 sequence/4 period designs ($\delta \neq 0$)

		$\delta = 0.2642$			$\delta = 0.2954$			$\delta = 0.5907$		
		$\sigma_{BT} = 0.1, \sigma_{BR} = 0.1$			$\sigma_{BT} = 0.2, \sigma_{BR} = 0.2$			$\sigma_{BT} = 0.4, \sigma_{BR} = 0.4$		
		$\sigma_{WT} = 0.1, \sigma_{WR} = 0.1$			$\sigma_{WT} = 0.1, \sigma_{WR} = 0.1$			$\sigma_{WT} = 0.2, \sigma_{WR} = 0.2$		
n	ρ	HSCI	MLS	GPV	HSCI	MLS	GPV	HSCI	MLS	GPV
12	0.8	0.0400	0.0400	0.0536	0.0202	0.0212	0.0694	0.0186	0.0214	0.0656
	0.9	0.0348	0.0346	0.0480	0.0108	0.0134	0.0630	0.0102	0.0110	0.0596
18	0.8	0.0412	0.0392	0.0540	0.0168	0.0170	0.0562	0.0200	0.0196	0.0582
	0.9	0.0384	0.0380	0.0550	0.0138	0.0152	0.0642	0.0142	0.0144	0.0548
24	0.8	0.0390	0.0360	0.0502	0.0194	0.0196	0.0568	0.0198	0.0200	0.0606
	0.9	0.0436	0.0418	0.0552	0.0132	0.0134	0.0594	0.0168	0.0168	0.0584

		$\delta = 0.5604$			$\delta = 0.8861$		
		$\sigma_{BT} = 0.3, \sigma_{BR} = 0.3$			$\sigma_{BT} = 0.6, \sigma_{BR} = 0.6$		
		$\sigma_{WT} = 0.3, \sigma_{WR} = 0.3$			$\sigma_{WT} = 0.3, \sigma_{WR} = 0.3$		
n	ρ	HSCI	MLS	GPV	HSCI	MLS	GPV
12	0.8	0.0302	0.0294	0.0636	0.0186	0.0194	0.0652
	0.9	0.0294	0.0276	0.0626	0.0110	0.0120	0.0546
18	0.8	0.0348	0.0320	0.0624	0.0214	0.0218	0.0582
	0.9	0.0298	0.0272	0.0614	0.0162	0.0168	0.0592
24	0.8	0.0400	0.0370	0.0636	0.0234	0.0226	0.0568
	0.9	0.0354	0.0322	0.0572	0.0158	0.0158	0.0552

		$\delta = 0.2954$			$\delta = 0.5907$					
		$\sigma_{BT} = 0.1732, \sigma_{BR} = 0.2$			$\sigma_{BT} = 0.3873. \sigma_{BR} = 0.4$			$\sigma_{BT} = 0.4, \sigma_{BR} = 0.3873$		
		$\sigma_{WT} = 0.1414, \sigma_{WR} = 0.1$			$\sigma_{WT} = 0.2236, \sigma_{WR} = 0.2$			$\sigma_{WT} = 0.2, \sigma_{WR} = 0.2236$		
n	ρ	HSCI	MLS	GPV	HSCI	MLS	GPV	HSCI	MLS	GPV
12	0.8	0.0260	0.0258	0.0644	0.0166	0.0184	0.0550	0.0230	0.0240	0.0684
	0.9	0.0168	0.0192	0.0570	0.0130	0.0144	0.0582	0.0114	0.0132	0.0596
18	0.8	0.0252	0.0250	0.0594	0.0226	0.0230	0.0658	0.0226	0.0232	0.0598
	0.9	0.0256	0.0260	0.0690	0.0108	0.0122	0.0522	0.0146	0.0152	0.0562
24	0.8	0.0256	0.0248	0.0600	0.0232	0.0228	0.0548	0.0264	0.0254	0.0642
	0.9	0.0244	0.0244	0.0570	0.0172	0.0168	0.0568	0.0156	0.0160	0.0624

		$\delta = 0.8861$					
		$\sigma_{BT} = 0.6, \sigma_{BR} = 0.5$			$\sigma_{BT} = 0.5, \sigma_{BR} = 0.6$		
		$\sigma_{WT} = 0.3, \sigma_{WR} = 0.4472$			$\sigma_{WT} = 0.4472, \sigma_{WR} = 0.3$		
n	ρ	HSCI	MLS	GPV	HSCI	MLS	GPV
12	0.8	0.0258	0.0248	0.0690	0.0254	0.0260	0.0642
	0.9	0.0200	0.0206	0.0606	0.0196	0.0212	0.0552
18	0.8	0.0280	0.0266	0.0634	0.0284	0.0274	0.0596
	0.9	0.0242	0.0232	0.0644	0.0258	0.0254	0.0612
24	0.8	0.0286	0.0280	0.0606	0.0300	0.0288	0.0544
	0.9	0.0228	0.0208	0.0618	0.0252	0.0250	0.0540

Table 4b. Estimated sizes of the test procedures HSCI, MLS, and GPV for testing PBE using 2 sequence/4 period design ($\delta = 0$)

		$\sigma_{BT} = 0.2825, \sigma_{BR} = 0.1$ $\sigma_{WT} = 0.1, \sigma_{WR} = 0.1$			$\sigma_{BT} = 0.3567, \sigma_{BR} = 0.2$ $\sigma_{WT} = 0.1, \sigma_{WR} = 0.1$			$\sigma_{BT} = 0.4238, \sigma_{BR} = 0.2$ $\sigma_{WT} = 0.2, \sigma_{WR} = 0.2$		
n	ρ	HSCI	MLS	GPV	HSCI	MLS	GPV	HSCI	MLS	GPV
12	0.8	0.029	0.044	0.039	0.013	0.017	0.063	0.017	0.023	0.054
	0.9	0.029	0.042	0.040	0.006	0.008	0.068	0.013	0.016	0.051
18	0.8	0.034	0.047	0.042	0.010	0.012	0.066	0.019	0.023	0.055
	0.9	0.030	0.039	0.040	0.005	0.008	0.067	0.010	0.014	0.047
24	0.8	0.029	0.038	0.037	0.010	0.012	0.060	0.016	0.022	0.050
	0.9	0.034	0.040	0.041	0.003	0.004	0.060	0.016	0.018	0.055
		$\sigma_{BT} = 0.7134, \sigma_{BR} = 0.4$ $\sigma_{WT} = 0.2, \sigma_{WR} = 0.2$			$\sigma_{BT} = 0.6357, \sigma_{BR} = 0.3$ $\sigma_{WT} = 0.3, \sigma_{WR} = 0.3$			$\sigma_{BT} = 1.0701, \sigma_{BR} = 0.6$ $\sigma_{WT} = 0.3, \sigma_{WR} = 0.3$		
n	ρ	HSCI	MLS	GPV	HSCI	MLS	GPV	HSCI	MLS	GPV
12	0.8	0.006	0.009	0.055	0.016	0.023	0.053	0.006	0.007	0.051
	0.9	0.002	0.003	0.051	0.013	0.018	0.048	0.002	0.003	0.054
18	0.8	0.007	0.009	0.053	0.017	0.022	0.058	0.008	0.010	0.055
	0.9	0.002	0.003	0.049	0.012	0.016	0.055	0.002	0.002	0.050
24	0.8	0.005	0.006	0.051	0.019	0.021	0.054	0.010	0.012	0.052
	0.9	0.002	0.003	0.048	0.013	0.016	0.052	0.003	0.004	0.055

In most cases, both HSCI and MLS are far more conservative than GPV. The GPV method shows a tendency to be somewhat anti-conservative, but none of observed test-levels exceeded 0.07. In any case, both HSCI and MLS appear to exhibit a level of conservatism far beyond what would seem to be reasonable. When a test with a nominal size of $\alpha = 0.05$ exhibits actual sizes below 0.005, its powers is apt to be affected. As shown below, this appears to be the case with HSCI and MLS.

Table 5 summarizes the simulation results for the power of the three tests. It shows the smallest n for which a particular test achieves 80% power and the actual estimate of the power for that test at that point. Here correlations of $\rho = 0.9$ and $\rho = 0.95$ are examined. These values were deemed reasonable for examining power because formulations that are bioequivalent should exhibit a very high degree of correlation.

Table 5. Sample size comparisons and estimated power for PBE of the test procedures HSCI, MLS, and GPV for 2 sequence/4 period designs

		$\delta = 0.05$							
		$\sigma_{WT} = 0.15, \sigma_{WR} = 0.15$				$\sigma_{WT} = 0.23, \sigma_{WR} = 0.23$			
		$\sigma_{BT} = 0.15$ $\sigma_{BR} = 0.15$		$\sigma_{BT} = 0.3$ $\sigma_{BR} = 0.3$		$\sigma_{BT} = 0.23$ $\sigma_{BR} = 0.23$		$\sigma_{BT} = 0.46$ $\sigma_{BR} = 0.46$	
Method	ρ	n	Power	n	Power	n	Power	n	Power
HSCI	0.90	8	0.811	10	0.820	10	0.831	10	0.833
	0.95	8	0.823	10	0.851	10	0.839	10	0.863
MLS	0.90	7	0.831	9	0.804	9	0.842	9	0.811
	0.95	7	0.834	9	0.835	9	0.840	9	0.841
GPV	0.90	6	0.835	6	0.854	7	0.813	6	0.860
	0.95	6	0.861	5	0.847	7	0.824	5	0.841
		$\sigma_{WT} = 0.3, \sigma_{WR} = 0.3$				$\sigma_{WT} = 0.5, \sigma_{WR} = 0.5$			
		$\sigma_{BT} = 0.3$ $\sigma_{BR} = 0.3$		$\sigma_{BT} = 0.6$ $\sigma_{BR} = 0.6$		$\sigma_{BT} = 0.5$ $\sigma_{BR} = 0.5$		$\sigma_{BT} = 1.0$ $\sigma_{BR} = 1.0$	
Method	ρ	n	Power	n	Power	n	Power	n	Power
HSCI	0.90	10	0.837	10	0.834	10	0.846	10	0.832
	0.95	10	0.856	10	0.864	9	0.810	10	0.862
MLS	0.90	9	0.843	9	0.810	9	0.843	9	0.818
	0.95	9	0.848	9	0.842	8	0.801	9	0.838
GPV	0.90	7	0.819	6	0.852	7	0.821	6	0.861
	0.95	7	0.840	5	0.839	7	0.830	5	0.842
		$\delta = 0.1$							
		$\sigma_{WT} = 0.15, \sigma_{WR} = 0.15$				$\sigma_{WT} = 0.23, \sigma_{WR} = 0.23$			
		$\sigma_{BT} = 0.15$ $\sigma_{BR} = 0.15$		$\sigma_{BT} = 0.3$ $\sigma_{BR} = 0.3$		$\sigma_{BT} = 0.23$ $\sigma_{BR} = 0.23$		$\sigma_{BT} = 0.46$ $\sigma_{BR} = 0.46$	
Method	ρ	n	Power	n	Power	n	Power	n	Power
HSCI	0.90	10	0.821	11	0.836	11	0.826	10	0.812
	0.95	10	0.831	10	0.804	11	0.850	10	0.833
MLS	0.90	9	0.829	10	0.826	10	0.827	10	0.854
	0.95	8	0.805	10	0.851	9	0.803	9	0.840
GPV	0.90	7	0.833	6	0.831	8	0.822	6	0.845
	0.95	7	0.854	6	0.872	8	0.825	5	0.805
		$\sigma_{WT} = 0.3, \sigma_{WR} = 0.3$				$\sigma_{WT} = 0.5, \sigma_{WR} = 0.5$			
		$\sigma_{BT} = 0.3$ $\sigma_{BR} = 0.3$		$\sigma_{BT} = 0.6$ $\sigma_{BR} = 0.6$		$\sigma_{BT} = 0.5$ $\sigma_{BR} = 0.5$		$\sigma_{BT} = 1.0$ $\sigma_{BR} = 1.0$	
Method	ρ	n	Power	n	Power	n	Power	n	Power
HSCI	0.90	10	0.821	10	0.823	10	0.833	10	0.836
	0.95	10	0.822	10	0.857	10	0.842	10	0.857
MLS	0.90	9	0.825	9	0.811	9	0.835	9	0.810
	0.95	9	0.818	9	0.831	9	0.838	9	0.842
GPV	0.90	8	0.845	6	0.841	7	0.811	6	0.855
	0.95	7	0.821	5	0.833	7	0.845	5	0.833

Table 5 shows how that GPV has smaller sample size requirements than the other 2 methods across the board. The principal reason for this appears to be that the proposed GPV method is the only one that accounts for the intra-subject dependence between $\overline{Y}_{ijT}$ and $\overline{Y}_{ijR}$. As one might expect, the difference in power between GPV and the other

methods becomes more pronounced as the correlation increases. GPV also apparently has more power when the between-subject variances are greater than the within-subject variances. Both these aspects are further discussed below. It is worth noting that the sample size required to yield 80% power for HSCI always results in at least 90% power for GPV, and, in most cases, the power for GPV is in excess of 95%.

3.7 Discussion

The GPV methods proposed in this investigation appear to perform reasonably well compared to the FDA-recommended HSCI approach and the MLS methods. For IBE, the HSCI and GPV methods appear to have very similar performance characteristics, with HSCI being slightly more powerful for what seems to be a substantial portion of the parameter space. However, the HSCI and MLS bounds are discontinuous with respect to σ_{WR} at the point of changeover (*i.e.*, σ^2_{W0}) from constant- to reference-scaled criterion. On the other hand, the GPV test can be shown to be continuous at the changeover point, which is certainly a desirable feature of any decision rule. However, the GPV test is also slightly anti-conservative in this region.

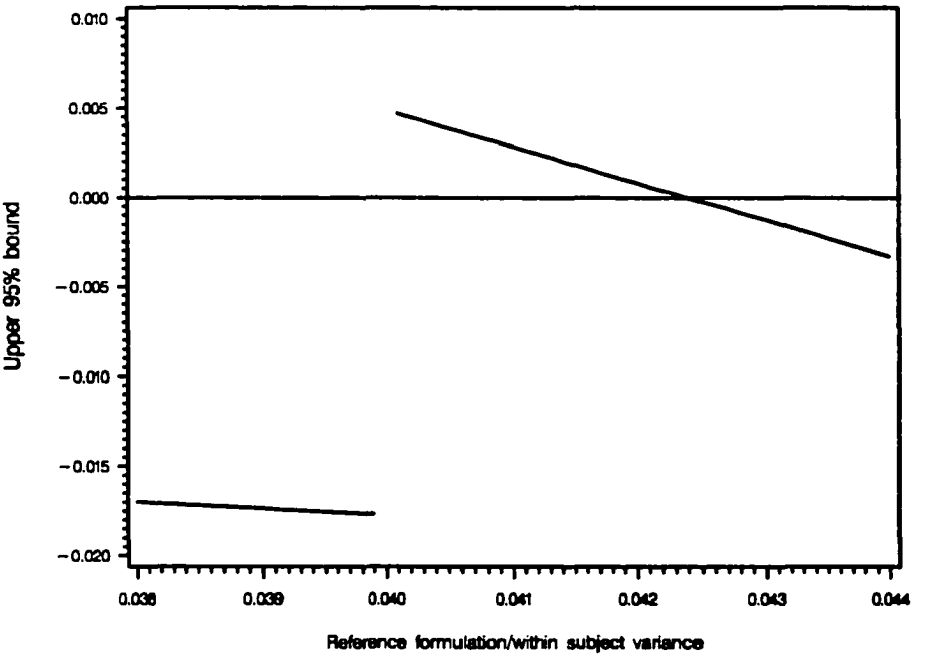
For PBE, the GPV method appears to be clearly more powerful than HSCI, although slightly anti-conservative. The impact of this is likely to be attenuated by the regulatory requirement that the test formulation be average BE with respect to the reference formulation, since the most anti-conservative cases are those where $|\delta|$ (the magnitude of difference between the means) is greater than $log(1.25)$. GPV makes up for its slight anti-conservatism by having substantially increased power compared to the other two tests.

The correlation of responses within subjects is key to the practice of assessing BE. When two formulations of the same drug are truly BE, we expect that the responses of a subject to the two formulations will be very highly correlated. In the example crossover studies we cited earlier, the method of moments estimates $\hat{\rho}$ were greater than 1 for the two examples that clearly showed BE (verapamil and methylphenidate/AUC), and $\hat{\rho}$ was high even in the case where BE was not demonstrated ($\hat{\rho} = 0.87$ for methylphenidate/Cmax).

This appears to justify the high correlations we used in our simulation studies. The seeming anomaly of $\hat{\rho} > 1$ is the fallout of using the method of moments to estimate variance components, as opposed to some other method that restricts the estimates to the parameter space.

The other supposition of the simulation studies – that the between-formulation variances are as large or larger than the within-formulation variances – is also a tenet of crossover designs in particular, and split-plot designs in general, *i.e.*, that the between-plot variances are usually greater than the within-plot variances. For example, Schumaker and Metzler (1998) observed that between-formulation variances are often much larger than within-formulation variances. In the IBE problem, the correlation is reflected in σ_D^2 - high values of ρ correspond to smaller subject-formulation interactions. The influence of correlation is not overtly apparent in the PBE problem, but it appears to have an impact on the variation of $\hat{\sigma}_R^2$ (remember $\sigma_R^2 = \sigma_{BR}^2 + 0.5\sigma_{WR}^2$ is the variance of $\overline{Y}_{ijR}$, or U_{ijR} in the notation of the FDA guidance document). The development of the HSCI method for PBE seems to have ignored this correlation which may explain why this method is rather conservative for establishing PBE. The same appears to be the case for the MLS method. The GPV method does properly account for this correlation and the use of the GPV method will give adequate power to assess PBE in situations where the Type II error rate of the other two methods may be as high as 70%. The substantial power gained by using the GPV method can also help ensure adequate power when there are dropouts in a study, something that is likely to be more of a problem in 4-period studies than it was when most BE studies were conducted with 2-period designs.

Figure 3.1: Discontinuity of HSCI at changeover point

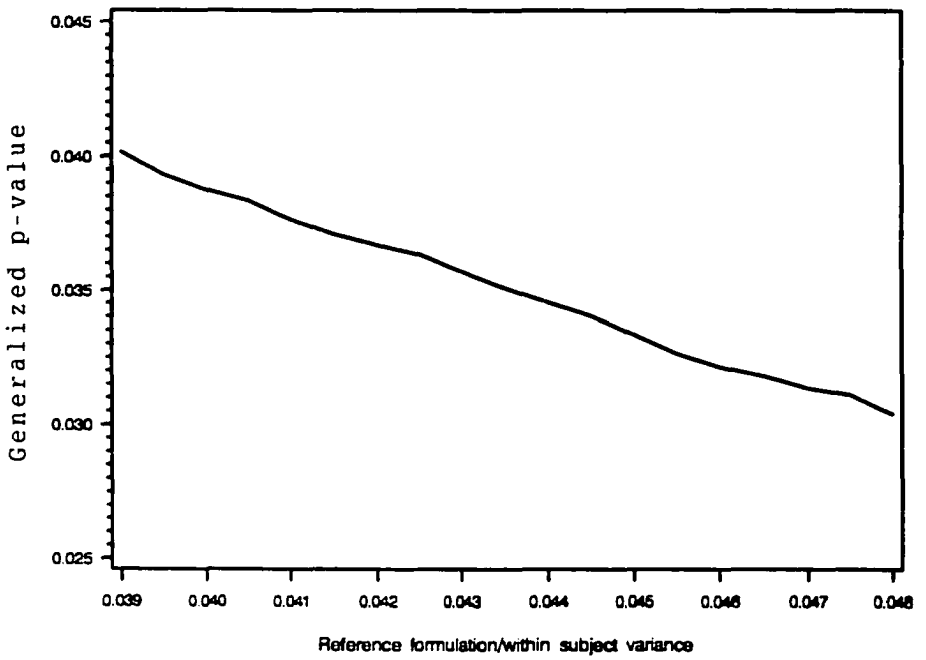


The procedure that we have applied for IBE in this examination uses the strict output of the function

$$max\left\{\frac{ss_{WR}}{U_{WR}}, \sigma_{W0}^2\right\}$$

to produce a single mixed-scaled test statistic. It is apparent from the literature, both statistical and pharmacological, that this approach is not universal (cf. Meyer *et al.* (2000) where both reference-scaled and constant-scaled HSCI bounds were reported for each pharmacokinetic parameter). In fact, the FDA draft guidance document on IBE presents it as two criteria [7]. The reason for this is that the HSCI bound is discontinuous at the changeover point.

Figure 3.2: Continuity of GPV at changeover point



An example illustrates the potential impact of the strict use of the mixed-scaled statistic with the HSCI method. Suppose that in a 2 sequence study with $n = 10$ in each sequence we have observed values of $\hat{\delta} = 0.15$, $\hat{\sigma}^2_{WT} = 0.05$, and $\hat{\sigma}^2_D = 0$. Now let $\hat{\sigma}^2_{WR}$ vary between 0.039 and 0.049, and apply the HSCI criterion with the mixed-scaled approach. The discontinuity can be seen clearly in Figure 3.1, which, in this case, has the disturbing consequence that the HSCI bound jumps from the region for concluding IBE to the region for not concluding IBE. That is, the HSCI concludes IBE when $\hat{\sigma}^2_{WR}$ is less than 0.04 and greater than 0.0475, but not at the intermediate values. Figure 3.2 shows the same testing situation using the GPV method: no discontinuity is seen (this is shown analytically in Chapter 4). It should be noted that the MLS method suffers from the same discontinuity problem as does the HSCI method.

The current regulatory position on handling the discontinuity in the HSCI upper bound is summarized in the following statement from the FDA guidance document (2001).

> This guidance recommends that sponsors applying the individual BE approach may use either reference-scaling or constant scaling at either side of the changeover point. With this approach, the multiple testing inflates the type I error rate slightly, to approximately 6.5% [*n.b.*, when we applied this to our simulation data, we observed a maximum type I error rate of 6.9%] but only over a small interval of σ_{WR} (about 0.18-0.20). [pg. 15]

However, consider the type I error rates given in Tables 2a-c. Over most of the parameter space, the size of the test is larger for HSCI than for GPV, the exception being when σ_{WR} is close to 0.2, *i.e.*, at the changeover point σ_{W0}. This is applying the strict mixed-scaled criterion to both GPV and HSCI. The GPV is slightly anti-conservative at $\sigma_{WR} = \sigma_{W0} = 0.2$. Based on a separate simulation study we have found that, when the 'play-the-winner' rule cited above is applied for testing IBE, the type-I error rates and the power for the HSCI test are very similar to that of the GPV test.

This discussion illustrates the main advantage of using the GPV method for IBE. It eliminates the need for *ad hoc* rules for choosing between the reference- and constant-scaled criteria. Examples of such rules include the FDA-sanctioned 'play-the-winner' rule, using a cutoff of $\sigma_{WR} = 0.25$ (suggested in an earlier draft of the FDA guidance document, but noticably absent from the final version), and even a formal test of $H_0 : \sigma_{WR} > 0.2$, suggested by Chow and Liu (2000), a strategy that raises additional multiple testing issues. Instead, by using the GPV test, we apply a criterion that does not penalize the analyst too much for using either reference- or constant- scaling in a particular region of the parameter space.

Given the discussion above, it may be appropriate for an investigator with extensive experience with a particular compound to specify the GPV testing procedure *a priori* if he feels that σ^2_{WR} is in the neighborhood of $0.04 - 0.07$ (this is the range for which some preliminary results suggest that GPV has greater power than HSCI). Such prior knowledge

could well exist in a large pharmaceutical company that is comparing from clinical to production batches of a given compound, or changing part of the manufacturing process. The discontinuity at the changeover point does not appear to be as great of an issue with respect to PBE, since σ^2_{TR} tends to be much larger than σ^2_{WR}.

Another advantage of the GPV method in general is that, by directly testing the null hypothesis, it provides analysts with an objective measure of the strength of the evidence against the null. This is illustrated by the 2 examples above. In assessing IBE with respect to the AUC, the HSCI method gives 95% upper bounds of -0.040 and -0.048 for verapamil and methylphenidate, respectively. These values may suggest some similarity in the degree to which the respective formulations are BE. However, the respective p-values are 0.018 and 0.003, which seems to suggest that the degree of evidence for IBE in the two studies may be quite different.

One supposed advantage of MLS over HSCI cited by Quiroz *et al.* (2002) is that it produces upper bounds for the bioequivalence criteria Θ_{IBE} and Θ_{PBE} themselves rather than on linearized versions of these. This point will be elaborated upon when we discuss generalized confidence intervals (Weerahandi, 1993) for Θ_{IBE} and Θ_{PBE} in Chapter 7.

It should be noted that, in general, the MLS methods for confidence intervals on variance components tend to be conservative (Burdick and Graybill (1992) give coverage probability for various classes of intervals which are very impressive and in line with what we have observed here). This conservatism may be desirable in certain applications, but, as we have shown, this also results in larger sample size requirements for a given level of power. Investigators who are contemplating using these methods need to assess the potential cost of using more conservative methods when testing for bioequivalence, especially when viable alternatives exist.

In conclusion, the results of this chapter show that the GPV method is a viable alternative to the HSCI method for IBE. These two methods are comparable with respect to type I error rate and power, while the MLS method appears to have somewhat less power in most cases. The GPV method is continuous at the changeover point from the reference-scaled criterion to the constant-scaled criterion whereas both HSCI and MLS are

discontinuous. Consequently, there is no need for *ad hoc* testing methods for parameter estimates in this region. For PBE, GPV is clearly more powerful in conditions that are very likely to occur in BE testing, *i.e.*, high correlation of intra-subject responses and the ratio of between-formulation to within-formulation variances being greater than one. The slight anti-conservatism of the GPV method is likely to be attenuated by the regulatory requirement that all test formulations show at least average BE with respect to reference formulations. In any case, the examples and simulation studies we have presented show that HSCI and MLS methods are overly conservative for PBE.

Chapter 4

TESTS FOR INDIVIDUAL AND POPULATION BIOEQUIVALENCE USING 3-PERIOD CROSSOVER DESIGNS

SUMMARY

The US Food and Drug Administration (FDA) has implemented new bioequivalence regulations that introduce two new criteria, population bioequivalence (PBE) and individual bioequivalence (IBE). The FDA recommends studies with a 4-period crossover designs to assess these criteria, and also recommends a testing procedure developed by Hyslop *et al.*, the Howe-type small-sample confidence interval (HSCI) An alternative testing procedure has been developed in Chapter 3 based on the the generalized p-value (GPV). This chapter compares the HSCI and GPV testing procedures for evaluating both IBE and PBE in 3-period crossover studies. We also compare the efficiency of the two designs we consider, the 3-period design given in the FDA guidance and an alternative design in which only the reference formulation is replicated, for both IBE and PBE. We show that the two tests are very similar for IBE, and that the GPV is significantly more powerful for PBE. We also show that a shortcoming of the HSCI test, the discontinuity in the test criteria when changing from constant to reference-scaling, is not present in the GPV test for IBE. We further find that the alternative 3-period crossover design is more efficient than the one in the FDA guidance for both IBE and PBE.

4.1 Introduction

The recently-adopted US Food and Drug Administration guidance document (2001) on statistical approaches to bioequivalence studies focuses on using 4-period crossover study designs to assess two new criteria, population bioequivalence (PBE) and individual bioequivalence (IBE). This same document alludes to the use of 3-period crossover designs for these purposes, but only gives a suggested study design with no other details. The objective of this investigation is to adapt the generalized p-value (GPV) methodology proposed in Chapter 3 for 4-period crossovers to 3-period designs, and to compare the GPV tests to Howe-type small sample confidence interval (HSCI) tests similar to those proposed by both Hyslop and Iglewicz (2001) and by Chow, Shao, and Wang (2002).

Section 4.2 introduces the designs, statistical model, and bioequivalence criteria that we will use. Section 4.3 details the basics of GPV tests, the summary statistics, and the statistical tests, GPV and HSCI, that we will compare. Section 4.4 uses simulations to assess and to compare the size and power of the two testing procedures. In Section 4.5 we discuss our findings and make some conclusions. Included in the discussion are some strategies for assessing the non-identifiable parameters in the extra-reference design.

4.2 Statistical Model

The decision criteria given in the current FDA guidance document on population and individual bioequivalence are based on four-period crossover designs. In this paper we consider the 2-sequence 3-period design (2S3P) and the 3-sequence 3-period design (3S3P) described in Tables 1a and 1b, respectively. The first design will hereafter be referred to as Design A or the *dual* design. The second design will be referred to as Design B or the *extra-reference* design.

Table 1a. Design A: 2-sequence 3-period *dual* design

Sequence 1	T	R	T
Sequence 2	R	T	R

Table 1b. Design B: 3-sequence 3-period *extra-reference* design

Sequence 1	T	R	R
Sequence 2	R	T	R
Sequence 3	R	R	T

The T and the R in the Table 1 refer to the test and reference formulations, respectively, of the drug product under consideration. In the context of an Abbreviated New Drug Application (ANDA), T is the generic product which we are trying to show is bioequivalent to the innovator (*i.e.*, brand-name) product R. The description of Design A as a *dual* designs refers to the fact that Sequence 2 has a R where Sequence 1 has a T, and vice-versa. Design B is called an *extra-reference* design because the reference formulation is replicated in each sequence. Hyslop and Iglewicz (2001) have considered this design for individual bioequivalence, and a similar 2-sequence extra reference design, with the first 2 sequences of Design B, has been considered by both Chow, Shao, and Wang (2002), and by Wang (1999). The 3-sequence extra reference design may be preferable to the 2-sequence version, since it guards against bias from any differential responses that may arise in either pharmacokinetic or clinical parameters in period three.

The observational model for a 3-period crossover design is given below.

$$Y_{ijkl} = \mu_k + \gamma_{ikl} + \eta_{ijk} + \epsilon_{ijkl} \quad i = 1, \ldots, s; \quad j = 1, \ldots, n_i; \quad k = T, R;$$

where Y_{ijkl} is the response from subject j in sequence i to the l^{th} application of treatment k, and s is the number of sequences. Here we encounter the first difference between 4 and 3-period designs: the value of l is either 1 or 2, but its range depends on the design, sequence, and treatment. The parameters μ_T and μ_R are the population means for treatments T and R, and γ_{ikl} is the fixed effect corresponding to the l^{th} application of treatment k in sequence i. The following estimability conditions are imposed on the γ_{ikl} (see Chinchilli and Esinhart, 1996).

$$\sum_{i=1}^{s} \sum_{l} \gamma_{ikl} = 0 \quad \text{for } k = T, R$$

Furthermore, η_{ijk} is the random effect of subject j in sequence i when receiving treatment k, and ϵ_{ijkl} is the within-subject random error term. The random variables ϵ_{ijkl} are jointly independent and distributed normally with mean 0 and variance σ^2_{Wk} $(k = R, T)$, and the vectors $\boldsymbol{\eta}_{ij} = [\eta_{ijT}, \eta_{ijR}]'$ are mutually independent bivariate normal with zero means and a covariance matrix given by

$$\boldsymbol{V} = \begin{pmatrix} \sigma^2_{BT} & \rho\sigma_{BT}\sigma_{BR} \\ \rho\sigma_{BT}\sigma_{BR} & \sigma^2_{BR} \end{pmatrix}.$$

Finally, $\{\boldsymbol{\eta}_{ij}\}$ and $\{\epsilon_{ijkl}\}$ are mutually independent. For convenience, we also define

$\delta = \mu_T - \mu_R$,

$\sigma^2_D = \sigma^2_{BT} + \sigma^2_{BR} - 2\rho\sigma_{BT}\sigma_{BR}$, $\sigma^2_{TT} = \sigma^2_{BT} + \sigma^2_{WT}$,

$\sigma_{TR} = \rho\sigma_{BT}\sigma_{BR}$, and $\sigma^2_{RR} = \sigma^2_{BR} + \sigma^2_{WR}$.

Recall that, for IBE, the parameter of interest is

$$\Theta_{IBE} = \begin{cases} \dfrac{\delta^2 + \sigma^2_D + \sigma^2_{WT} - \sigma^2_{WR}}{\sigma^2_{WR}} & \text{when } \sigma^2_{WR} > \sigma^2_0 \\ & \text{(reference-scaled criterion)} \\ \dfrac{\delta^2 + \sigma^2_D + \sigma^2_{WT} - \sigma^2_{WR}}{\sigma^2_{W0}} & \text{when } \sigma^2_{WR} \le \sigma^2_0 \\ & \text{(constant-scaled criterion)} \end{cases}$$

and, for PBE, the parameter of interest is

$$\Theta_{PBE} = \begin{cases} \dfrac{\delta^2 + \sigma^2_{TT} - \sigma^2_{RR}}{\sigma^2_{RR}} & \text{when } \sigma^2_{RR} > \sigma^2_0 \\ & \text{(reference-scaled criterion)} \\ \dfrac{\delta^2 + \sigma^2_{TT} - \sigma^2_{RR}}{\sigma^2_0} & \text{when } \sigma^2_{RR} \le \sigma^2_0 \\ & \text{(constant-scaled criterion)} \end{cases}$$

At present, the FDA recommends a value of $\sigma^2_0 = 0.04$, and upper limits of $\theta_I = 2.4948$, and $\theta_P = 1.7448$ for Θ_{IBE} and Θ_{PBE}, respectively.

4.3 The Generalized p-Value

We base our approach for testing for IBE and PBE on the generalized p-value (GPV) methodology of Tsui and Weerahandi (1989). This methodology is discussed extensively in Chapter 2. The idea is that we create a generalized test variable (GTV) with the following properties:

1. $T(\boldsymbol{x};\boldsymbol{x},\xi) = t$ is free of ξ.
2. For fixed $\boldsymbol{x}$ and ξ, the distribution of $T(\boldsymbol{X};\boldsymbol{x},\xi)$ is free of the nuisance parameter ζ.
3. For fixed $\boldsymbol{x}$ and ζ, $Pr(T(\boldsymbol{X};\boldsymbol{x},\xi) \geq t \mid \theta)$ is non-decreasing in θ.

From this test function we derive a test of the hypothesis of non-equivalence $H_0 : \theta > \theta_0$ versus the hypothesis of equivalance $H_a : \theta \leq \theta_0$ based on the generalized p-value

$$p = Pr[T(\boldsymbol{X};\boldsymbol{x},\xi) > t|\theta = \theta_0\,].$$

For a nominal α level test of H_0 against H_a one rejects H_0 whenever $p \leq \alpha$. See Chapter 2 for details.

We will compare the GPV test for PBE and IBE to the Howe-type small-sample confidence interval (HSCI) approach introduced by Hyslop *et al.* (2000) for assessing IBE in 4-period crossovers. The extention of HSCI tests for IBE to 3-period crossover studies have been developed by Hyslop and Iglewicz (2001) and Chow *et al.* (2002). We will use the same test as these authors for Design B, and a modification of the Chow HSCI test for Design A. Chapter 3 showed that for 4-period crossover designs the performance of the GPV is comparable to HSCI for IBE and is more powerful for PBE. The GPV approach is also free of an undesirable property of the HSCI in this problem: the discontinuity in the test criterion when one changes from constant to reference scaling. This last problem was discussed in detail in Chapter 3.

A third approach for assessing IBE and PBE was developed by Quiroz *et al.* (2002), who adapted the modified large sample methodology for confidence intervals on variance components introduced by Graybill and Wang (1980). Chapter 3 showed that, for 4-period

studies, this approach did not perform as well as HSCI and GPV approaches over much of the parameter space for IBE, and also did not perform as well as the GPV method for PBE. Therefore, we do not consider this approach in this investigation, but only our proposed method and the designated method of the FDA.

In both of the approaches above it is necessary to reparameterize the test criterion, expressing it in terms of $\sigma_I^2 = Var(\hat{\delta})$ instead of σ_D^2. The form of σ_I^2 depends on the design and, on occasion, the sequences of treatments within a design. The latter is the case with Design A. The summary statistics are also dependent on the design. Therefore, we present different parameterizations and sets of summary statistics for Design A and Design B.

4.3.1 Summary Statistics for Design A

Consider a bioequivalence study with the 2-sequence, 3-period crossover design described in Table 1a (Design A), where subjects are treated with the test and reference formulation twice in sequences 1 and 2, respectively. As before, let Y_{ijkl} denote the response of subject j in sequence i to the l^{th} application of treatment k ($k = T, R$). In the statistical model of Section 2, all of the statistics and parameter definitions that are needed to compute the IBE and the PBE criteria are given below.

$$\overline{Y}_{1jT} = \frac{1}{2}(Y_{1jT1} + Y_{1jT2}) \qquad \overline{Y}_{1jR} = Y_{1jR1} \qquad \overline{Y}_{2jT} = Y_{2jT1}$$

$$\overline{Y}_{2jR} = \frac{1}{2}(Y_{2jR1} + Y_{2jR2}) \qquad D_{ij} = \overline{Y}_{ijT} - \overline{Y}_{ijR} \qquad \overline{D}_i = \frac{1}{n_i}\sum_j D_{ij}$$

$$D = \frac{D_1 + D_2}{2} \qquad E_{1jT} = \frac{1}{\sqrt{2}}(Y_{1jT1} - Y_{1jT2}) \qquad E_{2jR} = \frac{1}{\sqrt{2}}(Y_{2jR1} - Y_{2jR2})$$

$$\overline{E}_{i\bullet k} = \frac{1}{n_i}\sum_{j=1}^{n_i} E_{ijk} \qquad \overline{Y}_{i\bullet k} = \frac{1}{n_i}\sum_{j=1}^{n_i} \overline{Y}_{ijk} \qquad SS_{Ii} = \textstyle\sum_{j=1}^{n_i}(D_{ij} - \overline{D}_i)^2$$

$$SS_{WT} = \textstyle\sum_{j=1}^{n_1}(E_{1jT} - \overline{E}_{1\bullet T})^2 \qquad SS_{WR} = \textstyle\sum_{j=1}^{n_2}(E_{2jR} - \overline{E}_{2\bullet R})^2 \qquad SS_{ki} = \textstyle\sum_{j=1}^{n_i}(\overline{Y}_{ijk} - \overline{Y}_{i\bullet k})^2$$

$$SS_{TRi} = \textstyle\sum_{j=1}^{n_i}(\overline{Y}_{ijT} - \overline{Y}_{i\bullet T})(\overline{Y}_{ijR} - \overline{Y}_{i\bullet R}) \qquad SS_{Ri|Ti} = SS_{Ri} - \frac{SS_{TRi}^2}{SS_{Ti}}$$

$$B_i = \frac{SS_{TR}}{SS_{Ti}} \qquad \delta = \mu_T - \mu_R \qquad \beta_i = \frac{\sigma_{TR}}{\sigma_{Ti}^2}$$

$$\sigma_D^2 = \sigma_{BT}^2 + \sigma_{BR}^2 - 2\rho\sigma_{BT}\sigma_{BR} \qquad \sigma_{T1}^2 = \sigma_{BT}^2 + \frac{1}{2}\sigma_{WT}^2$$

$$\sigma_{T2}^2 = \sigma_{BT}^2 + \sigma_{WT}^2 \qquad \sigma_{R1}^2 = \sigma_{BR}^2 + \sigma_{WR}^2 \qquad \sigma_{R2}^2 = \sigma_{BR}^2 + \frac{1}{2}\sigma_{WR}^2$$

$$\sigma_{TR} = \rho\sigma_{BT}\sigma_{BR} \qquad \sigma_{Ri|Ti}^2 = \sigma_{Ri}^2 - \frac{\sigma_{TRi}^2}{\sigma_{Ti}^2}$$

$$\sigma_{Ii}^2 = \sigma_{Ti}^2 + \sigma_{Ri}^2 - 2\sigma_{TRi} \qquad \nu_i = n_i - 1 \qquad w_i = \frac{\nu_i}{\nu_1 + \nu_2}$$

It is easily seen that the statistics $D, SS_{Ii}, SS_{WT}, SS_{WR}, SS_{Ti}, SS_{Ri|Ti}$, and B_i are associated with the following pivotal quantities with known distributions as indicated.

$$Z_D = \frac{D - \delta}{\frac{1}{2}\sqrt{\frac{\sigma_{I1}^2}{n_1} + \frac{\sigma_{I2}^2}{n_2}}} \sim N(0,1) \quad U_{Ii} = \frac{SS_{Ii}}{\sigma_{Ii}^2} \sim \chi_{\nu_i}^2 \quad U_{WT} = \frac{SS_{WT}}{\sigma_{WT}^2} \sim \chi_{\nu_1}^2$$

$$U_{WR} = \frac{SS_{WR}}{\sigma_{WR}^2} \sim \chi_{\nu_2}^2 \qquad U_{Ti} = \frac{SS_{Ti}}{\sigma_{Ti}^2} \sim \chi_{\nu_i}^2 \quad U_{Ri|Ti} = \frac{SS_{Ri|Ti}}{\sigma_{Ri|Ti}^2} \sim \chi_{\nu_i - 1}^2$$

$$Z_{Bi} = (B_i - \beta_i)\sqrt{\frac{SS_{Ti}}{\sigma_{Ri|Ti}^2}} \sim N(0,1)$$

The collection of random variables $Z_D, U_{I1}, U_{I2}, U_{WT}, U_{WR}$ are mutually independent. Also, the collection of random variables $Z_D, U_{WT}, U_{WR}, U_{T1}, U_{T2}, U_{R1|T1}, U_{R2|T2}, Z_{B1}, Z_{B2}$

are mutually independent. Note that, the quantities B_i and $SS_{R|Ti}$ are, respectively, the linear regression coefficient and the residual sum of squares from the regression of $\overline{Y}_{ijR}$ on $\overline{Y}_{ijT}$ within sequence i. In the development that follows, quantities in lower case, *viz.*, $d, ss_{Ii}, ss_{ki}, ss_{Wk}$, and b_i, are used to denote the observed values of the respective random variables, $D, SS_{Ii}, SS_{ki}, SS_{Wk}$, and B_i, which are given in upper case.

4.3.2 Tests of Individual Bioequivalence for Design A

Before giving the tests of IBE for Design A, it is useful to reparameterize the criteria given above.

$$\Theta_{IBE} = \frac{\delta^2 + w_1\left(\sigma_{I1}^2 + \frac{\sigma_{WT}^2}{2}\right) + w_2\left(\sigma_{I2}^2 + \frac{\sigma_{WR}^2}{2}\right) - 2\sigma_{WR}^2}{max(\sigma_{WR}^2, \sigma_0^2)}.$$

When $n_1 = n_2$, $w_1 = w_2 = \frac{1}{2}$, so the criteria becomes

$$\Theta_{IBE} = \frac{\delta^2 + \frac{\sigma_{I1}^2 + \sigma_{I2}^2}{2} + \frac{\sigma_{WT}^2}{4} - \frac{7\sigma_{WR}^2}{4}}{max(\sigma_{WR}^2, \sigma_0^2)}.$$

A generalized test variable for calculating the generalized p-value (GPV) is

$$T(X; x, \xi) = \frac{\left(d - \frac{Z_D}{2}\sqrt{\frac{ss_{I1}}{n_1 U_{I1}} + \frac{ss_{I2}}{n_2 U_{I2}}}\right)^2 + \left(\frac{w_1 ss_{I1}}{U_{I1}} + \frac{w_2 ss_{I2}}{U_{I2}}\right) + \frac{w_1}{2}\frac{ss_{WT}}{U_{WT}} + \left(\frac{w_2}{2} - 2\right)\frac{ss_{WR}}{U_{WR}}}{max\left(\frac{ss_{WR}}{U_{WR}}, \sigma_0^2\right)}$$

where $\xi = (\Theta_{IBE}, \delta, \sigma_{I1}^2, \sigma_{I2}^2, \sigma_{WT}^2, \sigma_{WR}^2)$. From this function we then calculate the area of the generalized extreme region, which here amounts to calculating the probability $Pr(T(X; x, \xi) \geq \theta_I)$. This probability is the GPV. We can compute this probability using Monte Carlo integration using a large number ($10000 - 250000$ or more) randomly-generated sets of the values $Z_D, U_{I1}, U_{I2}, U_{WT}$, and U_{WR}. We could then conclude IBE if the GPV is less than the designated α level. Typical values for the constants are $\sigma_0^2 = 0.04$ and $\theta_I = 2.4948$.

We compare this GPV to a test based on the HSCI developed by Hyslop *et al.* for 4-period designs. The $100(1-\alpha)\%$ upper bound for Θ_{IBE} for the Howe-type small sample confidence interval (HSCI) under the reference-scaled criterion is

$$d^2 + w_1\left(\frac{ss_{I1}}{\nu_1} + \frac{1}{2}\frac{ss_{WT}}{\nu_1}\right) + w_2\frac{ss_{I2}}{\nu_2} + \left(\frac{w_2}{2} - 2 - \theta_I\right)\frac{ss_{WR}}{\nu_2} + (H_D^2 + H_{I1}^2 + H_{I2}^2 + H_{WT}^2 + H_{WRR}^2)^{0.5}.$$

The upper bound for the constant-scaled criterion is

$$d^2+w_1\left(\frac{ss_{I1}}{\nu_1}+\frac{1}{2}\frac{ss_{WT}}{\nu_1}\right)+w_2\frac{ss_{I2}}{\nu_2}+\left(\frac{w_2}{2}-2\right)\frac{ss_{WR}}{\nu_2}-\sigma_0^2\theta_I+(H_D^2+H_{I1}^2+H_{I2}^2+H_{WT}^2+H_{WRC}^2)^{0.5}.$$

The terms in the variance expression are

$$H_D=\left(|d|+\frac{1}{2}t_{1-\alpha,\nu_*}\sqrt{\frac{ss_{I1}}{\nu_1 n_1}+\frac{ss_{I2}}{\nu_2 n_2}}\right)^2-d^2$$

$$H_{Ii}=w_i ss_{Ii}\left(\frac{1}{\chi^2_{\alpha,\nu_i}}-\frac{1}{\nu_i}\right) \qquad H_{WT}=\frac{w_1}{2}ss_{WT}\left(\frac{1}{\chi^2_{\alpha,\nu_1}}-\frac{1}{\nu_1}\right)$$

$$H_{WRR}=\left(\frac{w_2}{2}-2-\theta_I\right)ss_{WR}\left(\frac{1}{\chi^2_{1-\alpha,\nu_2}}-\frac{1}{\nu_2}\right) \quad H_{WRC}=\left(\frac{w_2}{2}-2\right)ss_{WR}\left(\frac{1}{\chi^2_{1-\alpha,\nu_2}}-\frac{1}{\nu_2}\right)$$

Here ν_* is the Satterthwaite-corrected degrees of freedom (see Steel, Torrie, and Dickey, 1997). The testing procedure is to conclude IBE when the relevant upper bound is less than zero. These intervals are similar to those of Chao *et al.* (2002), except that these authors do not use the Satterthwaite-corrected degrees of freedom.

4.3.3 Tests of Population Bioequivalence for Design A

For PBE, we assume that $n_1 = n_2$ to simplify the derivation of these test, so that the weights $w_1 = w_2 = 0.5$ are equal. Note that Chow *et al* used these weights in general without any justification for their HSCI test of IBE. The random variables SS_{Ti} and SS_{Ri} associated with the variance components σ^2_{Ti} and σ^2_{Ri} are not statistically independent. Therefore, we use the following reparameterization of the PBE criterion for the GPV method.

$$\Theta_{PBE}=\frac{\delta^2+\frac{1}{2}\left(\sigma^2_{T1}(1-\beta_1^2)+\frac{1}{2}\sigma^2_{WT}+\sigma^2_{T2}(1-\beta_2^2)-\left(\sigma^2_{R1|T1}+\sigma^2_{R2|T2}+\frac{1}{2}\sigma^2_{WR}\right)\right)}{max\left(\frac{1}{2}\left(\beta_1^2\sigma^2_{T1}+\sigma^2_{R1|T1}+\beta_2^2\sigma^2_{T2}+\sigma^2_{R2|T2}+\frac{1}{2}\sigma^2_{WR}\right),\sigma_0^2\right)}$$

A generalized test variable for calculating the generalized p-value is

$$T(X;x,\xi) = \frac{\left(d - \frac{Z_D}{2}\sqrt{\frac{T_D}{n}}\right)^2 + \frac{1}{2}\left(\frac{ss_{T1}}{U_{T1}} + \frac{ss_{WT}}{2U_{WT}} + \frac{ss_{T2}}{U_{T2}} - T_R\right)}{max(T_R, \sigma_0^2)}$$

$$T_R = \frac{ss_{T1}}{U_{T1}}\left(b_1 - Z_{B1}\sqrt{\frac{ss_{R1|T1}}{ss_{T1}U_{R1|T1}}}\right)^2 + \frac{ss_{R1|T1}}{U_{R1|T1}} + \frac{ss_{T2}}{U_{T2}}\left(b_2 - Z_{B2}\sqrt{\frac{ss_{R2|T2}}{ss_{T2}U_{R2|T2}}}\right)^2 + \frac{ss_{R2|T2}}{U_{R2|T2}} + \frac{1}{2}\frac{ss_{WR}}{U_{WR}}$$

$$T_D = \left(\frac{ss_{T1}}{U_{T1}}\left(1 - \left(b_1 - Z_{B1}\sqrt{\frac{ss_{R1|T1}}{ss_{T1}U_{R1|T1}}}\right)\right)^2 + \frac{ss_{R1|T1}}{U_{R1|T1}}\right) + \left(\frac{ss_{T2}}{U_{T2}}\left(1 - \left(b_2 - Z_{B2}\sqrt{\frac{ss_{R2|T2}}{ss_{T2}U_{R2|T2}}}\right)\right)^2 + \frac{ss_{R2|T2}}{U_{R2|T2}}\right)$$

where $\xi = (\Theta_{IBE}, \delta, \sigma_{I1}^2, \sigma_{I2}^2, \sigma_{T1}^2, \sigma_{T2}^2, \sigma_{R1|T1}^2, \sigma_{R2|T2}^2, \sigma_{WT}^2, \sigma_{WR}^2, \beta_1, \beta_2)$. Again, the GPV is computed by Monte Carlo integration using randomly-generated values of $Z_D, Z_{B1}, Z_{B2}, U_{T1}, U_{T2}, U_{R1|T1}, U_{R2|T2}, U_{WT}$, and U_{WR}. Then PBE is concluded if the GPV is less than the designated level of α. Typical values for the constants are $\sigma_0^2 = 0.04$ and $\theta_P = 1.7448$.

To assess PBE using the HSCI, we used the following reparameterization of the PBE criterion

$$\Theta_{PBE} = \frac{\delta^2 + \frac{1}{2}\left(\sigma_{T1}^2 + \frac{1}{2}\sigma_{WT}^2 + \sigma_{T2}^2 - \left(\sigma_{R1}^2 + \sigma_{R2}^2 + \frac{1}{2}\sigma_{WR}^2\right)\right)}{max\left(\frac{1}{2}(\sigma_{R1}^2 + \sigma_{R2}^2 + \frac{1}{2}\sigma_{WR}^2), \sigma_0^2\right)}.$$

The HSCI upper bound for Θ_{PBE} is under the reference-scaled criterion is then

$$\hat{\gamma}_R + \delta^2 + (H_D^2 + H_{T1}^2 + H_{WT}^2 + H_{T2}^2 + H_{R1R}^2 + H_{R2R}^2 + H_{WRR}^2)^{0.5}.$$

where

$$\hat{\gamma}_R = \delta^2 + \frac{1}{2}\left(\frac{ss_{T1}}{\nu_1} + \frac{ss_{WT}}{2\nu_1} + \frac{ss_{T2}}{\nu_2} - (1+\theta_P)\left(\frac{ss_{R1}}{\nu_1} + \frac{ss_{R2}}{\nu_2} + \frac{ss_{WR}}{2\nu_2}\right)\right).$$

The upper bound under the constant-scaled criterion is

$$\hat{\gamma}_C + (H_D^2 + H_{T1}^2 + H_{WT}^2 + H_{T2}^2 + H_{R1C}^2 + H_{R2C}^2 + H_{WRC}^2)^{0.5}.$$

where

$$\hat{\gamma}_C = \delta^2 + \frac{1}{2}\left(\frac{ss_{T1}}{\nu_1} + \frac{ss_{WT}}{2\nu_1} + \frac{ss_{T2}}{\nu_2} - \left(\frac{ss_{R1}}{\nu_1} + \frac{ss_{R2}}{\nu_2} + \frac{ss_{WR}}{2\nu_2}\right)\right) - \sigma_0^2\theta_P.$$

The quantity H_D is the same as it is for IBE, and the other components of the variance expression are as follows

$$H_{T1} = 0.5ss_{T1}\left(\frac{1}{\chi^2_{\alpha,\nu_1}} - \frac{1}{\nu_1}\right) \qquad H_{R1R} = 0.5(1+\theta_P)ss_{R1}\left(\frac{1}{\chi^2_{1-\alpha,\nu_1}} - \frac{1}{\nu_1}\right)$$

$$H_{T2} = 0.5ss_{T2}\left(\frac{1}{\chi^2_{\alpha,\nu_2}} - \frac{1}{\nu_2}\right) \qquad H_{R1C} = 0.5ss_{R1}\left(\frac{1}{\chi^2_{1-\alpha,\nu_1}} - \frac{1}{\nu_1}\right)$$

$$H_{R2R} = 0.5(1+\theta_P)ss_{R2}\left(\frac{1}{\chi^2_{1-\alpha,\nu_2}} - \frac{1}{\nu_2}\right) \quad H_{R2C} = 0.5ss_{R2}\left(\frac{1}{\chi^2_{1-\alpha,\nu_2}} - \frac{1}{\nu_2}\right)$$

$$H_{WT} = \frac{1}{4}ss_{WT}\left(\frac{1}{\chi^2_{\alpha,\nu_1}} - \frac{1}{\nu_1}\right) \qquad H_{WRR} = \frac{(1+\theta_P)}{4}ss_{WR}\left(\frac{1}{\chi^2_{1-\alpha,\nu_2}} - \frac{1}{\nu_2}\right)$$

$$H_{WRC} = \frac{1}{4}ss_{WR}\left(\frac{1}{\chi^2_{1-\alpha,\nu_2}} - \frac{1}{\nu_2}\right).$$

The test procedure is to conclude PBE if the upper bound is less than zero.

4.3.4 Example: Drug 15a

The data set for this example is from the FDA bioequivalence data website www.fda.gov/cder/bioequivdata. Drug 15a is therein described as an immediate release formulation of a hormone. The design is a 4-sequence/3-period crossover with sequences RTT and TRR added to the design. There are a total of 40 subjects, 10 in each sequence, in this study. The additional sequences RTT and TRR have the same summary statistics as those for TRT and RTR, respectively, in Design A. The following estimates were

computed for log(AUC)

$d = -0.01437$	$ss_{I1} = 2.0439$	$ss_{I2} = 2.7030$	$ss_{WT} = 0.53246$
$ss_{WR} = 1.3555$	$ss_{T1} = 4.9657$	$ss_{T2} = 7.9220$	$ss_{R1} = 7.5180$
$ss_{R2} = 5.2440$	$b_1 = 1.0512$	$b_2 = 0.6604$	$ss_{R1\|T1} = 2.0309$
$ss_{R2\|T2} = 1.7893$	$\nu_1 = \nu_2 = 18$	$\nu_* = 35.32$	

For IBE, the HSCI 95% upper bound is -0.03144, which is less than zero, so that we can conclude IBE. The GPV, based on 50000 Monte Carlo points, is 0.0293 (which has a 95% CI $(0.0278, 0.0308)$, based on the large-sample approximation to the binomial distribution), which is less than $\alpha = 0.05$, so we again conclude IBE. For PBE, the HSCI upper bound is -0.8774, and the GPV, based on 250000 Monte Carlo points, is less than 0.0001. From either test we can conclude PBE. One may be surprised by the apparent disparity, respectively, between, ss_{T1} and ss_{T2}, and between ss_{R1} and ss_{R2}, but it must be kept in mind that these summary statistics are all estimating different things (see the summary statistics above).

4.3.5 Amended and Additional Summary Statistics for Design B

Consider a BE study with the 3-period, 3-sequence, crossover design shown in Table 1b (Design B), where subjects are treated with the test formulation once and the reference formulation twice. As before, let Y_{ijkl} denote the response of subject j in sequence i to the l^{th} application of treatment k $(k = T, R)$. Again, the statistics and parameter definitions

that are needed to compute the IBE and the PBE criteria are below.

$$\overline{Y}_{ijR} = \frac{1}{2}(Y_{ijR1} + Y_{ijR2}) \qquad \overline{Y}_{ijT} = Y_{ijT1} \qquad E_{ijR} = \frac{1}{\sqrt{2}}(Y_{ijR1} - Y_{ijR2})$$

$$D_{ij} = \overline{Y}_{ijT} - \overline{Y}_{ijR} \qquad D = \frac{1}{3}\sum_i \overline{D}_i \qquad \overline{E}_{i\bullet R} = \frac{1}{n_i}\sum_{j=1}^{n_i} E_{ijR}$$

$$SS_I = \sum_{i=1}^{3}\sum_{j=1}^{n_i}(D_{ij} - \overline{D}_i)^2 \qquad SS_{WR} = \sum_{i=1}^{3}\sum_{j=1}^{n_i}(E_{ijR} - \overline{E}_{i\bullet R})^2$$

$$SS_k = \sum_{i=1}^{3}\sum_{j=1}^{n_i}(\overline{Y}_{ijK} - \overline{Y}_{i\bullet K})^2 \qquad SS_{TR} = \sum_{i=1}^{3}\sum_{j=1}^{n_i}(\overline{Y}_{ijT} - \overline{Y}_{i\bullet T})(\overline{Y}_{ijR} - \overline{Y}_{i\bullet R})$$

$$SS_{T|R} = SS_R - \frac{SS_{TR}^2}{SS_T} \qquad B = \frac{SS_{TR}}{SS_T}$$

$$\sigma_T^2 = \sigma_{BT}^2 + \sigma_{WT}^2 \qquad \sigma_R^2 = \sigma_{BR}^2 + \frac{1}{2}\sigma_{WR}^2 \qquad \beta = \frac{\sigma_{TR}}{\sigma_T^2}$$

$$\sigma_{R|T}^2 = \sigma_R^2 - \frac{\sigma_{TR}^2}{\sigma_T^2} \qquad c = \sqrt{\frac{1}{9}\sum_i \frac{1}{n_i}}$$

$$\sigma_I^2 = \sigma_T^2 + \sigma_R^2 - 2\sigma_{TR} = \sigma_D^2 + \sigma_{WT}^2 + \frac{1}{2}\sigma_{WR}^2 \qquad \nu = \sum_{i=1}^{3} n_i - 3.$$

The quantities $\overline{D}_i, \overline{Y}_{i\bullet k}, \delta, \sigma_{TR}, \sigma_D^2$ are the same as in Design A. It is easily seen that the statistics $D, SS_I, SS_{WR}, SS_T, SS_{R|T}$, and B are associated with the following pivotal quantities with known distributions as indicated.

$$Z_D = \frac{D-\delta}{c\sigma_I} \sim N(0,1) \quad U_I = \frac{SS_I}{\sigma_I^2} \sim \chi_\nu^2 \quad U_{WR} = \frac{SS_{WR}}{\sigma_{WR}^2} \sim \chi_\nu^2$$

$$U_k = \frac{SS_k}{\sigma_k^2} \sim \chi_\nu^2 \qquad k = T, R \qquad U_{R|T} = \frac{SS_{R|T}}{\sigma_{R|T}^2} \sim \chi_{\nu-1}^2$$

$$Z_B = (B-\beta)\sqrt{\frac{SS_T}{\sigma_{R|T}^2}} \sim N(0,1)$$

The collection of random variables Z_D, U_I, U_{WT}, U_{WR} are mutually independent. Also, the collection of random variables $Z_D, U_{WT}, U_{WR}, U_T, U_{R|T}, Z_B$ are mutually independent. Note that, the quantities B and $SS_{R|T}$ are, respectively, the linear regression coefficient

and the residual sum of squares from the regression of $\overline{Y}_{ijR}$ on $\overline{Y}_{ijT}$, after accounting for the sequence effect. Note that, since σ^2_{WT} is non-identifiable in this design, the quantities $E_{ijT}, \overline{E}_{i\bullet T}$, and SS_{WT}, defined above for Design A, do not exist here. For Design B, quantities in lower case–d, ss_I, ss_k, ss_{WR}, and b–are used to denote the observed values of the random variables–D, SS_I, SS_k, SS_{WR}, and B–which are given in upper case.

The FDA guidance document (2001) mandates the method of moments be used to estimate the fixed effects and variance components in the IBE/PBE model. We would like to note that, unlike Design A, the method of moments estimators for Design B are also the GLS/REML estimates. Chinchilli and Esinhart (1996) showed that this is the case for sequence-balanced designs (they used the term *uniform within sequence*) in which each treatment appears at least twice. A sequence-balanced design is one in which each treatment appear the same number of times in each sequence. An advantage of sequence-balanced designs is that each sequence generates the same information concerning the parameters in the model. Design B is an example of a sequence-balanced design, whereas Design A is not sequence-balanced. We prove the following proposition in the appendix.

Proposition. The method of moment estimators of the identifiable parameters in Design B, in particular $\delta, \sigma^2_T, \sigma^2_R, \sigma^2_I, \sigma^2_{WR}$, and β, are the same as the (unconstrained) GLS/REML estimates for these parameters.

4.3.6 Tests of Individual Bioequivalence for Design B

For Design B, we have the following reparameterization of the IBE criterion

$$\Theta_{IBE} = \frac{\delta^2 + \sigma^2_I - \frac{3}{2}\sigma^2_{WR}}{max(\sigma^2_{WR}, \sigma^2_0)}$$

A generalized test variable for computing the GPV in Design B is

$$T(X; x, \xi) = \frac{\left(d - cZ_D\sqrt{\frac{ss_I}{U_I}}\right)^2 + \frac{ss_I}{U_I} - \frac{3}{2}\frac{ss_{WR}}{U_{WR}}}{max\left(\frac{ss_{WR}}{U_{WR}}, \sigma^2_0\right)}$$

where $\xi = (\Theta_{IBE}, \delta, \sigma_I^2, \sigma_{WR}^2)$. The GPV is the area of the generalized extreme region for this function, which is calculated using Monte Carlo integration, and IBE is concluded if the GPV is less than α.

For comparison, we consider the HSCI intervals taken from Hyslop and Iglewicz (2001). The HSCI $100(1-\alpha)\%$ upper bound for the reference-scaled IBE criterion is

$$d^2 + \frac{ss_I}{\nu} - \left(\frac{3}{2} + \theta_I\right)\frac{ss_{WR}}{\nu} + (H_D^2 + H_I^2 + H_{WRR}^2)^{0.5}$$

The HSCI $100(1-\alpha)\%$ upper bound for the constant-scaled IBE criterion is

$$d^2 + \frac{ss_I}{\nu} - \frac{3}{2}\frac{ss_{WR}}{\nu} - \sigma_0^2\theta_I + (H_D + H_I^2 + H_{WRC}^2)^{0.5}$$

Individual BE is concluded if the appropriate upper bound is less than zero. The components of the variance expression are

$$H_D = \left(|d| + ct_{1-\alpha,\nu}\sqrt{\frac{ss_I}{\nu}}\right)^2 - d^2 \qquad H_I = ss_I\left(\frac{1}{\chi^2_{\alpha,\nu}} - \frac{1}{\nu}\right)$$

$$H_{WRR} = \left(\frac{3}{2} + \theta_I\right)ss_{WR}\left(\frac{1}{\chi^2_{1-\alpha,\nu}} - \frac{1}{\nu}\right) \qquad H_{WRC} = \frac{3}{2}ss_{WR}\left(\frac{1}{\chi^2_{1-\alpha,\nu}} - \frac{1}{\nu}\right)$$

Typical values for the constants in all these tests are $\sigma_0^2 = 0.04$ and $\theta_I = 2.4948$.

4.3.7 Test of PBE for Design B

As with Design A, we need a reparameterization of the PBE criteria that reflects the statistical dependence between the random variables SS_T and SS_R.

$$\Theta_{PBE} = \frac{\delta^2 + \sigma_T^2(1-\beta^2) - \left(\sigma_{R|T}^2 + \frac{1}{2}\sigma_{WR}^2\right)}{max\left(\sigma_T^2\beta^2 + \sigma_{R|T}^2 + \frac{1}{2}\sigma_{WR}^2, \sigma_0^2\right)}$$

The generalized test variable for computing the GPV in Design B is

$$T(X;x,\xi) = \frac{\left(d - cZ_D\sqrt{T_D}\right)^2 + \frac{ss_T}{U_T} - T_R}{max(T_R, \sigma_0^2)}$$

$$T_D = \frac{ss_{R|T}}{U_{R|T}} + \frac{ss_T}{U_T}\left(1 - \left(b - Z_B\sqrt{\frac{ss_{R|T}}{ss_T U_{R|T}}}\right)\right)^2$$

$$T_R = \frac{ss_T}{U_T}\left(b - Z_B\sqrt{\frac{ss_{R|T}}{ss_T U_{R|T}}}\right)^2 - \frac{ss_{R|T}}{U_{R|T}} - \frac{1}{2}\frac{ss_{WR}}{U_{WR}}$$

where $\xi = (\Theta_{IBE}, \delta, \sigma_I^2, \sigma_{R|T}^2, \sigma_{WR}^2, \beta)$. The GPV is the area of generalized extreme region for this function, which is computed using Monte Carlo integration, and PBE is concluded if the GPV is less than α. Typical values for the constants in all these tests are $\sigma_0^2 = 0.04$ and $\theta_P = 1.7448$.

The reparameterization of the PBE appropriate for the HSCI test is

$$\Theta_{PBE} = \frac{\delta^2 + \sigma_T^2 - \left(\sigma_R^2 + \frac{1}{2}\sigma_{WR}^2\right)}{max\left(\sigma_R^2 + \frac{1}{2}\sigma_{WR}^2, \sigma_0^2\right)}$$

The HSCI $100(1-\alpha)\%$ upper bound for the reference-scaled PBE criterion is

$$d^2 + \frac{ss_T}{\nu} - (1+\theta_P)\left(\frac{ss_R}{\nu} + \frac{1}{2}\frac{ss_{WR}}{\nu}\right) + (H_D^2 + H_T^2 + H_{RR}^2 + H_{WRR}^2)^{0.5}$$

The HSCI $100(1-\alpha)\%$ upper bound for the constant-scaled PBE criterion is

$$d^2 + \frac{ss_T}{\nu} - \left(\frac{ss_R}{\nu} + \frac{1}{2}\frac{ss_{WR}}{\nu}\right) - \sigma_0^2\theta_P + (H_D^2 + H_T^2 + H_{RR}^2 + H_{WRC}^2)^{0.5}$$

Population BE is concluded if the appropriate upper bound is less than zero. The components of the variance expression are

$$H_D = \left(|d| + ct_{1-\alpha,\nu}\sqrt{\frac{ss_I}{\nu}}\right)^2 - d^2 \qquad H_T = ss_T\left(\frac{1}{\chi^2_{\alpha,\nu}} - \frac{1}{\nu}\right)$$

$$H_{RR} = (\frac{1}{2} + \theta_P)ss_R\left(\frac{1}{\chi^2_{1-\alpha,\nu}} - \frac{1}{\nu}\right) \qquad H_{RC} = ss_R\left(\frac{1}{\chi^2_{1-\alpha,\nu}} - \frac{1}{\nu}\right)$$

$$H_{WRR} = \frac{1}{2}(1+\theta_P)ss_{WR}\left(\frac{1}{\chi^2_{1-\alpha,\nu}} - \frac{1}{\nu}\right) \quad H_{WRC} = \frac{1}{2}ss_{WR}\left(\frac{1}{\chi^2_{1-\alpha,\nu}} - \frac{1}{\nu}\right)$$

4.3.8 Example: Methylphenidate meta-data

Meyer *et al.* (2000), conducted a study to characterize the intra-subject variability of methylphenidate. In this study 20 healthy male subjects were randomized to one of four 4-period sequences. We adapted the data from this study in the following way:

1. Period 4 was deleted from Sequences 1 and 4.
2. Period 3 data in Sequence 2 was replaced with the data from Period 4.
3. Sequence 3 was completely deleted.

This gives data from a pseudo-study with treatment assignments conforming to Design B. This pseudo-study is constructed only for demonstration purposes and is not intended to be used to draw any conclusions about the IBE or PBE of methylphenidate. The observed values for the summary statistics for log(AUC) are

$$d = 0.10348 \qquad ss_I = 0.30837 \quad ss_T = 1.3758 \qquad ss_R = 1.3067$$
$$ss_{WR} = 0.52145 \quad b = 0.8628 \qquad ss_{R|T} = 0.2825 \quad \nu = 12$$

For IBE, we generated a HSCI upper bound of -0.052975 and a GPV of 0.0012. The GPV is based on 50000 Monte Carlo points and has a 95% confidence interval of $(0.0009, 0.0015)$. From either test we can conclude IBE. For PBE, the HSCI upper bound is -0.06926 and the GPV, based on 100000 Monte Carlo points, is 0.0004.

4.4 Simulation Studies

Simulation studies were conducted to compare the two methods – HSCI and GPV – for population bioequivalence and for individual bioequivalence. The methods were examined for their ability to maintain the nominal Type I error rate and for their power to reject the null hypothesis of bioinequivalence when the alternative hypothesis of bioequivalence is actually true. Since Design A has 2 sequences and Design B has 3, to facilitate comparison we let n represent the total number of subjects in a study. Simulation parameters were chosen based on the power tables given in the appendix of the FDA guidance document

(2001). For IBE, we used $\sigma_D = 0.01, 0.1, 0.15$, and, for assessing the sizes of the 2 test methods, chose the other parameter values so that $\theta_I = 2.4948$ (the boundary of the null hypothesis). The properties of these tests for PBE were examined in a similar manner. Five thousand simulated data sets were generated for each parameter combination. GPVs were estimated based on 10,000 Monte Carlo runs. All tests were done at the nominal $\alpha = 0.05$ level. All simulations were done using SAS® Version 8.2 for Windows. The results are summarized below for Design A and Design B, respectively, first for IBE, then for PBE.

4.4.1 Individual Bioequivalence for Design A

The examination of the size of the two procedures in Design A for IBE was done at a number of points on the boundary surface of the null hypothesis, i.e., where $\theta_I = 2.4948$ (the value proposed by the FDA). We also used the FDA-specified value $\sigma_0^2 = 0.04$ for constant-scaling. The estimated sizes of the procedures are summarized in Tables 2a-c.

Table 2a. Estimated sizes of the test procedures HSCI and GPV for Design A/IBE

	$\sigma_D = 0.01$									
	$\delta = 0.3157$				$\delta = 0.3631$		$\delta = 0.4737$		$\delta = 0.7897$	
	$\sigma_{WT} = \sigma_{WR} = 0.15$		$\sigma_{WT} = \sigma_{WR} = 0.2$		$\sigma_{WT} = \sigma_{WR} = 0.23$		$\sigma_{WT} = \sigma_{WR} = 0.3$		$\sigma_{WT} = \sigma_{WR} = 0.5$	
n	HSCI	GPV	HSCI	GPV	HSCI	GPV	HSCI	GPV	HSCI	GPV
24	0.042	0.040	0.045	0.047	0.043	0.037	0.039	0.029	0.039	0.028
36	0.043	0.038	0.058	0.060	0.048	0.041	0.051	0.042	0.052	0.042
48	0.047	0.042	0.059	0.065	0.046	0.040	0.044	0.037	0.046	0.038
	$\sigma_D = 0.10$									
	$\delta = 0.2997$				$\delta = 0.3493$		$\delta = 0.4631$		$\delta = 0.7834$	
	$\sigma_{WT} = \sigma_{WR} = 0.15$		$\sigma_{WT} = \sigma_{WR} = 0.2$		$\sigma_{WT} = \sigma_{WR} = 0.23$		$\sigma_{WT} = \sigma_{WR} = 0.3$		$\sigma_{WT} = \sigma_{WR} = 0.5$	
n	HSCI	GPV	HSCI	GPV	HSCI	GPV	HSCI	GPV	HSCI	GPV
24	0.040	0.033	0.045	0.043	0.044	0.035	0.045	0.034	0.040	0.032
36	0.041	0.035	0.051	0.055	0.045	0.040	0.047	0.039	0.049	0.038
48	0.044	0.038	0.049	0.053	0.047	0.039	0.053	0.046	0.048	0.040
	$\sigma_D = 0.15$									
	$\delta = 0.2780$				$\delta = 0.3309$		$\delta = 0.4495$		$\delta = 0.7754$	
	$\sigma_{WT} = \sigma_{WR} = 0.15$		$\sigma_{WT} = \sigma_{WR} = 0.2$		$\sigma_{WT} = \sigma_{WR} = 0.23$		$\sigma_{WT} = \sigma_{WR} = 0.3$		$\sigma_{WT} = \sigma_{WR} = 0.5$	
n	HSCI	GPV	HSCI	GPV	HSCI	GPV	HSCI	GPV	HSCI	GPV
24	0.034	0.028	0.042	0.040	0.038	0.031	0.044	0.032	0.042	0.032
36	0.044	0.038	0.050	0.048	0.041	0.033	0.043	0.033	0.048	0.039
48	0.040	0.033	0.052	0.057	0.049	0.041	0.047	0.038	0.048	0.040

Table 2b. Estimated sizes of the test procedures HSCI and GPV for Design A/IBE

	$\delta = 0.05$									
	$\sigma_D = 0.01$									
	$\sigma_{WT} = 0.3460$ $\sigma_{WR} = 0.15$		$\sigma_{WT} = 0.3704$ $\sigma_{WR} = 0.2$		$\sigma_{WT} = 0.4269$ $\sigma_{WR} = 0.23$		$\sigma_{WT} = 0.5585$ $\sigma_{WR} = 0.3$		$\sigma_{WT} = 0.9333$ $\sigma_{WR} = 0.5$	
n	HSCI	GPV	HSCI	GPV	HSCI	GPV	HSCI	GPV	HSCI	GPV
24	0.011	0.006	0.017	0.011	0.021	0.012	0.017	0.009	0.023	0.014
36	0.014	0.009	0.030	0.026	0.025	0.017	0.028	0.016	0.027	0.019
48	0.019	0.015	0.031	0.029	0.029	0.018	0.032	0.021	0.033	0.019
	$\sigma_D = 0.10$									
	$\sigma_{WT} = 0.3313$ $\sigma_{WR} = 0.15$		$\sigma_{WT} = 0.3568$ $\sigma_{WR} = 0.2$		$\sigma_{WT} = 0.4152$ $\sigma_{WR} = 0.23$		$\sigma_{WT} = 0.5496$ $\sigma_{WR} = 0.3$		$\sigma_{WT} = 0.9280$ $\sigma_{WR} = 0.5$	
n	HSCI	GPV	HSCI	GPV	HSCI	GPV	HSCI	GPV	HSCI	GPV
24	0.013	0.009	0.025	0.018	0.023	0.015	0.019	0.012	0.021	0.011
36	0.013	0.010	0.028	0.022	0.026	0.017	0.025	0.016	0.029	0.018
48	0.018	0.013	0.031	0.026	0.032	0.021	0.030	0.021	0.030	0.018
	$\sigma_D = 0.15$									
	$\sigma_{WT} = 0.3119$ $\sigma_{WR} = 0.15$		$\sigma_{WT} = 0.3388$ $\sigma_{WR} = 0.2$		$\sigma_{WT} = 0.3999$ $\sigma_{WR} = 0.23$		$\sigma_{WT} = 0.5381$ $\sigma_{WR} = 0.3$		$\sigma_{WT} = 0.9213$ $\sigma_{WR} = 0.5$	
n	HSCI	GPV	HSCI	GPV	HSCI	GPV	HSCI	GPV	HSCI	GPV
24	0.007	0.006	0.021	0.014	0.019	0.010	0.019	0.013	0.021	0.012
36	0.015	0.011	0.027	0.022	0.025	0.018	0.025	0.016	0.021	0.011
48	0.016	0.013	0.034	0.030	0.026	0.017	0.028	0.020	0.030	0.020

Table 2c. Estimated sizes of the test procedures HSCI and GPV for Design A/IBE

	$\delta = 0.15$									
	$\sigma_D = 0.01$									
	$\sigma_{WT} = 0.3157$ $\sigma_{WR} = 0.15$		$\sigma_{WT} = 0.3423$ $\sigma_{WR} = 0.2$		$\sigma_{WT} = 0.4269$ $\sigma_{WR} = 0.23$		$\sigma_{WT} = 0.5403$ $\sigma_{WR} = 0.3$		$\sigma_{WT} = 0.9225$ $\sigma_{WR} = 0.5$	
n	HSCI	GPV	HSCI	GPV	HSCI	GPV	HSCI	GPV	HSCI	GPV
24	0.013	0.010	0.027	0.020	0.024	0.016	0.024	0.015	0.021	0.012
36	0.021	0.016	0.033	0.029	0.027	0.016	0.026	0.018	0.029	0.017
48	0.024	0.017	0.039	0.040	0.033	0.023	0.032	0.022	0.034	0.023
	$\sigma_D = 0.10$									
	$\sigma_{WT} = 0.2997$ $\sigma_{WR} = 0.15$		$\sigma_{WT} = 0.3276$ $\sigma_{WR} = 0.2$		$\sigma_{WT} = 0.3903$ $\sigma_{WR} = 0.23$		$\sigma_{WT} = 0.5311$ $\sigma_{WR} = 0.3$		$\sigma_{WT} = 0.9172$ $\sigma_{WR} = 0.5$	
n	HSCI	GPV	HSCI	GPV	HSCI	GPV	HSCI	GPV	HSCI	GPV
24	0.016	0.013	0.027	0.022	0.021	0.013	0.025	0.013	0.025	0.015
36	0.020	0.016	0.033	0.031	0.027	0.018	0.029	0.021	0.029	0.017
48	0.022	0.018	0.038	0.034	0.034	0.024	0.030	0.020	0.029	0.021
	$\sigma_D = 0.15$									
	$\sigma_{WT} = 0.2780$ $\sigma_{WR} = 0.15$		$\sigma_{WT} = 0.3079$ $\sigma_{WR} = 0.2$		$\sigma_{WT} = 0.3740$ $\sigma_{WR} = 0.23$		$\sigma_{WT} = 0.5192$ $\sigma_{WR} = 0.3$		$\sigma_{WT} = 0.9103$ $\sigma_{WR} = 0.5$	
n	HSCI	GPV	HSCI	GPV	HSCI	GPV	HSCI	GPV	HSCI	GPV
24	0.015	0.013	0.026	0.020	0.024	0.014	0.025	0.013	0.019	0.012
36	0.018	0.015	0.032	0.029	0.031	0.021	0.032	0.021	0.030	0.018
48	0.026	0.021	0.039	0.038	0.036	0.027	0.030	0.021	0.027	0.016

The above tables indicate that, for the most part, both the GPV and HSCI tests hold the nominal level of $\alpha = 0.05$ with very few exceptions, most notably when $\delta > log(1.25)$ and $\sigma_{WR} = 0.2$. In many cases, both tests appear to be very conservative, which may have implications for the power in this design.

4.4.2 Power for IBE in Design A

Power of the two procedures was examined for $\delta = 0.05$, $\sigma_D = 0.01, 0.1, 0.15$, and $\sigma_{WT} = \sigma_{WR} = 0.15, 0.2, 0.23, 0.3, 0.5$. We have summarized these results by reporting the estimated minimum total sample size needed to achieve 80% power, and the estimate of the actual power at the given sample size for the 2 methods. These results are summarized in Table 3.

Table 3. Estimated power for IBE of the test procedures HSCI and GPV for Design A

		$\delta = 0.05$									
		$\sigma_{WT} = 0.15$ $\sigma_{WR} = 0.15$		$\sigma_{WT} = 0.2$ $\sigma_{WR} = 0.2$		$\sigma_{WT} = 0.23$ $\sigma_{WR} = 0.23$		$\sigma_{WT} = 0.3$ $\sigma_{WR} = 0.3$		$\sigma_{WT} = 0.5$ $\sigma_{WR} = 0.5$	
Method	σ_D	n	Power	n	Power	n	Power	n	Power	n	Power
HSCI	0.01	18	0.829	34	0.819	48	0.809	56	0.808	56	0.811
	0.10	24	0.807	44	0.813	62	0.807	64	0.816	58	0.812
	0.15	36	0.808	64	0.804	88	0.809	76	0.822	62	0.814
GPV	0.01	20	0.827	34	0.811	46	0.819	62	0.818	62	0.813
	0.10	28	0.847	44	0.805	58	0.807	68	0.802	64	0.815
	0.15	40	0.816	64	0.814	82	0.804	82	0.806	68	0.811

From the power comparisons it appears that GPV does somewhat better for values of σ_{WR} close to $\sigma_0 = 0.2$, and HSCI doing better when the σ_{WR} is relatively small or large. The differences between the sample size requirements for these 2 tests tended to be trivial for 4-period crossovers, which does not appear to be the case here. But what may strike the reader is the large number of subjects required even when the variation is modest. Bioequivalence studies rarely enroll more than 30 or 40 subjects. But there appears to be little advantage to running a 3-period study if Design A is used. This will be discussed further below.

4.4.3 Population Bioequivalence for Design A

Tables 4a and 4b show simulation results comparing the respective sizes of the two tests. All of the parameter combinations in these tables correspond to $\theta_P = 1.7448$, with σ_0^2 taken to be 0.04. Both values are as recommended in the FDA guidance document. Here, correlations of $\rho = 0.8$ and $\rho = 0.9$ are examined. The reasoning for examining cases with such seemingly high correlation is that, the responses within a subject to different formulations of the same drug are still likely to have a strong degree of correlation even if the formulations are only borderline bioequivalent.

Table 4a. Estimated sizes of the test procedures HSCI and GPV for Design A/PBE ($\delta \neq 0$)

		$\delta = 0.2642$		$\delta = 0.2954$		$\delta = 0.3736$	
		$\sigma_{BT} = 0.1, \sigma_{BR} = 0.1$		$\sigma_{BT} = 0.2, \sigma_{BR} = 0.2$		$\sigma_{BT} = 0.2, \sigma_{BR} = 0.2$	
		$\sigma_{WT} = 0.1, \sigma_{WR} = 0.1$		$\sigma_{WT} = 0.1, \sigma_{WR} = 0.1$		$\sigma_{WT} = 0.2, \sigma_{WR} = 0.2$	
n	ρ	HSCI	GPV	HSCI	GPV	HSCI	GPV
24	0.8	0.0016	0.0036	0.0250	0.0522	0.0158	0.0488
	0.9	0.0010	0.0036	0.0218	0.0392	0.0136	0.0432
36	0.8	0.0016	0.0012	0.0206	0.0444	0.0178	0.0386
	0.9	0.0020	0.0014	0.0208	0.0392	0.0144	0.0342
48	0.8	0.0008	0.0004	0.0206	0.0394	0.0138	0.0344
	0.9	0.0012	0.0012	0.0238	0.0358	0.0140	0.0322
		$\delta = 0.5907$		$\delta = 0.5604$		$\delta = 0.8861$	
		$\sigma_{BT} = 0.4, \sigma_{BR} = 0.4$		$\sigma_{BT} = 0.3, \sigma_{BR} = 0.3$		$\sigma_{BT} = 0.6, \sigma_{BR} = 0.6$	
		$\sigma_{WT} = 0.2, \sigma_{WR} = 0.2$		$\sigma_{WT} = 0.3, \sigma_{WR} = 0.3$		$\sigma_{WT} = 0.3, \sigma_{WR} = 0.3$	
n	ρ	HSCI	GPV	HSCI	GPV	HSCI	GPV
24	0.8	0.0186	0.0464	0.0166	0.0482	0.0208	0.0444
	0.9	0.0254	0.0480	0.0130	0.0474	0.0208	0.0420
36	0.8	0.0190	0.0362	0.0190	0.0418	0.0232	0.0378
	0.9	0.0218	0.0372	0.0158	0.0396	0.0212	0.0378
48	0.8	0.0208	0.0368	0.0126	0.0290	0.0200	0.0380
	0.9	0.0196	0.0326	0.0146	0.0292	0.0220	0.0346
		$\delta = 0.2954$				$\delta = 0.5907$	
		$\sigma_{BT} = 0.2, \sigma_{BR} = 0.1732$		$\sigma_{BT} = 0.1732, \sigma_{BR} = 0.2$		$\sigma_{BT} = 0.3873, \sigma_{BR} = 0.4$	
		$\sigma_{WT} = 0.1, \sigma_{WR} = 0.1414$		$\sigma_{WT} = 0.1414, \sigma_{WR} = 0.1$		$\sigma_{WT} = 0.2236, \sigma_{WR} = 0.2$	
n	ρ	HSCI	GPV	HSCI	GPV	HSCI	GPV
24	0.8	0.0194	0.0490	0.0182	0.0550	0.0228	0.0514
	0.9	0.0192	0.0408	0.0194	0.0498	0.0204	0.0442
36	0.8	0.0196	0.0346	0.0224	0.0508	0.0174	0.0364
	0.9	0.0194	0.0304	0.0158	0.0392	0.0252	0.0418
48	0.8	0.0202	0.0294	0.0158	0.0410	0.0208	0.0374
	0.9	0.0166	0.0264	0.0152	0.0348	0.0184	0.0320
		$\delta = 0.5907$		$\delta = 0.8861$			
		$\sigma_{BT} = 0.4, \sigma_{BR} = 0.3873$		$\sigma_{BT} = 0.6, \sigma_{BR} = 0.5$		$\sigma_{BT} = 0.5, \sigma_{BR} = 0.6$	
		$\sigma_{WT} = 0.2, \sigma_{WR} = 0.2236$		$\sigma_{WT} = 0.3, \sigma_{WR} = 0.4472$		$\sigma_{WT} = 0.4472, \sigma_{WR} = 0.3$	
n	ρ	HSCI	GPV	HSCI	GPV	HSCI	GPV
24	0.8	0.0188	0.0426	0.0184	0.0402	0.0168	0.0494
	0.9	0.0220	0.0380	0.0178	0.0380	0.0172	0.0490
36	0.8	0.0202	0.0408	0.0186	0.0348	0.0174	0.0460
	0.9	0.0212	0.0328	0.0166	0.0306	0.0144	0.0408
48	0.8	0.0248	0.0410	0.0158	0.0290	0.0138	0.0428
	0.9	0.0176	0.0282	0.0194	0.0272	0.0164	0.0370

Table 4b. Estimated sizes of the test procedures HSCI and GPV for Design A/PBE ($\delta = 0$)

		$\sigma_{BT} = 0.2825, \sigma_{BR} = 0.1$ $\sigma_{WT} = 0.1, \sigma_{WR} = 0.1$		$\sigma_{BT} = 0.3567, \sigma_{BR} = 0.2$ $\sigma_{WT} = 0.1, \sigma_{WR} = 0.1$		$\sigma_{BT} = 0.4238, \sigma_{BR} = 0.2$ $\sigma_{WT} = 0.2, \sigma_{WR} = 0.2$	
n	ρ	HSCI	GPV	HSCI	GPV	HSCI	GPV
24	0.8	0.0192	0.0244	0.0098	0.0616	0.0156	0.0642
	0.9	0.0226	0.0290	0.0086	0.0534	0.0110	0.0556
36	0.8	0.0254	0.0268	0.0154	0.0600	0.0182	0.0542
	0.9	0.0242	0.0286	0.0058	0.0584	0.0142	0.0480
48	0.8	0.0302	0.0290	0.0124	0.0588	0.0172	0.0544
	0.9	0.0302	0.0308	0.0070	0.0454	0.0144	0.0526
		$\sigma_{BT} = 0.7134, \sigma_{BR} = 0.4$ $\sigma_{WT} = 0.2, \sigma_{WR} = 0.2$		$\sigma_{BT} = 0.6357, \sigma_{BR} = 0.3$ $\sigma_{WT} = 0.3, \sigma_{WR} = 0.3$		$\sigma_{BT} = 1.0701, \sigma_{BR} = 0.6$ $\sigma_{WT} = 0.3, \sigma_{WR} = 0.3$	
n	ρ	HSCI	GPV	HSCI	GPV	HSCI	GPV
24	0.8	0.0104	0.0548	0.0180	0.0590	0.0130	0.0530
	0.9	0.0038	0.0538	0.0128	0.0554	0.0044	0.0500
36	0.8	0.0102	0.0516	0.0166	0.0572	0.0096	0.0534
	0.9	0.0038	0.0466	0.0154	0.0558	0.0078	0.0436
48	0.8	0.0112	0.0492	0.0192	0.0510	0.0108	0.0534
	0.9	0.0072	0.0426	0.0122	0.0488	0.0064	0.0438

Again, both tests appear to hold the nominal level of $\alpha = 0.05$ when $\delta \neq 0$. The GPV appears to be slightly anti-conservative when $\delta = 0$, whereas the HSCI test may viewed as somewhat over-conservative, since its estimated level is frequently lower than 0.02, quite a bit lower than $\alpha = 0.05$.

4.4.4 Power for PBE in Design A

Table 5 summarizes the simulation results for the power of the two tests. It shows the smallest n for which a particular test achieves 80% power and the actual estimate of the power for that test at that point. Here correlations of $\rho = 0.9$ and $\rho = 0.95$ are examined. These values were deemed reasonable for examining power because formulations that are bioequivalent should exhibit a very high degree of correlation.

Table 5. Sample size comparisons and estimated power for PBE of the test procedures HSCI and GPV for Design A

		$\delta = 0.05$							
		$\sigma_{WT} = 0.15, \sigma_{WR} = 0.15$				$\sigma_{WT} = 0.23, \sigma_{WR} = 0.23$			
		$\sigma_{BT} = 0.15$ $\sigma_{BR} = 0.15$		$\sigma_{BT} = 0.3$ $\sigma_{BR} = 0.3$		$\sigma_{BT} = 0.23$ $\sigma_{BR} = 0.23$		$\sigma_{BT} = 0.46$ $\sigma_{BR} = 0.46$	
Method	ρ	n	Power	n	Power	n	Power	n	Power
HSCI	0.90	22	0.801	30	0.831	30	0.835	28	0.811
	0.95	22	0.800	30	0.828	28	0.808	28	0.805
GPV	0.90	18	0.821	22	0.837	22	0.835	22	0.839
	0.95	18	0.831	22	0.829	22	0.832	22	0.826
		$\sigma_{WT} = 0.3, \sigma_{WR} = 0.3$				$\sigma_{WT} = 0.5, \sigma_{WR} = 0.5$			
		$\sigma_{BT} = 0.3$ $\sigma_{BR} = 0.3$		$\sigma_{BT} = 0.6$ $\sigma_{BR} = 0.6$		$\sigma_{BT} = 0.5$ $\sigma_{BR} = 0.5$		$\sigma_{BT} = 1.0$ $\sigma_{BR} = 1.0$	
Method	ρ	n	Power	n	Power	n	Power	n	Power
HSCI	0.90	28	0.808	28	0.809	28	0.821	28	0.803
	0.95	28	0.817	28	0.809	28	0.810	28	0.804
GPV	0.90	22	0.832	20	0.801	20	0.801	20	0.805
	0.95	22	0.829	20	0.800	22	0.845	20	0.801

Table 5 shows how that GPV has smaller sample size requirements than the HSCI method for most of the parameter space. The principal reason for this appears to be that the proposed GPV method accounts for the intra-subject dependence between $\overline{Y}_{ijT}$ and $\overline{Y}_{ijR}$.

4.4.5 Individual Bioequivalence for Design B

The examination of the size of the two procedures in Design B for IBE was done at a number of points on the boundary surface of the null hypothesis, *i.e.*, where $\theta_I = 2.4948$ (the value proposed by the FDA). We also used the FDA-specified value $\sigma_0^2 = 0.04$ for constant-scaling. The estimated sizes of the procedures are summarized in Tables 6a-c.

Table 6a. Estimated sizes of the test procedures HSCI and GPV for Design B/IBE

	$\sigma_D = 0.01$									
	$\delta = 0.3157$				$\delta = 0.3631$		$\delta = 0.4737$		$\delta = 0.7897$	
	$\sigma_{WT} = \sigma_{WR} = 0.15$		$\sigma_{WT} = \sigma_{WR} = 0.2$		$\sigma_{WT} = \sigma_{WR} = 0.23$		$\sigma_{WT} = \sigma_{WR} = 0.3$		$\sigma_{WT} = \sigma_{WR} = 0.5$	
n	HSCI	GPV	HSCI	GPV	HSCI	GPV	HSCI	GPV	HSCI	GPV
24	0.050	0.049	0.052	0.063	0.047	0.046	0.048	0.045	0.048	0.044
36	0.054	0.051	0.059	0.075	0.050	0.048	0.047	0.044	0.049	0.045
48	0.057	0.054	0.061	0.075	0.052	0.050	0.046	0.043	0.053	0.050
	$\sigma_D = 0.10$									
	$\delta = 0.2997$				$\delta = 0.3493$		$\delta = 0.4631$		$\delta = 0.7834$	
	$\sigma_{WT} = \sigma_{WR} = 0.15$		$\sigma_{WT} = \sigma_{WR} = 0.2$		$\sigma_{WT} = \sigma_{WR} = 0.23$		$\sigma_{WT} = \sigma_{WR} = 0.3$		$\sigma_{WT} = \sigma_{WR} = 0.5$	
n	HSCI	GPV	HSCI	GPV	HSCI	GPV	HSCI	GPV	HSCI	GPV
24	0.052	0.053	0.058	0.069	0.047	0.048	0.047	0.043	0.047	0.041
36	0.051	0.049	0.059	0.072	0.046	0.045	0.046	0.043	0.049	0.045
48	0.051	0.048	0.052	0.067	0.051	0.048	0.052	0.047	0.050	0.049
	$\sigma_D = 0.15$									
	$\delta = 0.2780$				$\delta = 0.3309$		$\delta = 0.4495$		$\delta = 0.7754$	
	$\sigma_{WT} = \sigma_{WR} = 0.15$		$\sigma_{WT} = \sigma_{WR} = 0.2$		$\sigma_{WT} = \sigma_{WR} = 0.23$		$\sigma_{WT} = \sigma_{WR} = 0.3$		$\sigma_{WT} = \sigma_{WR} = 0.5$	
n	HSCI	GPV	HSCI	GPV	HSCI	GPV	HSCI	GPV	HSCI	GPV
24	0.053	0.051	0.053	0.065	0.050	0.051	0.047	0.043	0.046	0.042
36	0.056	0.055	0.056	0.069	0.048	0.046	0.049	0.046	0.051	0.048
48	0.046	0.042	0.058	0.070	0.044	0.044	0.051	0.047	0.049	0.045

Table 6b. Estimated sizes of the test procedures HSCI and GPV for Design B/IBE

	$\delta = 0.05$									
	$\sigma_D = 0.01$									
	$\sigma_{WT} = 0.3460$ $\sigma_{WR} = 0.15$		$\sigma_{WT} = 0.3704$ $\sigma_{WR} = 0.2$		$\sigma_{WT} = 0.4269$ $\sigma_{WR} = 0.23$		$\sigma_{WT} = 0.5585$ $\sigma_{WR} = 0.3$		$\sigma_{WT} = 0.9333$ $\sigma_{WR} = 0.5$	
n	HSCI	GPV	HSCI	GPV	HSCI	GPV	HSCI	GPV	HSCI	GPV
24	0.031	0.030	0.045	0.049	0.035	0.036	0.034	0.031	0.038	0.033
36	0.038	0.035	0.048	0.053	0.042	0.040	0.038	0.035	0.037	0.034
48	0.039	0.036	0.054	0.058	0.040	0.039	0.038	0.034	0.040	0.037
	$\sigma_D = 0.10$									
	$\sigma_{WT} = 0.3313$ $\sigma_{WR} = 0.15$		$\sigma_{WT} = 0.3568$ $\sigma_{WR} = 0.2$		$\sigma_{WT} = 0.4152$ $\sigma_{WR} = 0.23$		$\sigma_{WT} = 0.5496$ $\sigma_{WR} = 0.3$		$\sigma_{WT} = 0.9280$ $\sigma_{WR} = 0.5$	
n	HSCI	GPV	HSCI	GPV	HSCI	GPV	HSCI	GPV	HSCI	GPV
24	0.038	0.035	0.048	0.052	0.040	0.039	0.034	0.031	0.040	0.035
36	0.039	0.037	0.050	0.056	0.039	0.038	0.037	0.033	0.038	0.033
48	0.041	0.037	0.047	0.055	0.050	0.048	0.038	0.035	0.041	0.037
	$\sigma_D = 0.15$									
	$\sigma_{WT} = 0.3119$ $\sigma_{WR} = 0.15$		$\sigma_{WT} = 0.3388$ $\sigma_{WR} = 0.2$		$\sigma_{WT} = 0.3999$ $\sigma_{WR} = 0.23$		$\sigma_{WT} = 0.5381$ $\sigma_{WR} = 0.3$		$\sigma_{WT} = 0.9213$ $\sigma_{WR} = 0.5$	
n	HSCI	GPV	HSCI	GPV	HSCI	GPV	HSCI	GPV	HSCI	GPV
24	0.032	0.030	0.044	0.047	0.036	0.035	0.034	0.031	0.039	0.034
36	0.038	0.034	0.048	0.053	0.044	0.043	0.034	0.030	0.041	0.037
48	0.038	0.036	0.054	0.061	0.041	0.039	0.041	0.037	0.045	0.040

Table 6c. Estimated sizes of the test procedures HSCI and GPV for Design B/IBE

	$\delta = 0.15$									
	$\sigma_D = 0.01$									
	$\sigma_{WT} = 0.3157$ $\sigma_{WR} = 0.15$		$\sigma_{WT} = 0.3423$ $\sigma_{WR} = 0.2$		$\sigma_{WT} = 0.4269$ $\sigma_{WR} = 0.23$		$\sigma_{WT} = 0.5403$ $\sigma_{WR} = 0.3$		$\sigma_{WT} = 0.9225$ $\sigma_{WR} = 0.5$	
n	HSCI	GPV	HSCI	GPV	HSCI	GPV	HSCI	GPV	HSCI	GPV
24	0.034	0.033	0.054	0.059	0.041	0.040	0.039	0.035	0.043	0.038
36	0.044	0.042	0.054	0.060	0.036	0.037	0.042	0.038	0.040	0.036
48	0.042	0.040	0.051	0.062	0.042	0.042	0.036	0.033	0.041	0.036
	$\sigma_D = 0.10$									
	$\sigma_{WT} = 0.2997$ $\sigma_{WR} = 0.15$		$\sigma_{WT} = 0.3276$ $\sigma_{WR} = 0.2$		$\sigma_{WT} = 0.3903$ $\sigma_{WR} = 0.23$		$\sigma_{WT} = 0.5311$ $\sigma_{WR} = 0.3$		$\sigma_{WT} = 0.9172$ $\sigma_{WR} = 0.5$	
n	HSCI	GPV	HSCI	GPV	HSCI	GPV	HSCI	GPV	HSCI	GPV
24	0.037	0.035	0.047	0.053	0.039	0.038	0.041	0.037	0.031	0.028
36	0.042	0.041	0.050	0.058	0.043	0.041	0.045	0.041	0.039	0.035
48	0.039	0.037	0.053	0.062	0.042	0.041	0.041	0.038	0.042	0.038
	$\sigma_D = 0.15$									
	$\sigma_{WT} = 0.2780$ $\sigma_{WR} = 0.15$		$\sigma_{WT} = 0.3079$ $\sigma_{WR} = 0.2$		$\sigma_{WT} = 0.3740$ $\sigma_{WR} = 0.23$		$\sigma_{WT} = 0.5192$ $\sigma_{WR} = 0.3$		$\sigma_{WT} = 0.9103$ $\sigma_{WR} = 0.5$	
n	HSCI	GPV	HSCI	GPV	HSCI	GPV	HSCI	GPV	HSCI	GPV
24	0.039	0.038	0.048	0.056	0.042	0.040	0.035	0.031	0.035	0.030
36	0.043	0.041	0.050	0.055	0.037	0.036	0.037	0.034	0.040	0.037
48	0.041	0.040	0.056	0.063	0.046	0.045	0.044	0.040	0.043	0.039

Both tests appear to hold the nominal level of $\alpha = 0.05$ for most of the parameter space. The exceptions are mostly when σ_{WR} is close to 0.2, especially for the GPV and when $\delta > log(1.25)$. On the whole, the tests appear to be very similar.

4.4.6 Power for IBE in Design B

Power of the two procedures was examined for $\delta = 0.05$, $\sigma_D = 0.01, 0.1, 0.15$, and $\sigma_{WT} = \sigma_{WR} = 0.15, 0.2, 0.23, 0.3, 0.5$. We have summarized these results by reporting the estimated minimum total sample size needed to achieve 80% power, and the estimate of the actual power at the given sample size for the 2 methods. These results are summarized in Table 7.

Table 7. Estimated power for IBE of the test procedures HSCI and GPV for Design B

		$\delta = 0.05$									
		$\sigma_{WT} = 0.15$ $\sigma_{WR} = 0.15$		$\sigma_{WT} = 0.2$ $\sigma_{WR} = 0.2$		$\sigma_{WT} = 0.23$ $\sigma_{WR} = 0.23$		$\sigma_{WT} = 0.3$ $\sigma_{WR} = 0.3$		$\sigma_{WT} = 0.5$ $\sigma_{WR} = 0.5$	
Method	σ_D	n	Power	n	Power	n	Power	n	Power	n	Power
HSCI	0.01	15	0.889	24	0.835	30	0.813	36	0.846	33	0.819
	0.10	18	0.820	30	0.811	39	0.809	39	0.815	36	0.823
	0.15	27	0.820	45	0.808	54	0.805	45	0.818	39	0.837
GPV	0.01	15	0.879	24	0.852	30	0.840	36	0.837	33	0.805
	0.10	18	0.804	30	0.827	36	0.809	39	0.804	36	0.812
	0.15	27	0.808	42	0.802	51	0.815	45	0.806	39	0.823

From the power comparisons it appears that HSCI and GPV are very similar with respect to power and sample size requirements, with GPV doing somewhat better for values of σ_{WR} close to $\sigma_0 = 0.2$, and HSCI doing marginally better when the σ_{WR} is relatively small or large. This is the same pattern observed in the previous chapter with respect to 4-period crossover designs. The differences between 2 methods may appear to be trivial, with the cases where GPV has a lower sample size requirement only amounting to one subject per sequence. But, as discussed in Chapter 3, there may be other reasons for preferring GPV over HSCI. The sample size requirements for Design B appear to be much more reasonable than for Design A, and these requirements also compare favorably to those for 4-period crossover. This will be discussed further in Section 5.

4.4.7 Population Bioequivalence for Design B

Tables 8a and 8b show simulation results comparing the respective sizes of the 2 tests. All of the parameter combinations in these tables correspond to $\theta_P = 1.7448$, with σ_0^2 taken to be 0.04. Both values are as recommended in the FDA guidance document. Correlations of $\rho = 0.8$ and $\rho = 0.9$ are examined, as we did for Design A.

Table 8a. Estimated sizes of the test procedures HSCI and GPV for Design B/PBE ($\delta \neq 0$)

		$\delta = 0.2642$		$\delta = 0.2954$		$\delta = 0.3736$	
		$\sigma_{BT} = 0.1, \sigma_{BR} = 0.1$		$\sigma_{BT} = 0.2, \sigma_{BR} = 0.2$		$\sigma_{BT} = 0.2, \sigma_{BR} = 0.2$	
		$\sigma_{WT} = 0.1, \sigma_{WR} = 0.1$		$\sigma_{WT} = 0.1, \sigma_{WR} = 0.1$		$\sigma_{WT} = 0.2, \sigma_{WR} = 0.2$	
n	ρ	HSCI	GPV	HSCI	GPV	HSCI	GPV
24	0.8	0.0446	0.0606	0.0202	0.0676	0.0334	0.0714
	0.9	0.0372	0.0522	0.0120	0.0634	0.0352	0.0726
36	0.8	0.0434	0.0576	0.0274	0.0710	0.0362	0.0650
	0.9	0.0410	0.0528	0.0152	0.0618	0.0370	0.0714
48	0.8	0.0492	0.0590	0.0218	0.0550	0.0396	0.0674
	0.9	0.0446	0.0558	0.0202	0.0628	0.0342	0.0640
		$\delta = 0.5907$		$\delta = 0.5604$		$\delta = 0.8861$	
		$\sigma_{BT} = 0.4, \sigma_{BR} = 0.4$		$\sigma_{BT} = 0.3, \sigma_{BR} = 0.3$		$\sigma_{BT} = 0.6, \sigma_{BR} = 0.6$	
		$\sigma_{WT} = 0.2, \sigma_{WR} = 0.2$		$\sigma_{WT} = 0.3, \sigma_{WR} = 0.3$		$\sigma_{WT} = 0.3, \sigma_{WR} = 0.3$	
n	ρ	HSCI	GPV	HSCI	GPV	HSCI	GPV
24	0.8	0.0236	0.0652	0.0328	0.0686	0.0214	0.0654
	0.9	0.0142	0.0624	0.0294	0.0692	0.0126	0.0654
36	0.8	0.0220	0.0592	0.0388	0.0680	0.0202	0.0550
	0.9	0.0136	0.0646	0.0316	0.0640	0.0154	0.0568
48	0.8	0.0284	0.0692	0.0348	0.0636	0.0248	0.0622
	0.9	0.0196	0.0608	0.0378	0.0662	0.0194	0.0588
		$\delta = 0.2954$				$\delta = 0.5907$	
		$\sigma_{BT} = 0.2, \sigma_{BR} = 0.1732$		$\sigma_{BT} = 0.1732, \sigma_{BR} = 0.2$		$\sigma_{BT} = 0.3873, \sigma_{BR} = 0.4$	
		$\sigma_{WT} = 0.1, \sigma_{WR} = 0.1414$		$\sigma_{WT} = 0.1414, \sigma_{WR} = 0.1$		$\sigma_{WT} = 0.2236, \sigma_{WR} = 0.2$	
n	ρ	HSCI	GPV	HSCI	GPV	HSCI	GPV
24	0.8	0.0202	0.0736	0.0244	0.0688	0.0180	0.0576
	0.9	0.0210	0.0778	0.0190	0.0608	0.0154	0.0644
36	0.8	0.0272	0.0652	0.0250	0.0610	0.0238	0.0600
	0.9	0.0212	0.0610	0.0240	0.0666	0.0168	0.0640
48	0.8	0.0316	0.0624	0.0310	0.0628	0.0290	0.0646
	0.9	0.0244	0.0608	0.0250	0.0628	0.0174	0.0556
		$\delta = 0.5907$		$\delta = 0.8861$			
		$\sigma_{BT} = 0.4, \sigma_{BR} = 0.3873$		$\sigma_{BT} = 0.6, \sigma_{BR} = 0.5$		$\sigma_{BT} = 0.5, \sigma_{BR} = 0.6$	
		$\sigma_{WT} = 0.2, \sigma_{WR} = 0.2236$		$\sigma_{WT} = 0.3, \sigma_{WR} = 0.4472$		$\sigma_{WT} = 0.4472, \sigma_{WR} = 0.3$	
n	ρ	HSCI	GPV	HSCI	GPV	HSCI	GPV
24	0.8	0.0206	0.0658	0.0258	0.0682	0.0294	0.0662
	0.9	0.0120	0.0618	0.0196	0.0728	0.0232	0.0698
36	0.8	0.0250	0.0642	0.0254	0.0596	0.0266	0.0574
	0.9	0.0176	0.0616	0.0256	0.0632	0.0254	0.0594
48	0.8	0.0214	0.0580	0.0318	0.0676	0.0344	0.0670
	0.9	0.0186	0.0608	0.0248	0.0622	0.0280	0.0604

Table 8b. Estimated sizes of the test procedures HSCI and GPV for Design B/PBE ($\delta = 0$)

		$\sigma_{BT} = 0.2825, \sigma_{BR} = 0.1$ $\sigma_{WT} = 0.1, \sigma_{WR} = 0.1$		$\sigma_{BT} = 0.3567, \sigma_{BR} = 0.2$ $\sigma_{WT} = 0.1, \sigma_{WR} = 0.1$		$\sigma_{BT} = 0.4238, \sigma_{BR} = 0.2$ $\sigma_{WT} = 0.2, \sigma_{WR} = 0.2$	
n	ρ	HSCI	GPV	HSCI	GPV	HSCI	GPV
24	0.8	0.0330	0.0404	0.0102	0.0630	0.0212	0.0570
	0.9	0.0342	0.0474	0.0074	0.0722	0.0154	0.0596
36	0.8	0.0332	0.0378	0.0086	0.0638	0.0176	0.0534
	0.9	0.0368	0.0468	0.0052	0.0626	0.0198	0.0622
48	0.8	0.0328	0.0408	0.0110	0.0620	0.0222	0.0574
	0.9	0.0338	0.0420	0.0040	0.0568	0.0160	0.0592
		$\sigma_{BT} = 0.7134, \sigma_{BR} = 0.4$ $\sigma_{WT} = 0.2, \sigma_{WR} = 0.2$		$\sigma_{BT} = 0.6357, \sigma_{BR} = 0.3$ $\sigma_{WT} = 0.3, \sigma_{WR} = 0.3$		$\sigma_{BT} = 1.0701, \sigma_{BR} = 0.6$ $\sigma_{WT} = 0.3, \sigma_{WR} = 0.3$	
n	ρ	HSCI	GPV	HSCI	GPV	HSCI	GPV
24	0.8	0.0088	0.0588	0.0214	0.0584	0.0092	0.0584
	0.9	0.0034	0.0554	0.0158	0.0578	0.0032	0.0558
36	0.8	0.0080	0.0584	0.0228	0.0600	0.0074	0.0452
	0.9	0.0028	0.0510	0.0162	0.0520	0.0022	0.0524
48	0.8	0.0108	0.0594	0.0204	0.0564	0.0084	0.0574
	0.9	0.0026	0.0498	0.0196	0.0596	0.0030	0.0498

The size of the HSCI appears to hold the nominal level of the test, but is sometimes much smaller than $\alpha = 0.05$, raising concerns about excessive Type II errors. The GPV is somewhat anti-conservative, but this effect is most pronounced when $\delta > log(1.25)$. Since concluding PBE requires some form of ABE, the somewhat inflated estimated sizes in Table 8a may be of little practical concern.

4.4.8 Power for PBE in Design B

Table 9 summarizes the simulation results for the power of the two tests. It shows the smallest total number of subjects for which a particular test achieves 80% power and the actual estimate of the power for that test at that point. Here correlations of $\rho = 0.9$ and $\rho = 0.95$ are examined. As with Design A, these values were deemed reasonable for examining power because formulations that are bioequivalent should exhibit a very high degree of correlation.

Table 9. Estimated power for PBE of the test procedures HSCI and GPV for Design B

		$\delta = 0.05$							
		$\sigma_{WT} = 0.15, \sigma_{WR} = 0.15$				$\sigma_{WT} = 0.23, \sigma_{WR} = 0.23$			
		$\sigma_{BT} = 0.15$ $\sigma_{BR} = 0.15$		$\sigma_{BT} = 0.3$ $\sigma_{BR} = 0.3$		$\sigma_{BT} = 0.23$ $\sigma_{BR} = 0.23$		$\sigma_{BT} = 0.46$ $\sigma_{BR} = 0.46$	
Method	ρ	n	Power	n	Power	n	Power	n	Power
HSCI	0.90	24	0.865	24	0.840	27	0.860	24	0.842
	0.95	21	0.811	24	0.854	24	0.809	24	0.877
GPV	0.90	15	0.803	15	0.863	21	0.869	15	0.866
	0.95	15	0.813	12	0.821	18	0.821	12	0.826
		$\sigma_{WT} = 0.3, \sigma_{WR} = 0.3$				$\sigma_{WT} = 0.5, \sigma_{WR} = 0.5$			
		$\sigma_{BT} = 0.3$ $\sigma_{BR} = 0.3$		$\sigma_{BT} = 0.6$ $\sigma_{BR} = 0.6$		$\sigma_{BT} = 0.5$ $\sigma_{BR} = 0.5$		$\sigma_{BT} = 1.0$ $\sigma_{BR} = 1.0$	
Method	ρ	n	Power	n	Power	n	Power	n	Power
HSCI	0.90	27	0.869	24	0.854	24	0.814	24	0.852
	0.95	24	0.806	24	0.870	24	0.813	24	0.870
GPV	0.90	18	0.805	15	0.861	18	0.820	15	0.863
	0.95	18	0.822	12	0.833	18	0.822	12	0.832

As we already have observed for Design A, the GPV is more powerful than the HSCI for PBE, which we feel is due to GPV accounting for the intra-subject correlation between the estimates of the between-subject variances. Also, the sample size requirements are much lower for Design B than Design A, and comparable to the sample size requirements derived in Chapter 3 for assessing PBE in 4-period crossovers.

4.5 Discussion

The conclusions of this investigation can be summarized as follows

1. Design A is very inefficient, with respect to the number of subjects needed to conclude bioequivalence compared to Design B.

2. Most of the discussion of Chapter 3 with respect to 4-period crossover designs also appears to apply to 3-period crossovers, particularly to Design B.

3. Design B has comparable performance to 4-period crossovers, but some parameters are not identifiable.

Design A is given by the FDA guidance document (2001) as an example of a 3-period crossover design that can be used to assess individual bioequivalence. The guidance doc-

ument contains no other details concerning 3-period crossovers. In particular, the FDA guidance does not discuss how to analyze population and individual BE for 3-period crossovers, although it is relatively straightforward to do so by applying either the Hyslop methodology or the GPV. No justification is given as to why Design A should be preferred to any other 3-period design (the same can also be said of the 4-period designs that the guidance document recommends). Therefore, it may come as a surprise to some that Design A appears to have such poor efficiency compared to Design B when the number of subjects needed to yield a reasonable degree of power is assessed. This poor efficiency for IBE is apparent in Table 10, which gives the ratio of the estimated required sample sizes for 80% power when Table 3 is compared to Table 7. A similar, although less dramatic, effect is seen for PBE in Table 11, which is comparing Table 5 to Table 9. Later we will see similar comparisons of Design A and Design B to the 2-sequence/4-period (2S4P) design studied in the previous chapter which will show that the number of 'doses' (total number of subjects times the number of periods) required in Design A is $30 - 80\%$ higher than in the 2S4P design, whereas Design B and the 2S4P tend to requires the same number of 'doses'. Crossover designs are preferred for bioequivalence studies because we can derive precise estimates with relatively few subjects. The number of subjects required to assess IBE using Design A far surpasses what is typical for a bioequivalence study, despite it being a dual design. When structure of the summary statistics for designs are closely examined, the deficiencies of Design A become clear. On an elementary level, the information derived from the 2 sequences of Design A is different, so that there are only $n_i - 1$ degrees of freedom (one less the the number of subjects in a particular sequence) available to estimate the variance components $\sigma^2_{I1}, \sigma^2_{I2}, \sigma^2_{WT}$, and σ^2_{WR}. This was noted by Chen *et al.* (2000), who then stated that one should expect to need to enroll more subjects in a study with Design A then in a study with the recommended 4-period design. Now, it is to be expected that a study with fewer periods would require more subjects, but as noted above, the total number of treatments or doses required for adequate power can be much higher in Design A than in the 2S4P design, whereas Design B and the 4-period design are comparable with respect to total doses. On the other hand, the 3 sequences in Design

B all provide the same information, so that there are $n_1 + n_2 + n_3 - 3$ degrees of freedom (three less than the *total* number of subjects) to estimate σ_I^2 and σ_{WR}^2. But as a trade-off, the parameters σ_{WT}^2 and σ_D^2 cannot be estimated using Design B.

The ability to derive the same information from all sequences is a property of sequence-balanced designs. Design B and the 4-period designs in the FDA guidance document are sequence-balanced, whereas Design A is not. Since sequence-balanced designs require fewer subjects (or 'doses', when 3 and 4-period designs are compared) than the Design A, which is not sequence-balanced, and since the sequence-balanced designs are comparable with respect to numbers of 'doses', it would appear that sequence-balance is a desirable property for crossover designs to assess IBE and PBE.

Table 10. Ratio of estimated sample size requirements of Design A to Design B for the test procedures HSCI and GPV for IBE

		$\delta = 0.05$				
		$\sigma_{WT} = 0.15$ $\sigma_{WR} = 0.15$	$\sigma_{WT} = 0.2$ $\sigma_{WR} = 0.2$	$\sigma_{WT} = 0.23$ $\sigma_{WR} = 0.23$	$\sigma_{WT} = 0.3$ $\sigma_{WR} = 0.3$	$\sigma_{WT} = 0.5$ $\sigma_{WR} = 0.5$
Method	σ_D	A/B	A/B	A/B	A/B	A/B
GPV	0.01	1.333	1.417	1.533	1.722	1.879
	0.10	1.556	1.467	1.611	1.744	1.778
	0.15	1.481	1.524	1.608	1.822	1.744
HSCI	0.01	1.200	1.417	1.600	1.556	1.697
	0.10	1.333	1.467	1.590	1.641	1.611
	0.15	1.333	1.422	1.630	1.689	1.590

Table 11. Ratio of estimated sample size requirements of Design A to Design B of the test procedure GPV for PBE

		$\delta = 0.05$			
		$\sigma_{WT} = \sigma_{WR} = 0.15$		$\sigma_{WT} = \sigma_{WR} = 0.23$	
		$\sigma_{BT} = 0.15$ $\sigma_{BR} = 0.15$	$\sigma_{BT} = 0.3$ $\sigma_{BR} = 0.3$	$\sigma_{BT} = 0.23$ $\sigma_{BR} = 0.23$	$\sigma_{BT} = 0.46$ $\sigma_{BR} = 0.46$
Method	ρ	A/B	A/B	A/B	A/B
GPV	0.90	1.200	1.467	1.048	1.467
	0.95	1.200	1.833	1.222	1.833
		$\sigma_{WT} = 0.3, \sigma_{WR} = 0.3$		$\sigma_{WT} = 0.5, \sigma_{WR} = 0.5$	
		$\sigma_{BT} = 0.3$ $\sigma_{BR} = 0.3$	$\sigma_{BT} = 0.6$ $\sigma_{BR} = 0.6$	$\sigma_{BT} = 0.5$ $\sigma_{BR} = 0.5$	$\sigma_{BT} = 1.0$ $\sigma_{BR} = 1.0$
Method	ρ	A/B	A/B	A/B	A/B
GPV	0.90	1.222	1.333	1.111	1.333
	0.95	1.222	1.667	1.222	1.667

Most of the observations made in Chapter 3 for 4-period crossover designs can be applied here to 3-period crossovers, particularly with respect to Design B. For IBE, the number of subjects required for adequate power is roughly the same for HSCI and GPV, as is the case for 4-period crossovers. A problem with the HSCI is that the 95% upper bound is discontinuous at the changeover point from constant to reference-scaling. Chapter 3 gave an example where, when the strict form of the mixed-scaled criteria was applied to the HSCI, whether the estimate of σ_{WR}^2 was greater or less than σ_0^2 effected the decision on IBE. But there was no effect on this decision when the GPV was applied to the same example. What was heuristically demonstrated in the previous chapter is now formalized in the following theorem, which is proved in the appendix.

Theorem. The GPV is continuous for IBE with respect to ss_{WR} at $\nu\sigma_0^2$.

This continuity property of the GPV gives a testing procedure free of the necessity for *ad hoc* decision rules. Since it is not unlikely that an experimenter interested in the IBE of a product would have prior knowledge on the magnitude of σ_{WR}^2, he may specify a preference for the GPV *a priori* when he feels that σ_{WR}^2 will be close to σ_0^2. This not only circumvents the continuity issue, it also happens to be the part of the parameter space where the GPV has a slight power advantage over the HSCI.

For PBE, the GPV is clearly more powerful than the HSCI, as it was for 4-period crossovers. This advantage is manifest in Table 9, where using the HSCI over the GPV

precipitates increases in the required sample size for adequate power to conclude PBE that range from 28% to 100%. The GPV is slightly anti-conservative, especially in Design B, but the worse of this occurs when the mean difference δ is well outside of the acceptable range for average bioequivalence. The GPV seems to increase in power as the ratio of the between-subject to the within-subject variances increase, a circumstance very likely to occur in practice (this is part of the rationale for using crossover studies to assess bioequivalence). While the continuity argument above also appear to apply, is it less likely to be of any impact, since $\sigma^2_{RR} = \sigma^2_{BR} + \sigma^2_{WR}$ is usually much larger than σ^2_0.

The last observation to be made is that although Design B appears to be preferable to Design A for assessing the specific IBE and PBE criteria currently set forth by the FDA, it comes with a cost: the secondary parameters σ^2_D and σ^2_{WT} are not identifiable. Our approach to this problem treats all the variance components and even δ as nuisance parameters. This reflects an attitude that we are only interested in Θ_{IBE} and Θ_{PBE}, at least as long as we are able to conclude IBE and PBE, respectively. But some of our 'nuisance' parameters, or functions of these, have their own intrinsic interpretation. The best example is δ, on which we focus when we assess average bioequivalence, the previous standard for comparing bioavailability. Furthermore, the original discussions of IBE concentrated more on σ^2_D and $r^2 = \sigma^2_{WT}/\sigma^2_{WR}$ (*e.g.*, Ekbohm and Melander (1989)). The former, the subject-formulation interaction, is still considered very important by the FDA, and the general FDA guidance document on bioequivalence studies (2000) requests that sponsors give the point estimate of σ_D in their data summary. To address this issue, we suggest the following two strategies. The first is a sensitivity analysis approach, which we will illustrate with a hypothetical example. Suppose a bioequivalence study with Design B yielded parameter estimates of $\hat{\sigma}^2_I = 0.0835$ and $\hat{\sigma}^2_{WR} = 0.0475$. Table 12 shows the 'estimated' values of $\tilde{\sigma}^2_D$ and $\tilde{\sigma}_D$ when $r^2 = \tilde{\sigma}^2_{WT}/\hat{\sigma}^2_{WR}$ is allowed to vary over a likely range, here from 0.5 to 1.5 (tildes are used instead of hats for quantities that are not directly estimated).

Table 12. Sensitivity Analysis for $\tilde{\sigma}_D^2$ and $\tilde{\sigma}_D$

$r^2 = \tilde{\sigma}_{WT}^2/\hat{\sigma}_{WR}^2$

r^2	$\tilde{\sigma}_D^2$	$\tilde{\sigma}_D$
0.50	0.0360	0.1897
0.75	0.024125	0.1553
1.00	0.012250	0.1107
1.25	0.000375	0.019365
1.50	-0.01150	ND

Here, we have a stark demonstration of the tradeoff between parameters in the aggregate IBE criterion. When r^2 is low, $\tilde{\sigma}_D$ is relatively high, and vice-versa. In the former case, the proportion of subjects whose individual bioequivalence ratios (*i.e.*, μ_{ijT}/μ_{ijR}) are outside of the range $(0.8, 1.25)$ may cause some concern (it is about 14% when $\sigma_D = 0.15$), but the new formulation has considerably less intra-subject variation. In the latter case, a modest increase in intra-subject variation coincides with a negligible level of subject-formulation interaction. In some cases, we might let r^2 grow as large as 2, but it appears to be sensible to stop, as we did here, when the value of $\tilde{\sigma}_D^2$ becomes negative, since the value of $\tilde{\sigma}_D$ is not defined (*n.b.*, negative estimates of σ_D^2 are not unusual in IBE when the method of moments is used.)

The second approach to the non-identifiability of σ_D is to take $\hat{\sigma}_I^2 - 0.5\hat{\sigma}_{WR}^2$, which estimates $\sigma_D^2 + \sigma_{WT}^2$, as an upper bound of σ_D^2, which may in itself be sufficient to allay regulatory concerns about subject-formulation interaction. Using the methylphenidate meta-data as an example, we computed this upper bound on σ_D to be 0.0630. This suggests that proportion of subjects whose individual bioequivalence ratios are outside of the range $(0.8, 1.25)$ is very unlikely to be greater than 0.1%. In this case, the first approach does not yield much more information, since, even with r^2 as low as 0.1, the derived value of $\tilde{\sigma}_D^2$ is negative.

In conclusion, we hope that this investigation has answered some critical question concerning the use of 3-period crossover designs in assessing bioequivalence. The poor performance of Design A should preclude it from consideration as a viable method for

assessing IBE and PBE. Design B provides a highly efficient method for assessing PBE and IBE, but cannot be used to estimate secondary parameters that may be of interest to concerned parties. We recommend either a sensitivity analysis or an upper bound assessment to account for the lack of explicit estimates of σ_D and σ^2_{WT}. It is also recommended that the sponsor considering the merits of Designs B versus a 4-period crossover design may want to consult with the FDA before implementing this design in their study.

4.6 Appendix

4.6.1 Proof of Proposition

Proposition. The method of moment estimators of the identifiable parameters in Design B are the same as the unconstrained REML estimates for these parameters.

First we prove the proposition for the fixed parameters, μ_T, μ_R, the γ_{ikl}'s, and linear functions of these parameters, such as $\delta = \mu_T - \mu_R$. Consider the following linear model

$$E(\boldsymbol{y}) = \boldsymbol{X}\boldsymbol{\beta} \qquad cov(\boldsymbol{y}) = \boldsymbol{V}.$$

We will show that $\boldsymbol{H}\boldsymbol{V} = \boldsymbol{V}\boldsymbol{H}$, where $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T$ is the *projection* or *hat* matrix (we can also define $\boldsymbol{H}$ using a generalized inverse for $\boldsymbol{X}^T\boldsymbol{X}$). Puntanen and Styan (1989) call this *Zyskind's condition*. First, we use the estimability conditions from Section 4.2.1 to eliminate γ_{3T1} and γ_{3R2}, *i.e.*, we define γ_{3T1} and γ_{3R2} as follows

$$\gamma_{3T1} = -\gamma_{1T1} - \gamma_{2T1}$$

and

$$\gamma_{3R2} = -\gamma_{1R1} - \gamma_{2R1} - \gamma_{1R2} - \gamma_{2R2} - \gamma_{3R1}.$$

The design matrix $\boldsymbol{X}$ is

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_1 \otimes \mathbf{1} \\ \boldsymbol{X}_1 \otimes \mathbf{1}_{n_1} \\ \boldsymbol{X}_2 \otimes \mathbf{1}_{n_2} \\ \boldsymbol{X}_3 \otimes \mathbf{1}_{n_3} \end{pmatrix}$$

where $\mathbf{1}_{n_i}$ is a vector of ones and the $\boldsymbol{X}_i$'s are

$$\boldsymbol{X}_1 = \begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}$$

$$\boldsymbol{X}_2 = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

and

$$\boldsymbol{X}_3 = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & -1 & -1 & -1 & -1 & -1 \\ 1 & 0 & -1 & -1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Now the covariance matrix for Design B is

$$\boldsymbol{V} = \begin{pmatrix} \boldsymbol{V}_1 \otimes \boldsymbol{I}_{n_1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{V}_2 \otimes \boldsymbol{I}_{n_2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{V}_3 \otimes \boldsymbol{I}_{n_3} \end{pmatrix}$$

where

$$\boldsymbol{V}_1 = \begin{pmatrix} \sigma^2_{BT} + \sigma^2_{WT} & \sigma_{TR} & \sigma_{TR} \\ \sigma_{TR} & \sigma^2_{BR} + \sigma^2_{WR} & \sigma^2_{BR} \\ \sigma_{TR} & \sigma^2_{BR} & \sigma^2_{BR} + \sigma^2_{WR} \end{pmatrix}$$

$$\boldsymbol{V}_2 = \begin{pmatrix} \sigma^2_{BR} + \sigma^2_{WR} & \sigma_{TR} & \sigma^2_{BR} \\ \sigma_{TR} & \sigma^2_{BT} + \sigma^2_{WT} & \sigma_{TR} \\ \sigma^2_{BR} & \sigma_{TR} & \sigma^2_{BR} + \sigma^2_{WR} \end{pmatrix}$$

$$\boldsymbol{V}_3 = \begin{pmatrix} \sigma^2_{BR} + \sigma^2_{WR} & \sigma^2_{BR} & \sigma_{TR} \\ \sigma^2_{BR} & \sigma^2_{BR} + \sigma^2_{WR} & \sigma_{TR} \\ \sigma_{TR} & \sigma_{TR} & \sigma^2_{BT} + \sigma^2_{WT} \end{pmatrix}$$

are covariance matrices for the respective sequences, the $\boldsymbol{I}_{n_i}$ are identity matrices and $\mathbf{0}$ is a matrix of zeroes. The hat matrix is

$$\boldsymbol{H} = \begin{pmatrix} \frac{1}{n_1}\boldsymbol{I}_3 \otimes \boldsymbol{J}_{n_1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \frac{1}{n_2}\boldsymbol{I}_3 \otimes \boldsymbol{J}_{n_2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \frac{1}{n_3}\boldsymbol{I}_3 \otimes \boldsymbol{J}_{n_3} \end{pmatrix}$$

where $\boldsymbol{J}_{n_i}$ is a $n_i \times n_i$ matrix of 1's. It is now clear that $\boldsymbol{HV} = \boldsymbol{VH}$, since $\boldsymbol{V}$, has an identity matrix on the right side of the Kronecker product and $\boldsymbol{H}$, has an identity matrix on the left side of the Kronecker product. Therefore, the REML estimates of the fixed effects equals the method of moments (OLS) estimates for Design B.

To prove the proposition for estimable functions of the variance components (*i.e.*, $\sigma^2_T, \sigma^2_R, \sigma_{TR}, \sigma^2_{WR}$, and σ^2_I), we slightly modify the argument of Chinchilli and Esinhart (C&E, 1996). The key condition here is that the covariance matrices for each sequence $\boldsymbol{\Lambda}_i = cov(\boldsymbol{Y}_{ij})$ are equal for all three sequences, where $\boldsymbol{Y}_{ij} = (Y_{ijT1}\ Y_{ijR1}\ Y_{ijR2})^T$ is the vector of the individual responses. Since

$$\boldsymbol{\Lambda}_i = \begin{pmatrix} \sigma^2_{BT} + \sigma^2_{WT} & \sigma_{TR} & \sigma_{TR} \\ \sigma_{TR} & \sigma^2_{BR} + \sigma^2_{WR} & \sigma^2_{BR} \\ \sigma_{TR} & \sigma^2_{BR} & \sigma^2_{BR} + \sigma^2_{WR} \end{pmatrix}$$

for $i = 1, 2, 3$, Design B meets this condition (this actually follows from the sequence-balance property discussed above), so that we can apply C&E's subsequence argument. Now consider the following covariance matrix for the statistics defined in Section 3.5.

$$var \begin{bmatrix} \overline{Y}_{ijT} \\ \overline{Y}_{ijR} \\ E_{ijR} \end{bmatrix} = \begin{pmatrix} \sigma^2_{BT} + \sigma^2_{WT} & \sigma_{TR} & 0 \\ \sigma_{TR} & \sigma^2_{BR} + \frac{1}{2}\sigma^2_{WR} & 0 \\ 0 & 0 & \sigma^2_{WR} \end{pmatrix}.$$

We now have a covariance matrix in the same form as C&E equation (21), and so we follow their argument, constructing the usual sums of squares cross products matrix (C&E equation (23)), from which we apply C&E equation (25) to get the REML estimates of

σ_T^2, σ_R^2, and σ_{TR}. The REML estimate of σ_I^2 is a linear function of these estimates, and the REML estimate of σ_{WR}^2 is derived from C&E equation (26). Therefore, the REML estimators that we need for assessing Θ_{IBE} and Θ_{PBE} are equal to the method of moments estimators.

4.6.2 Proof of Theorem

We here prove the following theorem, which is similar to the one in Section 5.

Theorem. For the IBE parameter in Design B, GPV is continuous with respect to ss_{WR} at $\nu\sigma_0^2$.

First, consider the following generalized test variable (GTV)

$$T_*(X; x, \xi) = \frac{\theta_I max\left(\frac{ss_{WR}}{U_{WR}}, \sigma_0^2\right) + \frac{3}{2}\frac{ss_{WR}}{U_{WR}}}{\left(d - cZ_D\sqrt{\frac{ss_I}{U_I}}\right)^2 + \frac{ss_I}{U_I}}.$$

Note that $T_*(X; x, \xi)$ is a one-to-one transformation of the GTV is Section 3.5, and, along with the generalized extreme region

$$C_x = \{X : T_*(X; x, \xi) < 1\},$$

gives the same GPV value as the procedure in Section 3.5. Note the T_* is non-negative We now define the following functions

$$f(Z_D, U_I, U_{WR}, ss_{WR}) = T_*(X; x, \xi) \times 1_{\{T_* < 1\}}(Z_D, U_I, U_{WR}, ss_{WR})$$

$$\phi(ss_{WR}) = \int f(Z_D, U_I, U_{WR}, ss_{WR}) dF(Z_D, U_I, U_{WR})$$

where $1_S(X)$ is the indicator function, which take the value 1 when $X \in S$ and 0 otherwise, and $dF(*)$ is the probability measure with respect to the random variables in T_*. Here $f(Z_D, U_I, U_{WR}, ss_{WR})$ is a measurable and integrable function of (Z_D, U_I, U_{WR}) for each ss_{WR} in the interval (a, b) that contains $\nu\sigma_0^2$. Note that the definition of $\phi(ss_{WR})$ is equivalent to our definition of the GPV.

We now assert that the following conditions hold

1. For $(Z_D, U_I, U_{WR}) \in A$, where A satisfies condition (16.9) of Billingsley (1995, pg. 211), $f(Z_D, U_I, U_{WR}, ss_{WR})$ is continuous in ss_{WR} at $\nu\sigma_0^2$.

2. There exist δ and $g(*)$ such that $|f(Z_D, U_I, U_{WR}, ss_{WR})| \leq g(Z_D, U_I, U_{WR}) = 1$ for $(Z_D, U_I, U_{WR}) \in A$ and $|ss_{WR} - \nu\sigma_0^2| < \delta$, where δ is independent of (Z_D, U_I, U_{WR}) and $g(Z_D, U_I, U_{WR}) = 1$ is integrable.

The continuity of $\phi(ss_{WR})$ at $\nu\sigma_0^2$ then follows from Theorem 16.8(i) of Billingsley (pg. 212). Since $\phi(ss_{WR})$ is the GPV, the theorem above immediately follows.

Similar proofs can be applied to GPV tests of IBE for both Design A and the 4-period crossover design in Chapter 3.

Chapter 5

A TEST FOR POPULATION BIOEQUIVALENCE FOR PHARMACOKINETIC DATA FROM 2×2 CROSSOVER DESIGNS

SUMMARY

We propose a test for assessing population bioequivalence in 2×2 crossover designs based on the *generalized p-value* (GPV). We find that the GPV test for 2×2 crossovers is more powerful than the FDA test, yielding adequate power in cases where the FDA may have a false negative rate exceeding 50%.

5.1 Introduction

Individual bioequivalence is important in the assessment of new generic versions of drug products already on the market. The most likely application of PBE involves newly approved products (Chow, 1999), where it may be necessary to show that the drug product from large-scale batches intended for commercial distribution are equivalent to the product from the small-scale batches used in clinical trials. Here it is assumed that the vast majority of the potential patient population has yet to be exposed to this product.

As we have seen in previous chapters, the assessment of individual bioequivalence requires that treatments be replicated within subjects so that the bioequivalence criterion can be estimated. The guidance document is more ambiguous on the study design when the goal is to assess PBE. Appendix F to the FDA guidance document, which details how to conduct the test for PBE, assumes the use of the same 2×4 crossover design that is recommended for assessing individual bioequivalence. But the 2×2 crossover design that has long been the standard for evaluating average bioequivalence can also be used to assess PBE. It is straightforward to adapt the test in the guidance document to the 2×2 crossover design, but this test is very conservative for PBE, regardless of the study design. This chapter proposes an alternative test for assessing PBE for 2×2 crossover studies based on the *generalized p-value* (GPV) similar to the ones we proposed in Chapter 3 for 4-period crossover designs and Chapter 4 for 3-period crossover designs. We will also compare this test directly to the FDA test methodology.

In section 5.2, we define the model, summary statistics, and other details that we will need to construct tests for PBE. Section 5.3 contains the derivation of the GPV and FDA tests, and provides an example with actual bioequivalence data. Section 5.4 uses simulation to evaluate and to compare the two tests for size and power. Section 5.5 contains a short discussion and our conclusions.

5.2 Preliminaries

5.2.1 The Statistical Model

The linear model that the FDA has adopted for assessing individual and population bioequivalence (FDA 2001) is different from the sequence-period-treatment model used for assessing average bioequivalence (*i.e.*, the equivalence of the treatment means of two formulations) recommended by the previous statistical guidance document on bioequivalence (FDA, 1992). The observational model, taken from Chinchilli and Esinhart (1996), is given below.

$$Y_{ijkl} = \mu_k + \gamma_{ikl} + \eta_{ijk} + \epsilon_{ijkl} \quad i = 1, \ldots, s; \quad j = 1, \ldots, n_i; \quad k = T, R; \quad l = 1, \ldots, q;$$

where Y_{ijkl} is the response from subject j in sequence i to the l^{th} application of treatment k, and s is the number of sequences. The parameters μ_T and μ_R are the population means for the test (T) and the reference (R) treatments, and γ_{ikl} is the fixed effect corresponding to the l^{th} application of treatment k in sequence i. For the 2×2 crossover study design we have $s = 2$ and $q = 1$. Following Chinchilli and Esinhart, we impose the estimability conditions $\gamma_{1k1} = -\gamma_{2k1}, k = T, R$ on the γ_{ikl}'s. Note that the sequence and period effects that are often used to model data from 2×2 crossover studies are linear combinations of the above γ's, *i.e.*, the sequence effect is $\gamma_{1T1} + \gamma_{1R1}$ and the period effect is $\gamma_{1T1} - \gamma_{1R1}$.

Since $l = 1$ for all observations, the random between-subject (η_{ijk}) and within-subject (ϵ_{ijkl}) effects are confounded with each other. This is not a problem for PBE, since the criteria we will be using only requires the total variance (from the random effects η_{ijk} and ϵ_{ijkl} added together) for each formulation. So we then write the linear model as follows

$$Y_{ijk} = \mu_k + \gamma_{ik} + \epsilon_{ijk} \quad i = 1, \ldots, 2: \quad j = 1, \ldots, n_i; \quad k = T, R;$$

Now it only remains to specify the distribution of the vectors $\epsilon_{ij} = [\epsilon_{ijT}, \epsilon_{ijR}]^t$ for each subject: they are mutually independent bivariate normal with zero means and a variance matrix given by

$$\boldsymbol{V} = \begin{pmatrix} \sigma_T^2 & \sigma_{TR} \\ \sigma_{TR} & \sigma_R^2 \end{pmatrix}.$$

For convenience, we define $\delta = \mu_T - \mu_R$, since the fixed effect of primary interest is the treatment difference. We also need the variance of the estimate of the treatment difference, which is $\sigma_I^2 = \sigma_T^2 + \sigma_R^2 - 2\sigma_{TR}$.

The evaluation of PBE is based on the following parameter of interest

$$\Theta_{PBE} = \begin{cases} \dfrac{\delta^2 + \sigma_T^2 - \sigma_R^2}{\sigma_R^2} & \text{when } \sigma_R^2 > \sigma_0^2 \\ & \text{(reference-scaled criterion)} \\ \dfrac{\delta^2 + \sigma_T^2 - \sigma_R^2}{\sigma_0^2} & \text{when } \sigma_R^2 \le \sigma_0^2 \\ & \text{(constant-scaled criterion)} \end{cases}$$

This parameter has been discussed previously. Since σ_T^2 and σ_R^2 are estimated directly, reparameterization of Θ_{PBE} is not an issue here as it was in Chapters 3 and 4.

5.2.2 Summary Statistics

Consider a bioequivalence study with a 2×2 crossover design, where each subject is treated with both the test and reference formulations. As before, let Y_{ijk} denote the response of subject j in sequence i to treatment k ($k = T, R$). In the statistical model of Section 2.1, the parameters that are needed to compute the PBE criteria are μ_T, μ_R, σ_T^2, σ_R^2, and σ_{TR}. We now define the following quantities.

$$D_{ij} = Y_{ijT} - Y_{ijR} \qquad \overline{D}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} D_{ij} \qquad D = \frac{\overline{D}_1 + \overline{D}_2}{2}$$

$$\overline{Y}_{i\bullet k} = \frac{1}{n_i}\sum_{j=1}^{n_i} Y_{ijk} \qquad SS_k = \sum_{i=1}^{2}\sum_{j=1}^{n_i}(Y_{ijk} - \overline{Y}_{i\bullet k})^2 \quad k = T, R$$

$$SS_{TR} = \sum_{i=1}^{2}\sum_{j=1}^{n_i}(Y_{ijT} - \overline{Y}_{i\bullet T})(Y_{ijR} - \overline{Y}_{i\bullet R}) \quad B = \frac{SS_{TR}}{SS_T}$$

$$SS_{R|T} = SS_R - \frac{[SS_{TR}]^2}{SS_T} \quad SS_I = \sum_{i=1}^{2}\sum_{j=1}^{n_i}(D_{ij} - \overline{D}_i)^2$$

$$\delta = \mu_T - \mu_R \qquad \sigma_I^2 = \sigma_T^2 + \sigma_R^2 - 2\sigma_{TR}$$

$$\beta = \frac{\sigma_{TR}}{\sigma_T^2} \qquad \sigma_{R|T}^2 = \sigma_R^2 - \frac{\sigma_{TR}^2}{\sigma_T^2} = \sigma_R^2 - \beta^2\sigma_T^2$$

$$c = \frac{1}{4}\left(\frac{1}{n_1} + \frac{1}{n_2}\right) \qquad \nu = n_1 + n_2 - 2$$

It is easily seen that the statistics $D, SS_T, SS_R, SS_{R|T}$, and B are associated with the following pivotal quantities with known distributions as indicated.

$$Z_D = \frac{D-\delta}{\sqrt{c\sigma_I^2}} \sim N(0,1) \qquad U_T = \frac{SS_T}{\sigma_T^2} \sim \chi^2_\nu \qquad U_R = \frac{SS_R}{\sigma_R^2} \sim \chi^2_\nu$$

$$U_{R|T} = \frac{SS_{R|T}}{\sigma^2_{T|R}} \sim \chi^2_{\nu-1} \qquad Z_B = (B-\beta)\sqrt{\frac{SS_T}{\sigma^2_{R|T}}} \sim N(0,1)$$

The collection of random variables $Z_D, U_T, U_{R|T}, Z_B$ are mutually independent. Note that, the quantities B and $SS_{R|T}$ are, respectively, the linear regression coefficient and the residual sum of squares from the regression of Y_{ijR} on Y_{ijT}.

5.3 Generalized Test Variable and GPV for PBE

We first define the following: let $\boldsymbol{X} = (D, B, SS_T, SS_{R|T})$ be a vector of random quantities, let $\boldsymbol{x} = (d, b, ss_T, ss_{R|T})$ be the observed values of $\boldsymbol{X}$, and let $\boldsymbol{\xi} = (\Theta_{PBE}, \boldsymbol{\zeta})$, where Θ_{PBE} is our parameter of interest and $\boldsymbol{\zeta} = (\delta, \beta, \sigma_T^2, \sigma^2_{R|T})$ is a vector of nuisance parameters. The generalized test variable (GTV) for the PBE criteria is

$$T(\boldsymbol{X}; \boldsymbol{x}, \xi) = \frac{\left(d - Z_D\sqrt{cT_D}\right)^2 + \frac{ss_T}{U_T} - T_R}{max(T_R, \sigma_0^2)}$$

where

$$T_D = \frac{ss_T}{U_T}\left(1 - \left(b - Z_B\sqrt{\frac{ss_{R|T}}{ss_T U_{R|T}}}\right)\right)^2 + \frac{ss_{R|T}}{U_{R|T}}$$

and

$$T_R = \frac{ss_T}{U_T}\left(b - Z_B\sqrt{\frac{ss_{R|T}}{ss_T U_{R|T}}}\right)^2 + \frac{ss_{R|T}}{U_{R|T}}.$$

The observed values $d, ss_T, ss_{R|T}, b$ of $D, SS_T, SS_{R|T}, B$, respectively, are regarded as fixed constants. Note that $T(\boldsymbol{X}; \boldsymbol{x}, \xi)$ is distributed free of nuisance parameters, and that $T(\boldsymbol{x}; \boldsymbol{x}, \xi) = \Theta_{PBE}$. Also note that $T(\boldsymbol{X}; \boldsymbol{x}, \xi)$ increases as Θ_{PBE} increases. Therefore, the test function $T(\boldsymbol{X}; \boldsymbol{x}, \xi)$ we have defined has the required properties set forth by Tsui and Weerahandi (1989). The generalized p-value for testing

$$H_0 : \Theta_{PBE} \geq \theta_P \quad \text{versus} \quad H_a : \Theta_{PBE} < \theta_P$$

is given by $Pr(T(\boldsymbol{X};\boldsymbol{x},\xi) \geq \theta_P \mid \Theta_{PBE} = \theta_P)$. This probability can be easily computed using Monte Carlo integration.

In the example and simulation studies, we will compare this GPV test with the test recommended in the FDA guidance document (2001). This test is adapted from the test for individual bioequivalence derived by Hyslop, Hsuan, and Holder (2000). Since it is based on a confidence interval construct according to the methodology of Howe (1974), we call it the Howe-type small sample confidence interval (HSCI). Note that Howe developed his method for a linear combination of independent random variables. Since the derivation of the test for PBE in the FDA guidance document is for data from a four-period design, we must adapt the HSCI test to a 2-period design. Using the same notation as above (*i.e.*, lower case for the observed values of random quantities), the FDA (HSCI) 95% upper bound for the reference-scaled criterion is

$$d^2 + \frac{ss_T}{\nu} - (1+\theta_P)\frac{ss_R}{\nu} + \sqrt{H_D^2 + H_T^2 + H_{RR}^2}.$$

For the constant-scaled criterion the HSCI upper bound is

$$d^2 + \frac{ss_T}{\nu} - \frac{ss_R}{\nu} - \sigma_0^2\theta_P + \sqrt{H_D^2 + H_T^2 + H_{RC}^2}.$$

The quantities in the variance part of these expressions are

$$H_D = \left(|d| + t_{\nu,0.05}\sqrt{cs_I^2}\right)^2 - d^2 \qquad H_T = ss_T\left(\frac{1}{\chi^2_{\nu,0.05}} - \frac{1}{\nu}\right)$$

$$H_{RR} = ss_R(1+\theta_P)\left(\frac{1}{\chi^2_{\nu,0.95}} - \frac{1}{\nu}\right) \qquad H_{RC} = ss_R\left(\frac{1}{\chi^2_{\nu,0.95}} - \frac{1}{\nu}\right)$$

where $t_{\nu,0.05}$ and $\chi^2_{\nu,\gamma}$ are the tabled values of the t and χ^2 distributions, respectively, and s_I^2 is the sample variance of the estimate of the treatment difference δ computed in the usual way. Because of the linearization of the test criteria, the HSCI test concludes PBE if the calculated upper bound is less than 0. See Hyslop *et al.* (2000) or the FDA guidance document (2001) for details on the linearized test criteria and how the hypothesis test is modified.

5.3.1 Example

Chow and Liu (2000, pp. 71 *ff*) give data from a bioequivalence study with a 2 × 2 crossover design. Twenty-four (24) subjects were given 250 mg of an unspecified drug either in the form of five 50 mg tablets (T) or a 5mL oral suspension (R). The pharmacokinetic parameter under consideration is AUC_{0-32}, the area under the concentration-time curve from treatment adminstration to Hour 32. The following summary statistics were computed using the method of moments:

$$d = -0.02865 \quad ss_T = 1.67730 \quad ss_R = 1.66503$$
$$b = 0.50815 \quad ss_{R|T} = 1.23191 \quad s_I^2 = 0.074441 \quad \nu = 22.$$

Since $ss_R/\nu > 0.04$, we use the reference-scaled criterion. The linearized criterion (Hyslop *et al.*, 2000) has a point estimate of -0.13067, and the HSCI 95% upper bound is -0.03634, which is less than 0, so we conclude PBE. The point estimate of Θ_{PBE} is 0.01827, and the GPV, based on 50000 Monte Carlo points, is $p = 0.00376$, which is less than $\alpha = 0.05$, so we again conclude PBE. Here the GPV and HSCI tests agree in their conclusions, but we have previously seen examples where the GPV test concluded PBE, while the HSCI test did not. Given that out simulations show the HSCI to be very conservative, there are some concerns about the false negative rate of this test. This issue is addressed further below.

5.4 Simulation Study

A simulation study was conducted to compare the GPV method to the FDA method for population bioequivalence. The methods were examined for their ability to maintain the nominal type-I error rate and for their power to reject the null hypothesis of non-bioequivalence when the alternative hypothesis of bioequivalence is actually true. To estimate the sizes of the two tests, we chose the values of the nuisance parameters so that $\theta_P = 1.7448$, the boundary of the null hypothesis. For evaluating the power of the test, simulation parameters were chosen based on the power tables given in the appendix of the FDA guidance document (2001). The value of $\sigma_0^2 = 0.04$ was used for changeover point

from the constant to the reference-scaled criterion. The quantity n in the tables represent the number of subjects per sequence. The covariance parameters in these simulations were defined terms of the parameters from the original Chinchilli and Esinhart model (1996), $\sigma^2_{BT}, \sigma^2_{BR}, \sigma^2_{WT}, \sigma^2_{WR}$, and $\rho\sigma_{BT}\sigma_{BR}$, where a subscripted B signifies a between-subject effect arising from η_{ijk}, a subscripted W signifies a within-subject effect arising from ϵ_{ijkl}, and ρ is the correlation coefficient of the relationship between η_{ijT} and η_{ijR} (in our model, ρ is the correlation between ϵ_{ijT} and ϵ_{ijR}. In the notation of this paper, $\sigma^2_T = \sigma^2_{BT} + \sigma^2_{WT}$, $\sigma^2_R = \sigma^2_{BR} + \sigma^2_{WR}$, and $\sigma_{TR} = \rho\sigma_{BT}\sigma_{BR}$. This notational convention should facilitate comparisons to results from investigation examining PBE tests in other designs, such as Chapter 3, although it causes some cases to be repeated in Table 1a. Five thousand (5000) simulated data sets were generated for each parameter combination. GPVs were estimated based on 10000 Monte Carlo runs. All tests were done at the nominal $\alpha = 0.05$ level. All simulations were done using SAS® Version 8.2 for Windows. The results are summarized below.

Tables 1a and 1b show simulation results comparing the respective sizes of the two tests. Here, correlations of $\rho = 0.8$ and $\rho = 0.9$ are examined. The reasoning for examining cases with such seemingly high correlation is that, the responses within a subject to different formulations of the same drug are still likely to have a strong degree of correlation even if the formulations are only borderline bioequivalent.

Table 1a. Estimated sizes of the test procedures HSCI and GPV for PBE for 2 sequence/2 period designs ($\delta \neq 0$)

		$\delta = 0.2642$		$\delta = 0.2954$		$\delta = 0.3736$	
		$\sigma_{BT} = 0.1, \sigma_{BR} = 0.1$		$\sigma_{BT} = 0.2, \sigma_{BR} = 0.2$		$\sigma_{BT} = 0.2, \sigma_{BR} = 0.2$	
		$\sigma_{WT} = 0.1, \sigma_{WR} = 0.1$		$\sigma_{WT} = 0.1, \sigma_{WR} = 0.1$		$\sigma_{WT} = 0.2, \sigma_{WR} = 0.2$	
n	ρ	HSCI	GPV	HSCI	GPV	HSCI	GPV
8	0.8	0.0338	0.0552	0.0208	0.0730	0.0306	0.0680
	0.9	0.0292	0.0514	0.0148	0.0688	0.0280	0.0668
12	0.8	0.0342	0.0498	0.0234	0.0660	0.0378	0.0610
	0.9	0.0398	0.0554	0.0198	0.0696	0.0316	0.0610
16	0.8	0.0450	0.0554	0.0244	0.0700	0.0330	0.0604
	0.9	0.0370	0.0504	0.0176	0.0608	0.0324	0.0568
		$\delta = 0.5907$		$\delta = 0.5604$		$\delta = 0.8861$	
		$\sigma_{BT} = 0.4, \sigma_{BR} = 0.4$		$\sigma_{BT} = 0.3, \sigma_{BR} = 0.3$		$\sigma_{BT} = 0.6, \sigma_{BR} = 0.6$	
		$\sigma_{WT} = 0.2, \sigma_{WR} = 0.2$		$\sigma_{WT} = 0.3, \sigma_{WR} = 0.3$		$\sigma_{WT} = 0.3, \sigma_{WR} = 0.3$	
n	ρ	HSCI	GPV	HSCI	GPV	HSCI	GPV
8	0.8	0.0200	0.0658	0.0308	0.0736	0.0148	0.0612
	0.9	0.0114	0.0614	0.0274	0.0652	0.0086	0.0582
12	0.8	0.0220	0.0612	0.0340	0.0648	0.0214	0.0548
	0.9	0.0162	0.0550	0.0322	0.0652	0.0192	0.0606
16	0.8	0.0240	0.0582	0.0372	0.0648	0.0224	0.0596
	0.9	0.0168	0.0572	0.0334	0.0610	0.0214	0.0642
		$\delta = 0.2954$				$\delta = 0.5907$	
		$\sigma_{BT} = 0.2, \sigma_{BR} = 0.1732$		$\sigma_{BT} = 0.1732, \sigma_{BR} = 0.2$		$\sigma_{BT} = 0.3873, \sigma_{BR} = 0.4$	
		$\sigma_{WT} = 0.1, \sigma_{WR} = 0.1414$		$\sigma_{WT} = 0.1414, \sigma_{WR} = 0.1$		$\sigma_{WT} = 0.2236, \sigma_{WR} = 0.2$	
n	ρ	HSCI	GPV	HSCI	GPV	HSCI	GPV
8	0.8	0.0228	0.0718	0.0226	0.0716	0.0188	0.0620
	0.9	0.0200	0.0694	0.0194	0.0702	0.0150	0.0610
12	0.8	0.0258	0.0668	0.0264	0.0674	0.0222	0.0556
	0.9	0.0290	0.0768	0.0250	0.0648	0.0166	0.0566
16	0.8	0.0276	0.0646	0.0318	0.0648	0.0218	0.0546
	0.9	0.0282	0.0706	0.0270	0.0644	0.0198	0.0528
		$\delta = 0.5907$		$\delta = 0.8861$			
		$\sigma_{BT} = 0.4, \sigma_{BR} = 0.3873$		$\sigma_{BT} = 0.6, \sigma_{BR} = 0.5$		$\sigma_{BT} = 0.5, \sigma_{BR} = 0.6$	
		$\sigma_{WT} = 0.2, \sigma_{WR} = 0.2236$		$\sigma_{WT} = 0.3, \sigma_{WR} = 0.4472$		$\sigma_{WT} = 0.4472, \sigma_{WR} = 0.3$	
n	ρ	HSCI	GPV	HSCI	GPV	HSCI	GPV
8	0.8	0.0178	0.0562	0.0196	0.0628	0.0188	0.0534
	0.9	0.0164	0.0580	0.0150	0.0610	0.0268	0.0578
12	0.8	0.0258	0.0628	0.0260	0.0560	0.0296	0.0582
	0.9	0.0186	0.0590	0.0286	0.0664	0.0200	0.0622
16	0.8	0.0296	0.0632	0.0312	0.0602	0.0260	0.0640
	0.9	0.0226	0.0550	0.0330	0.0632	0.0298	0.0620

Table 1b. Estimated sizes of the test procedures HSCI and GPV for 2 sequence/2 period design ($\delta = 0$) for PBE

		$\sigma_{BT} = 0.2825, \sigma_{BR} = 0.1$ $\sigma_{WT} = 0.1, \sigma_{WR} = 0.1$		$\sigma_{BT} = 0.3567, \sigma_{BR} = 0.2$ $\sigma_{WT} = 0.1, \sigma_{WR} = 0.1$		$\sigma_{BT} = 0.4238, \sigma_{BR} = 0.2$ $\sigma_{WT} = 0.2, \sigma_{WR} = 0.2$	
n	ρ	HSCI	GPV	HSCI	GPV	HSCI	GPV
8	0.8	0.0312	0.0482	0.0160	0.0688	0.0194	0.0534
	0.9	0.0290	0.0436	0.0098	0.0714	0.0180	0.0618
12	0.8	0.0332	0.0432	0.0112	0.0686	0.0182	0.0470
	0.9	0.0380	0.0498	0.0074	0.0664	0.0184	0.0542
16	0.8	0.0338	0.0424	0.0154	0.0608	0.0254	0.0548
	0.9	0.0318	0.0426	0.0082	0.0654	0.0196	0.0528
		$\sigma_{BT} = 0.7134, \sigma_{BR} = 0.4$ $\sigma_{WT} = 0.2, \sigma_{WR} = 0.2$		$\sigma_{BT} = 0.6357, \sigma_{BR} = 0.3$ $\sigma_{WT} = 0.3, \sigma_{WR} = 0.3$		$\sigma_{BT} = 1.0701, \sigma_{BR} = 0.6$ $\sigma_{WT} = 0.3, \sigma_{WR} = 0.3$	
n	ρ	HSCI	GPV	HSCI	GPV	HSCI	GPV
8	0.8	0.0082	0.0506	0.0230	0.0592	0.0078	0.0548
	0.9	0.0058	0.0580	0.0142	0.0526	0.0052	0.0522
12	0.8	0.0110	0.0506	0.0226	0.0558	0.0066	0.0458
	0.9	0.0042	0.0478	0.0174	0.0476	0.0032	0.0502
16	0.8	0.0106	0.0508	0.0198	0.0508	0.0092	0.0484
	0.9	0.0038	0.0516	0.0164	0.0544	0.0052	0.0490

In most cases, the HSCI is far more conservative than GPV. The GPV method shows a tendency to be somewhat anti-conservative, but none of observed test-levels exceeds 0.08. The GPV appears to be most anti-conservative under the following two circumstances: (1) when δ exceeds $log(1.25)$, the limit for average bioequivalence, and (2) when $\sigma_R^2 = \sigma_{BR}^2 + \sigma_{WR}^2$ is close to $\sigma_0^2 = 0.04$. In regard to the first case, the FDA guidance document requires some form of average bioequivalence in order to conclude PBE, so the slight anti-conservatism of the GPV when $|\delta| > log(1.25)$ is not likely to have much of a practical impact. As for the second case, similar behavior was observed in the previous chapters for the GPV when testing both individual and population bioequivalence for 4-period crossover designs. The levels for individual bioequivalence, though slightly inflated, were similar to the levels for the HSCI when the 'play-the-winner' rule (FDA, 2001) was applied. As for the HSCI, the observed sizes are quite low for a test that is advertised as having a level of $\alpha = 0.05$, with the estimated size of the test going below $\alpha = 0.01$ (*i.e.*, 5 times smaller than the nominal level) in the most extreme cases. Given how conservative this test appears to be, we expect its power to be adversely effected.

Table 2 summarizes the simulation results for the power of the two tests. It shows the smallest n for which a particular test achieves 80% power and the actual estimate of the

power for that test at that point. Here correlations of $\rho = 0.9$ and $\rho = 0.95$ are examined. These values were deemed reasonable for examining power because formulations that are bioequivalent should exhibit a very high degree of correlation.

Table 2. Sample size comparisons and estimated power for PBE of the test procedures HSCI and GPV for 2 sequence/2 period designs

		$\delta = 0.05$							
		$\sigma_{WT} = 0.15, \sigma_{WR} = 0.15$				$\sigma_{WT} = 0.23, \sigma_{WR} = 0.23$			
		$\sigma_{BT} = 0.15$ $\sigma_{BR} = 0.15$		$\sigma_{BT} = 0.3$ $\sigma_{BR} = 0.3$		$\sigma_{BT} = 0.23$ $\sigma_{BR} = 0.23$		$\sigma_{BT} = 0.46$ $\sigma_{BR} = 0.46$	
Method	ρ	n	Power	n	Power	n	Power	n	Power
HSCI	0.90	12	0.826	13	0.836	15	0.819	12	0.808
	0.95	11	0.802	12	0.805	14	0.801	12	0.815
GPV	0.90	9	0.837	8	0.837	12	0.823	8	0.832
	0.95	8	0.807	7	0.826	12	0.819	7	0.819
		$\sigma_{WT} = 0.3, \sigma_{WR} = 0.3$				$\sigma_{WT} = 0.5, \sigma_{WR} = 0.5$			
		$\sigma_{BT} = 0.3$ $\sigma_{BR} = 0.3$		$\sigma_{BT} = 0.6$ $\sigma_{BR} = 0.6$		$\sigma_{BT} = 0.5$ $\sigma_{BR} = 0.5$		$\sigma_{BT} = 1.0$ $\sigma_{BR} = 1.0$	
Method	ρ	n	Power	n	Power	n	Power	n	Power
HSCI	0.90	15	0.823	12	0.805	14	0.806	12	0.815
	0.95	15	0.839	12	0.831	15	0.844	12	0.825
GPV	0.90	12	0.823	8	0.830	11	0.801	8	0.834
	0.95	12	0.828	7	0.819	11	0.805	7	0.813

Table 2 shows how that GPV has smaller sample size requirements than the HSCI across the board. The increase in the sample size requirement for the HSCI over the GPV is always at least 16.67%, and, in some cases, is as much as 71%. It is worth noting that, in some extreme cases, when GPV has an adequate sample size required to yield 80% power, the Type II error rate of the HSCI exceeds 50%.

5.5 Discussion

The FDA (HSCI) test assumes that Y_{ijT} and Y_{ijR} are independent, despite the fact that statistical model that includes a covariance term σ_{TR}. We feel that the independence assumption has a detrimental effect on the power of the HSCI test, since this test does well for individual bioequivalence, where the random variables involved are independent. On the other, the GPV test is able to account for the intra-subject dependence of the responses to different formulations. We feel that this is why the GPV test has better power than the

HSCI for PBE. The simulations only examined situations where the correlation coefficient ρ is high $(0.8-0.95)$, even when assessing the size of the test (*i.e.*, on the border of the null hypothesis). In our experience with bioequivalence data, we have found it not unusual to have estimates of ρ (using the method of moments) larger than 1. Even in cases where the evidence for individual and population bioequivalence did not appear to be overly convincing, we have observed $\hat{\rho}$ close to 0.9. Further support for using high levels of ρ comes from Gould (2000), who considered 15 data sets from 3 and 4-period studies. These data sets are available at the FDA bioequivalence website (www.fda.gov/cder/bioequivdata). In these data sets 11/15 had $\rho > 0.8$, and 14/15 had $\rho > 0.7$. Therefore, we feel justified in using what may appear to be somewhat large values of ρ in our simulations.

The anti-conservative tendencies of the GPV test is most pronounced in two cases. The first is where $|\delta|$ (the difference of the means) is greater than $log(1.25)$. In these cases, the conclusion of bioequivalence would not be made unless, at the very least, the point estimate $\hat{\delta}$ was in the range for average bioequivalence (*i.e.*, $\hat{\delta} \in (-log(1.25), log(1.25))$). The practical significance of the anti-conservativeness of the GPV may be attenuated by this. The second condition under which the GPV is most anti-conservative is when σ_R^2 is close to the changeover value for switching from the constant to the reference-scaled criterion. But observed values of σ_R^2 are not likely to be very common. To see why, consider that the total variation is $\sigma_R^2 = \sigma_{BR}^2 + \sigma_{WR}^2$. The estimate of the parameter σ_{WR}^2 is often itself greater than $\sigma_0^2 = 0.04$. And σ_{BR}^2 is usually just as large as, if not much larger than, σ_{WR}^2. Therefore, only products with very low total variation are likely to the affected by the increased anti-conservatism of the GPV test when σ_R^2 is close to 0.04.

One advantage of the GPV over the HSCI for both individual and population bioequivalence is that the GPV is continuous at the changeover point from the constant-scaled to the reference-scaled criterion. That does not appear to be as great of an issue with respect to PBE, since the comparator value for PBE criteria, $\sigma_R^2 = \sigma_{BR}^2 + \sigma_{WR}^2$, tends to be much larger (for reason cited above) than the comparator for the individual bioequivalence criteria σ_{WR}^2. Another possible advantage of the GPV is that it provides what could be interpreted as strength as evidence for a particular conclusion. Such an assessment with

the HSCI may be difficult, if not impossible in some cases, *e.g.*, if one has a publication that only gives the HSCI upper bound. Chapter 3 contains an extensive discussion of some of the advantages of the GPV over the HSCI.

In conclusion, GPV is clearly more powerful than the HSCI in conditions that are very likely to occur in testing for PBE, *i.e.*, high correlation of intra-subject responses and the ratio of between-formulation to within-formulation variances being greater than one. The slight anti-conservativeness of the GPV method is likely to be attenuated by the regulatory requirement that all test formulations show at least average BE with respect to reference formulations. In any case, we feel that our simulation studies provide sufficient evidence that using the FDA (HSCI) test for PBE can result in excessive false negative (Type II error) rates.

Chapter 6

THE EFFICIENCY OF DESIGNS FOR ASSESSING INDIVIDUAL AND POPULATION BIOEQUIVALENCE

SUMMARY

The FDA criteria for individual bioequivalence require crossover designs with more than 2 treatment periods. While 4-period designs provide more information about variance parameters, certain 3-period designs may be used. We compare the 4-period and 3-period designs (2S4P and 2S3P, respectively) from the new FDA guidance document on bioequivalence with each other and with an alternative 3-period design (3S3P) using both asymptotic and empirical methods. The 2S3P design is inefficient compared to the other 2 designs with respect to the number of 'doses' or 'administrations'. The 3S3P design has a modest advantage in efficiency over the 2S4P design for much of the parameter space. The disadvantage of the 3S3P design is that some of the secondary variance parameters are non-identifiable. An investigator should carefully weigh the pros and cons of both designs when choosing between the 2S4P and 3S3P designs.

Extending our methods to population bioequivalence, we found the standard 2 × 2 crossover already in wide use for evaluating average bioequivalence is also the most efficient, in terms of 'doses', for assessing population bioequivalence.

6.1 Introduction

Since the absorption of a drug product tends to be highly variable from individual to individual, crossover studies with a single dose per treatment period are the overwhelmingly preferred instrument for assessing bioequivalence. The evaluation of individual bioequivalence requires using a study design in which treatments are replicated within subjects. Therefore, instead of the standard 2×2 (*i.e.* two sequence/two period) design that has been the standard for evaluating average bioequivalence, three and four-period crossover designs are used to assess individual bioequivalence, with four-period designs being highly recommended by the FDA.

Obviously, complications that hamper the conduct and analysis of crossover studies are more likely to occur in a four-period study than in a two-period study. Chief among these is that subjects in a longer study are more likely to miss treatment periods, or even to discontinue from the study entirely. Therefore, an investigator must weight many factors when choosing a study design for assessing individual bioequivalence. One important factor is the relative efficiency of the candidate study designs that may be used for individual bioequivalence. This investigation has been initiated to examine the relatively efficiency of the designs designated by the FDA, as well as another design that has appeared in the statistical literature. The designs will be compared for efficiency using both large-sample and small-sample methods, the former using asymptotic variances computed by the delta method (Lehmann and Casella, 1998), the latter based on power calculations derived from simulations.

6.2 Methods

The statistical model in this investigation is the same as that recommended by the FDA guidance (2001) for assessing population and individual bioequivalence, which was originally proposed for experiments assessing intra-subject variability by Chinchilli and Esinhart (1996). It is as follows

$$Y_{ijkl} = \mu_k + \gamma_{ikl} + \eta_{ijk} + \epsilon_{ijkl} \quad i = 1, \ldots, s; \quad j = 1, \ldots, n_i; \quad k = T, R; \quad l = 1, q_i;$$

where Y_{ijkl} is the log-transformed response from subject j in sequence i to the l^{th} application of treatment k, s is the number of sequences, and q_i is the number of times treatment k appears in a particular sequence. The parameters μ_T and μ_R are the population means of the treatments, which here are the test (T) and reference (R) formulations of a particular drug product, and γ_{ikl} is the fixed effect corresponding to the l^{th} application of treatment k in sequence i. The following estimability conditions are imposed on the γ_{ikl}

$$\sum_{i=1}^{s}\sum_{l=1}^{q_i}\gamma_{ikl} = 0 \qquad \text{for } k = T, R.$$

Furthermore, η_{ijk} is the random effect of subject j in sequence i when receiving treatment k, and ϵ_{ijkl} represents the within-subject random error. The random variables ϵ_{ijkl} are jointly independent and distributed normally with mean 0 and variance σ^2_{Wk} $(k = R, T)$, and the vectors $\boldsymbol{\eta}_{ij} = [\eta_{ijT}, \eta_{ijR}]'$ are mutually independent bivariate normal with zero means and a covariance matrix given by

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma^2_{BT} & \rho\sigma_{BT}\sigma_{BR} \\ \rho\sigma_{BT}\sigma_{BR} & \sigma^2_{BR} \end{pmatrix}.$$

Finally, $\{\boldsymbol{\eta}_{ij}\}$ and $\{\epsilon_{ijkl}\}$ are mutually independent. For convenience, we also define $\delta = \mu_T - \mu_R$, and $\sigma^2_D = \sigma^2_{BT} + \sigma^2_{BR} - 2\rho\sigma_{BT}\sigma_{BR}$. The latter quantity has become known as the subject-formulation interaction [12], and is especially important in the assessment of individual bioequivalence [13].

The model above assumes that no residual effect of previous treatment periods (*i.e.*, carryover) is present. This assumption will be carried throughout the exposition that follows.

The parameter of interest for individual bioequivalence is

$$\Theta_{IBE} = \frac{\delta^2 + \sigma^2_D + \sigma^2_{WT} - \sigma^2_{WR}}{max(\sigma^2_{WR}, \sigma^2_0)}$$

where δ is the difference between formulation means, σ^2_D is the subject-formulation interaction mentioned above, and σ^2_{WT} and σ^2_{WR} are the within-subject variances for

test and reference formulations, respectively. The reference-scaled criterion for individual bioequivalence is to be applied when σ_{WR}^2 is greater than a FDA specified threshold value σ_0^2. Otherwise, the constant-scaled criterion is to be used, with a possible exception made for drugs with narrow therapeutic windows, for which we may set $\sigma_0^2 = 0$. At present, the FDA recommended value for σ_0^2 is 0.04. Individual bioequivalence is concluded if it can be demonstrated that $\Theta_{IBE} \leq \theta_I$. We are assuming a value of 2.4948 for θ_I, the top of the range of values recommended by the FDA. Since the estimate of σ_D^2 has a statistical distribution that is somewhat intractable, following reparametrization will be used for sequence-balanced (defined below) designs

$$\Theta_{IBE} = \frac{\delta^2 + \sigma_I^2 + c_1\sigma_{WT}^2 - c_2\sigma_{WR}^2}{max(\sigma_{WR}^2, \sigma_0^2)}.$$

Here $\sigma_I^2 = \sigma_D^2 + (1 - c_1)\sigma_{WT}^2 + (c_2 - 1)\sigma_{WR}^2$, $(n.b., \sigma_I^2 = Var(\overline{Y}_{ijT.} - \overline{Y}_{ijR.}))$, and c_1 and c_2 are design-dependent constants.

In the standard 2×2 crossover design that is widely used for the evaluation of average bioequivalence, the parameters $\sigma_D^2, \sigma_{WT}^2$, and σ_{WR}^2 are non-identifiable. Therefore, designs with treatment replication are needed to assess individual bioequivalence. The design preferred by the FDA (2001) is the following 2-sequence, 4-period (2S4P) design

Sequence 1	T	R	T	R
Sequence 2	R	T	R	T.

It is also possible to assess individual bioequivalence using certain 3-period designs. The one given in the FDA guidance (2S3P) is

Sequence 1	T	R	T
Sequence 2	R	T	R.

Chow and Liu (2000) note that this design is a truncated version (last period omitted) of the FDA's 4-period design.

The last design we will consider is another 3-period design (3S3P) used by Hyslop and Iglewicz (2001)

Sequence 1	T	R	R
Sequence 2	R	T	R
Sequence 3	R	R	T.

The 2S4P and 2S3P designs feature within-subject replication of both the reference (R) and test (T) treatments. The 3S3P design has replicates of only the reference treatment. The design parameter $c_1 = 0$ in the alternative parameterization of Θ_{IBE}, so it is not necessary (or possible) to estimate σ^2_{WT}. This design is called an extra-reference design. A similar extra-reference design, using the first 2 sequences of the 3S3P design, was used by both Chow, Shao, and Wang (2002), and Wang (1999) to assess individual bioequivalence. Since the absence of the test formulation in the third treatment period may bias the evaluation of some clinical parameters (*e.g.*, adverse drug reactions) against the reference formulation, we prefer the 3-sequence extra-reference design.

The 2S4P and 3S3P designs are *uniform within sequence* (Chinchilli and Esinhart, 1996), *i.e.*, each treatment appears the same number of times in each sequence. Two useful properties of such designs are (1) The same information is extracted from each sequence for each parameter, and (2) the least squares (for fixed parameters) and method of moments (for variance parameters) estimates are the same as the unconstrained REML estimates (Chinchilli and Esinhart, 1996). In this paper, we use the term *sequence-balanced* synonymously with uniform within sequence. Here 'balance' does not necessarily imply that each sequence has the same number of subjects. Since the 2S3P design is not sequence-balanced, we must define a slighly different parameterization for the individual bioequivalence parameter.

$$\Theta_{IBE} = \frac{\delta^2 + w_1\sigma^2_{I1} + w_2\sigma^2_{I2} + w_1\frac{\sigma^2_{WT}}{2} + \left(\frac{w_2}{2} - 2\right)\sigma^2_{WR}}{max(\sigma^2_{WR}, \sigma^2_0)}.$$

Here $\sigma_{I1}^2 = \sigma_D^2 + \frac{\sigma_{WT}^2}{2} + \sigma_{WR}^2$, $\sigma_{I2}^2 = \sigma_D^2 + \sigma_{WT}^2 + \frac{\sigma_{WR}^2}{2}$, and w_1 and w_2 are weights that add to 1. These weights can be based on the number of subjects in each sequence, as in Hyslop and Iglewicz (2001), or simply $w_1 = w_2 = 0.5$, as in Chow *et al.* (2002), or determined in some other way. Since the exposition that follows assumes an equal number of subjects in both sequences, we have $w_1 = w_2 = 0.5$ by either of the cited references.

All parameters in the individual bioequivalence criteria are estimated using the method of moments, in accordance with FDA guidelines (2001). For sequence-balanced designs, The parameter Θ_{IBE} is a function of the parameters δ, σ_I^2, σ_{WT}^2, and σ_{WR}^2. For the 2S4P design these parameters are estimated, respectively, by the following independent random variables

$$\overline{D} = \frac{1}{s}\sum_{i=1}^{s}\overline{D}_i \qquad S_I^2 = \frac{1}{s(n-1)}\sum_{i=1}^{s}\sum_{j=1}^{n}(D_{ij} - \overline{D}_i)^2$$

$$S_{WT}^2 = \frac{1}{s(n-1)}\sum_{i=1}^{s}\sum_{j=1}^{n}(U_{ijT} - \overline{U}_{iT})^2 \quad S_{WR}^2 = \frac{1}{s(n-1)}\sum_{i=1}^{s}\sum_{j=1}^{n}(U_{ijR} - \overline{U}_{iR})^2$$

where

$$D_{ij} = \frac{Y_{ijT1} + Y_{ijT2} - Y_{ijR1} - Y_{ijR2}}{2} \quad \overline{D}_i = \frac{1}{n}\sum_{j=1}^{n} D_{ij}$$

$$U_{ijT} = \frac{Y_{ijT1} - Y_{ijT2}}{\sqrt{2}} \qquad U_{ijR} = \frac{Y_{ijR1} - Y_{ijR2}}{\sqrt{2}}$$

$$\overline{U}_{iT} = \frac{1}{n}\sum_{j=1}^{n} U_{ijT} \qquad \overline{U}_{iR} = \frac{1}{n}\sum_{j=1}^{n} U_{ijR}.$$

Slight modifications are made to derive the method of moments estimators for the other designs. Then for a sequence-balanced design under the above model with the same number of subjects in each sequence, the covariance matrix for the estimates of these parameters is

$$V = Diag\left(\frac{\sigma_I^2}{sn}, \frac{2\sigma_I^4}{s(n-1)}, \frac{2\sigma_{WT}^4}{s(n-1)}, \frac{2\sigma_{WR}^4}{s(n-1)}\right).$$

Then we compute the vector of first partial derivatives of Θ_{IBE} with respect to its constituent parts

$$a^t = \left(\frac{2\delta}{\sigma_{WR}^2} \quad \frac{1}{\sigma_{WR}^2} \quad \frac{c_1}{\sigma_{WR}^2} \quad \frac{-(\delta^2 + \sigma_I^2 + c_1\sigma_{WT}^2)}{\sigma_{WR}^4} \right)$$

for reference-scaled Θ_{IBE}, and

$$a^t = \left(\frac{2\delta}{\sigma_0^2} \quad \frac{1}{\sigma_0^2} \quad \frac{c_1}{\sigma_0^2} \quad \frac{-c_2}{\sigma_0^2} \right)$$

for constant-scaled Θ_{IBE}. By the delta method (see Lehmann and Casella, 1998), the asymptotic variances are then $AVar(\hat{\Theta}_{IBE}) = a^t V a$. Since the 2S3P design is not sequence-balanced, slight modifications must be made to the covariance matrix and the vector a in order to compute the asymptotic variance of $\hat{\Theta}_{IBE}$. The covariance matrix is then

$$V = Diag\left(\frac{1}{4}\frac{\sigma_{I1}^2 + \sigma_{I2}^2}{n}, \frac{2\sigma_{I1}^4}{n-1}, \frac{2\sigma_{I2}^4}{n-1}, \frac{2\sigma_{WT}^4}{n-1}, \frac{2\sigma_{WR}^4}{n-1}\right).$$

The vector a for reference-scaled criterion is

$$a^t = \left(\frac{2\delta}{\sigma_{WR}^2} \quad \frac{1}{2\sigma_{WR}^2} \quad \frac{1}{2\sigma_{WR}^2} \quad \frac{1}{4\sigma_{WR}^2} \quad \frac{-\left(\delta^2 + \frac{\sigma_{I1}^2 + \sigma_{I2}^2}{2} + \frac{\sigma_{WT}^2}{4}\right)}{\sigma_{WR}^4} \right)$$

and for the constant-scaled criterion a is

$$a^t = \left(\frac{2\delta}{\sigma_0^2} \quad \frac{1}{2\sigma_0^2} \quad \frac{1}{2\sigma_0^2} \quad \frac{1}{4\sigma_0^2} \quad \frac{-7}{4\sigma_0^2} \right).$$

Note that for the 2S3P design, there are two parameters, σ_{I1}^2 and σ_{I2}^2, for the variance of the estimated mean difference between the formulations, one for each sequence.

The relative efficiency of design x to design y is then

$$ARE = \frac{AVar(\hat{\Theta}_{IBEx})}{AVar(\hat{\Theta}_{IBEy})}.$$

Since we are dealing with asymptotic variances, it seems appropriate to call this quantity the asymptotic relative efficiency (ARE) of the designs in question, although this term is usually associated with comparisons of estimators rather than with designs. Separate comparisons are made for the reference and constant-scaled criteria for each pair of designs. With the relatively small sample sizes that are typically used in bioequivalence studies, the asymptotic variance may not be very close to the actual variance of the estimate $\hat{\Theta}_{IBE}$. The purpose of computing $AVar(\hat{\Theta}_{IBE})$ is to assess qualitatively the relative efficiency of each pair. The asymptotic calculation may also give a feel for what factors are likely to

effect relative efficiency, so that the qualitative results from the large sample calculations may be useful when we do the small sample evaluations of relative efficiency. The asymptotic comparison also gives us a quick way of assessing the relative efficiency of designs over large segments of the parameter space. On the other hand, the empirical method described below requires considerable computational effort for just one comparison point. Plotting the graphs below, each containing hundreds of points, would be a Herculean task if we needed to calculate each point using the empirical method.

Empirical comparisons were made using the required sample size n in each sequence to achieve at least 80% power for a set of given design parameter. These power calculations were done by simulation. The empirical relative efficiency of design x to design y is then

$$\frac{s_x \times n_x \times p_x}{s_y \times n_y \times p_y}$$

where p_i is the number of periods in design i. Therefore, we are actually evaluating efficiency by comparing the total number of treatment periods or 'observations' (*i.e.*, $s \times n \times p$). The same assumption is being made in computing the asymptotic variances. The rationale behind this will be discussed below. Two tests are used to evaluate empirical relative efficiency, the Howe-type small-sample confidence interval (the method featured in the FDA guidance document) (FDA 2001, Howe 1974, Hyslop *et al* 2000) and the generalized p-value (Tsui and Weerahandi, 1989). An extensive comparison of these 2 methods for individual bioequivalence appears in Chapter 3, which also includes a detailed description of how the power simulations were performed. All power simulations were done by simulation using SAS® version 8 for Windows.

6.3 Results for IBE

For sequence-balanced designs (2S4P and 3S3P), the asymptotic variance of $\hat{\Theta}_{IBE}$ is

$$\frac{4(n-1)\delta^2\sigma_I^2 + 2n(\sigma_I^4 + c_1^2\sigma_{WT}^4 + (\delta^2 + \sigma_I^2 + c_1\sigma_{WT}^2)^2)}{sn(n-1)\sigma_{WR}^4}$$

for the reference-scaled criterion, and

$$\frac{4(n-1)\delta^2\sigma_I^2 + 2n(\sigma_I^4 + c_1^2\sigma_{WT}^4 + c_2^2\sigma_{WR}^4)}{sn(n-1)\sigma_0^4}$$

for the constant-scaled criterion. For the 2S4P design $c_1 = 0.5$ and for the 3S3P (extra-reference) design $c_1 = 0$. For both the 2S4P and 3S3P designs $c_2 = 1.5$.

The 2S3P design is not sequence-balanced, so the asymptotic variance of $\hat{\Theta}_{IBE}$ is different. It is

$$\frac{(n-1)\delta^2(\sigma_{I1}^2+\sigma_{I2}^2)+n(0.5\sigma_{I1}^4+0.5\sigma_{I2}^4+0.125\sigma_{WT}^4+2(\delta^2+0.5\sigma_{I1}^2+0.5\sigma_{I2}^2+0.25\sigma_{WT}^2)^2)}{n(n-1)\sigma_{WR}^4}$$

for the reference-scaled criterion, and

$$\frac{(n-1)\delta^2\sigma_I^2+n(0.5\sigma_{I1}^4+0.5\sigma_{I2}^4+0.125\sigma_{WT}^4+6.125\sigma_{WR}^4)}{n(n-1)\sigma_0^4}$$

for the constant-scaled criterion. Here $\sigma_{I1}^2=\sigma_D^2+0.5\sigma_{WT}^2+\sigma_{WR}^2$, and $\sigma_{I2}^2=\sigma_D^2+\sigma_{WT}^2+0.5\sigma_{WR}^2$.

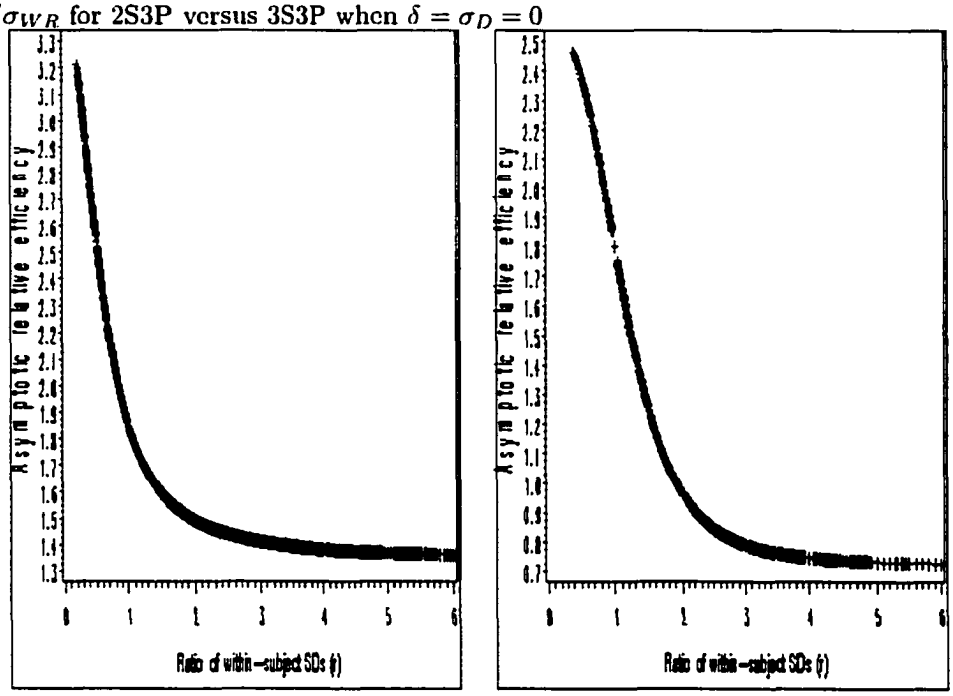
The relative efficiency results are summarized in Table 1 (large sample) and Table 2 (small sample). For the constant-scaled criterion, Table 1 uses $n = 9$ subjects per sequence for the 2S4P design, $n = 12$ per sequence for the 2S3P design, and $n = 8$ for the 3S3P design. For the reference-scaled criterion, the number of subjects per sequence is doubled for each design. Therefore, each design has 72 or 144 total 'observations' or 'doses' (Hyslop and Iglewicz (2001) used the term 'administrations'). The large sample results are based on the observation that, for δ and σ_D^2 close to 0, the asymptotic relative efficiency (ARE) is a function of the ratio $r = \sigma_{WT}/\sigma_{WR}$. This relationship is illustrated in Figure 6.1 for the comparison of the 2S3P design to the 3S3P design when δ and σ_D equal 0. Analogous graphs for comparing the other designs are similar. In Table 1, the first column shows the designs being compared, the second the individual bioequivalence criteria, either reference-scaled (R) or constant-scaled (C). The next 5 columns show the ARE's for various values of r^2, assuming $\delta = 0$ and $\sigma_D^2 = 0$. The last column shows the value of r for which $ARE = 1$, *i.e.*, when the two designs are equally efficient.

Table 1. Large Sample Results

Designs Compared	Criteria	ARE $\delta = 0, \sigma_D^2 = 0$ $r^2 = 0.5, 0.75, 1, 1.25, \infty$					r $ARE = 1$
2S4P/3S3P	R	1.075	1.046	1.0294	1.020	0.9926	1.689
2S4P/3S3P	C	1.161	1.087	1.0208	0.965	0.6563	1.0432
2S3P/3S3P	R	2.171	1.971	1.8478	1.764	1.3451	never
2S3P/3S3P	C	2.184	1.983	1.8030	1.648	0.7159	1.9096
2S3P/2S4P	R	2.018	1.885	1.7950	1.730	1.3551	never
2S3P/2S4P	C	1.881	1.825	1.7662	1.708	1.0909	never

$r = \sigma_{WT}/\sigma_{WR}$ R=Reference-scaled C=Constant-scaled

Figure 6.1: Relative efficiency versus ratio of within-subject standard deviations $r = \sigma_{WT}/\sigma_{WR}$ for 2S3P versus 3S3P when $\delta = \sigma_D = 0$



Left: Reference-scaled criterion Right: Constant-scaled criterion

(*Note that vertical scales differ*)

Table 1 shows that the 2S3P design can be expected to require more 'doses' to estimate Θ_{IBE} with the same precision as the sequence-balanced designs (3S3P and 2S4P). The exception is when the ratio $r = \sigma_{WT}/\sigma_{WR}$ is almost larger than 2 (*i.e.*,, the within-subject standard deviation is almost twice as large for the test formulation) for the comparison to the 3S3P design using the constant-scaled criterion, *i.e.*, when $\sigma^2_{WR} < 0.04$, but with such a large ratio for the within-subject standard deviation we would not want to conclude individual bioequivalence, since this corresponds to $\Theta_{IBE} = 2.646$, larger than the regulatory upper bound on this parameter. This lower precision for estimating Θ_{IBE} translates directly into a greater number of 'doses' (and, by extension, a greater number of subjects)

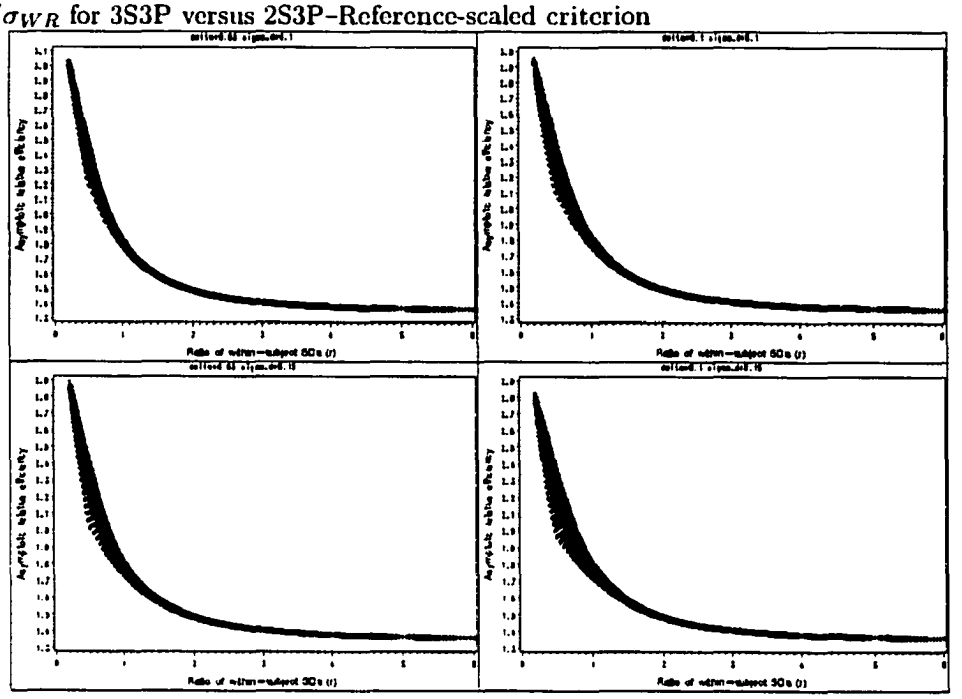
needed to assure adequate power for concluding individual bioequivalence. We will see the direct effect of this below in Table 2. In comparing the 2S4P and 3S3P designs we find that this 3-period design appears to have a modest advantage over the 4-period design over a large part of the parameter space. The exception again is for larger values of r, especially when the constant-scaled criterion comes into play.

Figures 6.2-6.7 show the effect of slightly varying values of $\delta(= 0.05, 0.1)$ and $\sigma_D^2(= 0.1, 0.15)$ on ARE. The horizon scale is the ratio of within-subject variation $r = \sigma_{WT}/\sigma_{WR}$, where σ_{WT} is allowed to vary from 0.2 to 1.2 for reference-scaling, and from 0.1 to 0.6 for constant-scaling. The reader should note that the vertical scales both between and within the figures. Figures 2 and 3 show the relative efficiencies for the comparison of the 3S3P design to the 2S3P design for the reference and constant-scaled criteria, respectively. We see the effect of δ and σ_D being greater than zero is minimal for reference-scaling. For constant-scaling, the data cloud appears to have a little more spread, but the relative efficiency values do not appear to deviate significantly from those in Table 1. Figures 4 and 5 show same comparison for 3S3P versus 2S4P for reference and constant-scaling, respectively. The data cloud in Figure 4 appears to be larger than in Figure 1 (partly due to the smaller vertical scale), but here the deviation from Table 1 would appear to favor the 3S3P design. Figure 5 would not likely change any of our observations from Table 1. Figures 6 and 7 show comparative relative efficiencies for 2S4P versus 2S3P for reference and constant-scaling, respectively. In Figure 6, the bottom edge of the data cloud shows an apparent attenuation of the efficiency advantage of the 2S4P over the 2S3P when r is close to 1, although the 4-period design still appears to be substantially better. It should be noted that on the lower edge of the data cloud we have $\sigma_{WR} = 0.2$, *i.e.*, we are at the changeover point from constant to reference-scaling. In Figure 7, we see that the ARE can go as low as 1.05 for $\delta = 0.1, \sigma_D = 0.15, \sigma_{WR} = 0.1$, and $r = 1$. It should be noted that the empirical relative efficiencies for this point are 1.167 and 1.333 for the FDA and generalized p-value tests, respectively.

In Table 2, the ratios of the total number of observations required for 80% power (*i.e.*, the of subjects required for 80% power times the number of sequences times the

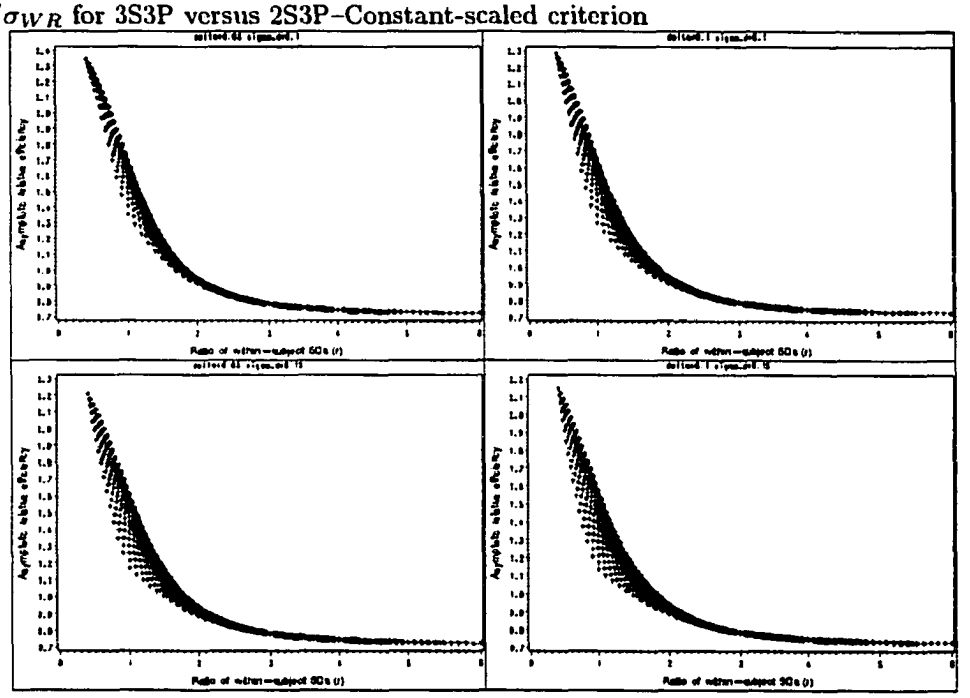
number of periods in a given design) are compared, assuming $\delta = 0.05$, $\sigma_D = 0.1, r^2 = 0.5, 0.75, 1, 1.25$, and $\sigma_{WR} = 0.15, 0.23, 0.5$, these three levels roughly corresponding to low, medium, and high intra-subject variances. The two tests used to compare the designs are the the Howe-type small-sample confidence interval (H) (Hyslop *et al* 2000, Chow *et al* 2002) and the generalized p-value (G) tests of Chapters 3 and 4. The column labelled A is the large sample asymptotic relative efficiency computed in the same manner as the relative efficiencies in Table 1.

Figure 6.2: Relative efficiency versus ratio of within-subject standard deviations $r = \sigma_{WT}/\sigma_{WR}$ for 3S3P versus 2S3P-Reference-scaled criterion



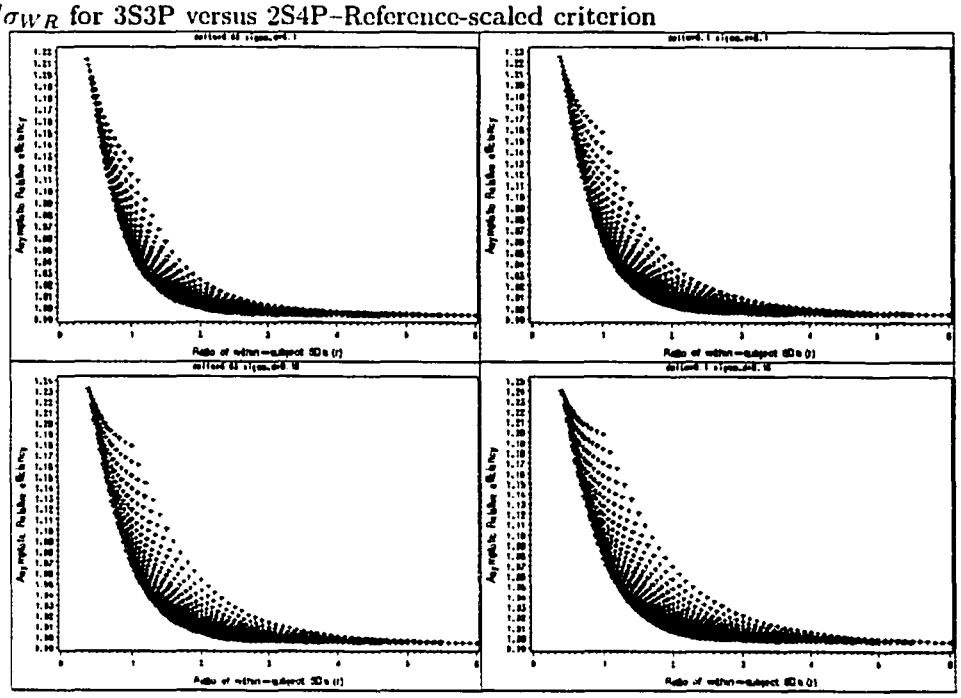
(*Note that vertical scales differ*)

Figure 6.3: Relative efficiency versus ratio of within-subject standard deviations $r = \sigma_{WT}/\sigma_{WR}$ for 3S3P versus 2S3P–Constant-scaled criterion



(*Note that vertical scales differ*)

Figure 6.4: Relative efficiency versus ratio of within-subject standard deviations $r = \sigma_{WT}/\sigma_{WR}$ for 3S3P versus 2S4P–Reference-scaled criterion



(*Note that vertical scales differ*)

Figure 6.5: Relative efficiency versus ratio of within-subject standard deviations $r = \sigma_{WT}/\sigma_{WR}$ for 3S3P versus 2S4P–Constant-scaled criterion

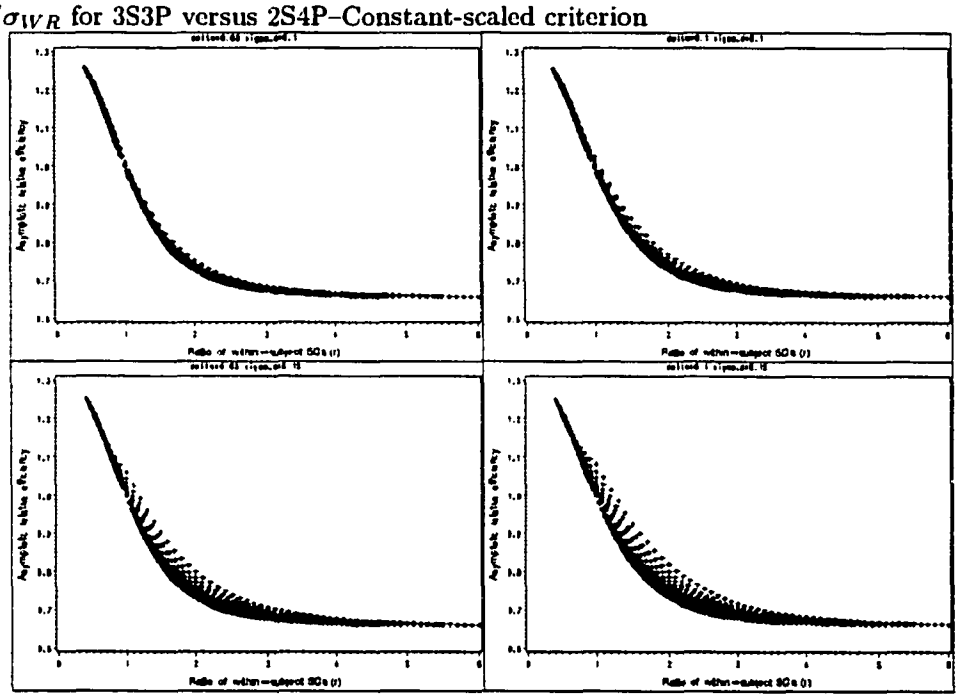
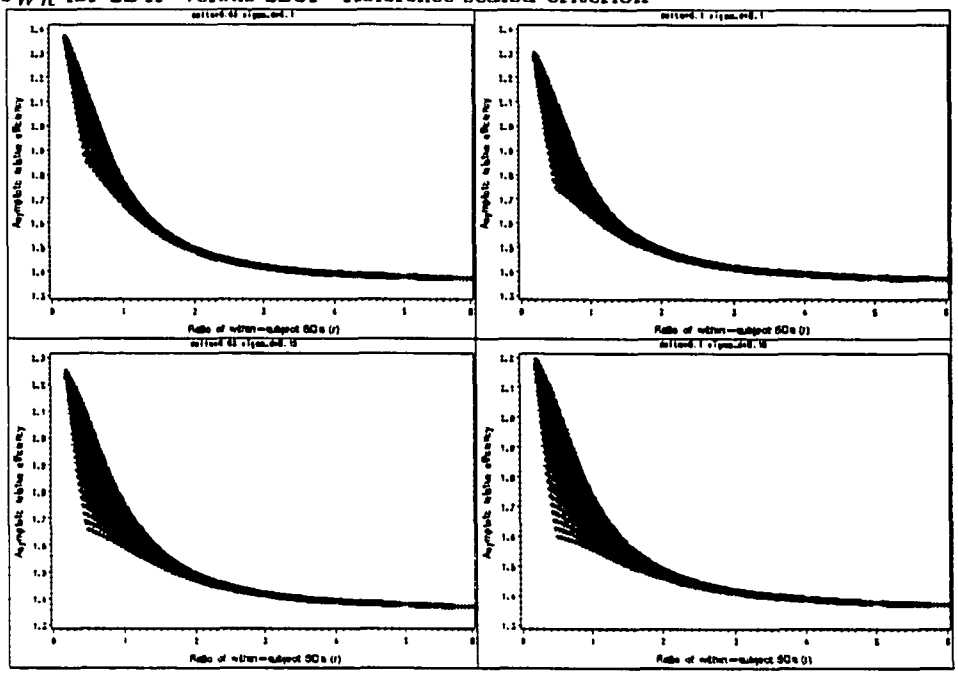
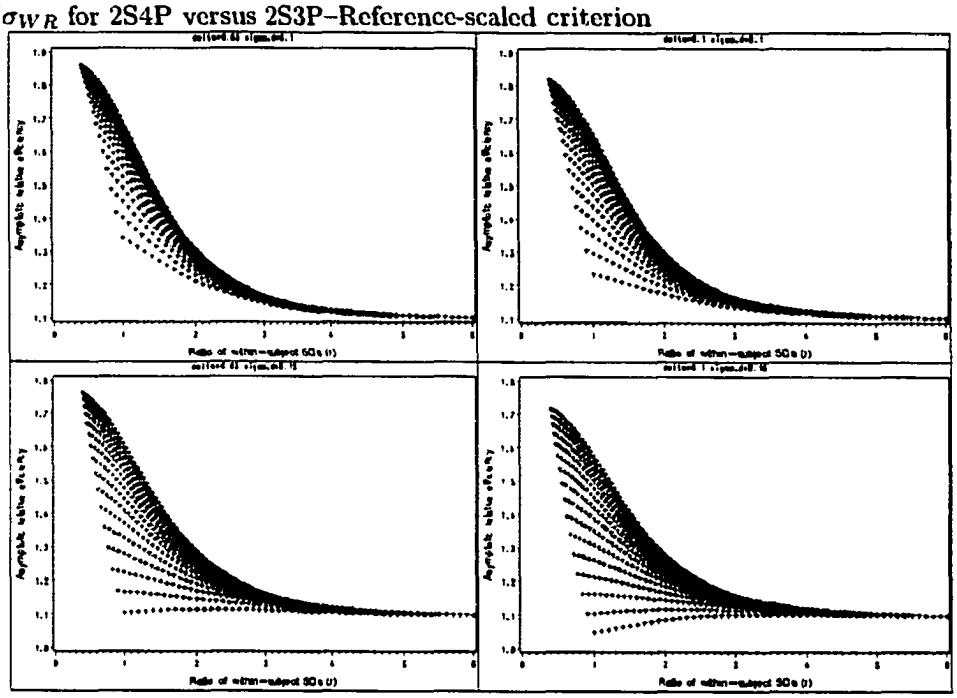


Figure 6.6: Relative efficiency versus ratio of within-subject standard deviations $r = \sigma_{WT}/\sigma_{WR}$ for 2S4P versus 2S3P–Reference-scaled criterion



(*Note that vertical scales differ*)

Figure 6.7: Relative efficiency versus ratio of within-subject standard deviations $r = \sigma_{WT}/\sigma_{WR}$ for 2S4P versus 2S3P–Reference-scaled criterion



(Note that vertical scales differ)

Table 2. Small Sample Results

	$\delta = 0.05$, $\sigma_D = 0.1$								
	$r^2 = 0.5$								
Designs	$\sigma_{WR} = 0.15$			$\sigma_{WR} = 0.23$			$\sigma_{WR} = 0.5$		
Compared	H	G	A	H	G	A	H	G	A
2S4P/3S3P	0.889	1.067	1.130	1.016	1.016	1.110	1.016	1.143	1.084
2S3P/3S3P	1.2	1.333	1.851	1.619	1.619	2.026	1.619	1.810	2.134
2S3P/2S4P	1.35	1.25	1.638	1.594	1.594	1.825	1.594	1.583	1.969
	$r^2 = 0.75$								
Designs	$\sigma_{WR} = 0.15$			$\sigma_{WR} = 0.23$			$\sigma_{WR} = 0.5$		
Compared	H	G	A	H	G	A	H	G	A
2S4P/3S3P	1.067	1.067	1.058	1.086	1.086	1.074	1.086	1.086	1.052
2S3P/3S3P	1.467	1.6	1.689	1.630	1.630	1.885	1.630	1.778	1.950
2S3P/2S4P	1.375	1.5	1.597	1.5	1.5	1.755	1.5	1.636	1.854
	$r^2 = 1.0$								
Designs	$\sigma_{WR} = 0.15$			$\sigma_{WR} = 0.23$			$\sigma_{WR} = 0.5$		
Compared	H	G	A	H	G	A	H	G	A
2S4P/3S3P	1.037	1.037	0.998	1.026	1.037	1.052	1.037	1.111	1.034
2S3P/3S3P	1.333	1.556	1.553	1.590	1.611	1.792	1.611	1.778	1.834
2S3P/2S4P	1.286	1.5	1.557	1.55	1.554	1.703	1.554	1.6	1.773
	$r^2 = 1.25$								
Designs	$\sigma_{WR} = 0.15$			$\sigma_{WR} = 0.23$			$\sigma_{WR} = 0.5$		
Compared	H	G	A	H	G	A	H	G	A
2S4P/3S3P	1.016	0.889	0.948	1.037	0.993	1.038	1.0	1.111	1.024
2S3P/3S3P	1.333	1.333	1.440	1.593	1.569	1.725	1.625	1.75	1.755
2S3P/2S4P	1.313	1.5	1.519	1.536	1.579	1.662	1.625	1.575	1.716

H=Howe-type small-sample CI G=Generalized p-value
A=Asymptotic relative efficiency calculation $r^2 = \sigma^2_{WT}/\sigma^2_{WR}$

For the most part, Table 2 appears to confirm the the effects on relative efficiency suggested by Table 1. With the exception of when the within-subject variance ratio is very low ($r^2 = 0.5$), we feel somewhat encouraged by the agreement between the theoretic and empirical values, given that the sample sizes are modest (40-56 doses for the more efficient designs (2S4P and 3S3P) with low variability, and around 120 doses for these designs with moderate and high variability). For $r^2 = 1$, it may be interesting to note that the 3S3P design appears to be empirically more efficient than the 2S4P design, barely the opposite of what the large sample calculation would suggest. This may be the result of the asymptotic calculation being based on the constant-scaled criterion, while the empirical result is based on power calculations that use mixed scaling, since the asymptotic calculation for the reference-scaled criterion here is 1.072, a value favoring the 3S3P design. This may also be an artifact caused by the discreteness in the empirical calculations. What we mean by 'discreteness' is that the estimated relative efficiencies here are the ratios of the number of observations, which are whole numbers that can only have a finite number of values. The reason for this is twofold: the required sample size per sequence is whole number, which when realized may yield an estimated power from 80.0% to 89%. The number of observations, assuming an equal number of subjects in each sequence, are multiples of 8, 9, and 6 for the 2S4P, 3S3P, and 2S3P designs, respectively, so that the estimated relative efficiencies tend to be quite grainy, especially when the sample size requirements are relatively low. So when the discreteness is taken into account, it appears that the large sample calculations do relatively well in predicting the small sample relative efficiencies, especially when r^2 is close to 1.

The important point to be made concerning Table 2 is that the 2S3P is never as efficient as the other two designs.

6.4 Extension to Population Bioequivalence

The methodology above can be extended to evaluate study designs with respect to their ability to assess Θ_{PBE}, the FDA parameter for population bioequivalence (PBE). Using the same statistical model that we had for IBE, the FDA parameter for PBE is

$$\Theta_{PBE} = \frac{\delta^2 + \sigma_{TT}^2 - \sigma_{RR}^2}{max(\sigma_{RR}^2, \sigma_0^2)}$$

where δ is the same as above, $\sigma_{TT}^2 = \sigma_{BT}^2 + \sigma_{WT}^2$ is the total variation of the test formulation, and $\sigma_{RR}^2 = \sigma_{BR}^2 + \sigma_{WR}^2$ is the total variation of the reference formulation. Here we again have in essence two criterion, reference and constant-scaled, depending on whether the variance of the reference formulation is greater or less than some changeover value σ_0^2.

In population bioequivalence, the estimability of parameters is no longer at issue, as it was for IBE: any crossover design will be adequate to assess Θ_{PBE} and it constituents, and even parallel-group studies could be used, if necessary. Therefore, we considered the 2S4P design, the 3S3P design, and the typical 2 × 2 (2S2P) design that has been the standard for assessing average bioequivalence. We do not consider the 2S3P design with respect to relative efficacy, since we have shown in a previous chapter [Chapter 4] that the 3S3P crossover design is superior with respect to PBE; and since with parameter estimability is no longer an issue, the 2S3P holds no advantage over the 3S3P design. In assessing relative efficiency of designs for PBE, we will be using both the asymptotic and empirical methods described above.

The evaluation of the asymptotic relative efficiency for PBE is simplified by the fact that all 3 of the designs we are considering are sequence-balanced, allowing us to generalize the PBE criteria and formulas in terms of the design constants c_1 and c_2. First we reparameterize Θ_{PBE} in terms of the estimable parameters and the design constants.

$$\Theta_{PBE} = \frac{\delta^2 + \sigma_T^2 + c_1\sigma_{WT}^2 - (\sigma_R^2 + c_2\sigma_{WR}^2)}{max(\sigma_R^2 + c_2\sigma_{WR}^2, \sigma_0^2)}$$

Here δ, σ^2_{WT}, and σ^2_{WR} are as before, $\sigma^2_T = \sigma^2_{BT} + (1-c_1)\sigma^2_{WT}$, and $\sigma^2_R = \sigma^2_{BR} + (1-c_2)\sigma^2_{WR}$. To estimate the necessary parameters, we need the following additional statistics

$$S^2_T = \frac{1}{s(n-1)} \sum_{i=1}^{s} \sum_{j=1}^{n} (V_{ijT} - \overline{V}_{iT})^2$$

$$S^2_R = \frac{1}{s(n-1)} \sum_{i=1}^{s} \sum_{j=1}^{n} (V_{ijR} - \overline{V}_{iR})^2$$

$$SS_{TR} = \sum_{i=1}^{s} \sum_{j=1}^{n} (V_{ijT} - \overline{V}_{iT})(V_{ijR} - \overline{V}_{iR})$$

where

$$V_{ijT} = \frac{Y_{ijT1} + Y_{ijT2}}{2} \qquad V_{ijR} = \frac{Y_{ijR1} + Y_{ijR2}}{2}$$

$$\overline{V}_{iT} = \frac{1}{n} \sum_{j=1}^{n} V_{ijT} \qquad \overline{V}_{iR} = \frac{1}{n} \sum_{j=1}^{n} V_{ijR}$$

for the 2S4P design. The statistics are suitably modified for the other designs. The design constants are $c_1 = c_2 = 0.5$ for the 2S4P design, $c_1 = 0$ and $c_2 = 0.5$ for the 3S3P design, and $c_1 = c_2 = 0$ for the 2S2P design. Note that when a design constant is zero, it is not necessary (or even possible) to estimate the corresponding within formulation variance for that design. That is, we do not need S^2_{WT} for the 3S3P and 2S2P designs, or S^2_{WR} for the 2S2P design.

To calculate $AVar(\hat{\Theta}_{PBE})$ we need the constituents of $a^t V a$. The covariance matrix is

$$V = \begin{pmatrix} \frac{\sigma^2_I}{sn} & 0 & 0 & 0 & 0 \\ 0 & 2\frac{\sigma^4_T}{s(n-1))} & 2\frac{\sigma^2_{TR}}{s(n-1)} & 0 & 0 \\ 0 & 2\frac{\sigma^2_{TR}}{s(n-1)} & 2\frac{\sigma^4_R}{s(n-1))} & 0 & 0 \\ 0 & 0 & 0 & 2\frac{\sigma^4_{WT}}{s(n-1))} & 0 \\ 0 & 0 & 0 & 0 & 2\frac{\sigma^4_{WR}}{s(n-1))} \end{pmatrix}$$

where $\sigma^2_I = Var(\hat{\delta}) = \sigma^2_T + \sigma^2_R - 2\sigma_{TR}$. Note that, unlike the covariance matrix that we used to get $AVar(\hat{\Theta}_{IBE})$, this covariance matrix has non-zero off-diagonal elements. This is due the covariance of V_{ijT} and V_{ijR} generally being non-zero. The vector a is

$$a^t = \begin{pmatrix} \frac{2\delta}{\sigma^2_R + c_2\sigma^2_{WR}} & \frac{1}{\sigma^2_R + c_2\sigma^2_{WR}} & -\frac{\delta^2 + \sigma^2_T + c_1\sigma^2_{WT}}{(\sigma^2_R + c_2\sigma^2_{WR})^2} & \frac{c_1}{\sigma^2_R + c_2\sigma^2_{WR}} & -c_2\frac{\delta^2 + \sigma^2_T + c_1\sigma^2_{WT}}{(\sigma^2_R + c_2\sigma^2_{WR})^2} \end{pmatrix}$$

for the reference-scaled criterion, and

$$a^t = \left(2\frac{\delta}{\sigma_0^2} \quad \frac{1}{\sigma_0^2} \quad -\frac{1}{\sigma_0^2} \quad \frac{c_1}{\sigma_0^2} \quad -\frac{c_2}{\sigma_0^2} \right)$$

for the constant-scaled criterion. The asymptotic variances of $\hat{\Theta}_{PBE}$ are then

$$2\left(\frac{2\delta^2\sigma_I^2}{sn(\sigma_R^2+c_2\sigma_{WR}^2)^2}+\frac{\sigma_T^4+c_1^2\sigma_{WT}^4}{s(n-1)(\sigma_R^2+c_2\sigma_{WR}^2)^2}-\frac{2(\delta^2+\sigma_T^4)\sigma_{TR}^2}{s(n-1)(\sigma_R^2+c_2\sigma_{WR}^2)^3}+\frac{(\delta^2+\sigma_T^4)^2(\sigma_R^4+c_2^2\sigma_{WR}^4)}{s(n-1)(\sigma_R^2+c_2\sigma_{WR}^2)^4}\right)$$

for the reference-scaled criterion, and

$$2\left(\frac{2\delta^2\sigma_I^2}{sn\sigma_0^4}+\frac{\sigma_T^4+c_1^2\sigma_{WT}^4-2\sigma_{TR}^2+\sigma_R^4+c_2^2\sigma_{WR}^4}{s(n-1)\sigma_0^4}\right)$$

for the constant-scaled criterion. We then compute the asymptotic relative efficiency (ARE) in the manner described in the Methods Section. Empirical relative efficiencies are also calculated as described above. The empirical results are based on power calculations using the generalized p-value test , since this test accounts for correlation between summary statistics, while the HSCI/FDA test does not.

One challenge with multi-parameter problems is discovering the relationships that affect whatever properties we are studying. In the IBE case, we found that the ratio σ_{WT}/σ_{WR} was key in the asymptotic relative efficiency of the designs that we considered, especially when the other parameters were close to 0. The PBE case has more parameters than IBE, so the situation is not so clear. But an examination of some scenarios suggests that the key relationship here is the ratio of the between-subject and within-subject standard deviations ($\kappa = \sigma_B/\sigma_W$). To facilitate comparisons between the empirical and asymptotic calculation, we varied the total number of doses in each comparison to try to match the total doses in the empirical calculations, where possible. Therefore the 2S4P/3S3P comparison is based on 72 doses, the 3S3P/2S2P comparison is based on 36 doses, and the 2S4P/2S2P comparison is based on 48 doses.

Table 3 shows empirical and asymptotic results for $\delta = 0.05, \sigma_{WT} = \sigma_{WR} = 0.15, 0.23, 0.5, \rho = 0.9, 0.95$, and $\kappa = 1.0, 2.5(0.5)$. The line with the value of ρ gives the empirical relative efficiencies, and the line marked 'Asym' gives the asymptotic values.

Table 3a. Relative Efficiency Results for PBE

2S4P vs. 3S3P

	$\delta = 0.05$											
	$\sigma_{WR} = 0.15$ $\sigma_{BT}/\sigma_{WT} = \sigma_{BR}/\sigma_{WR}$				$\sigma_{WR} = 0.23$ $\sigma_{BT}/\sigma_{WT} = \sigma_{BR}/\sigma_{WR}$				$\sigma_{WR} = 0.5$ $\sigma_{BT}/\sigma_{WT} = \sigma_{BR}/\sigma_{WR}$			
ρ	1.0	1.5	2.0	2.5	1.0	1.5	2.0	2.5	1.0	1.5	2.0	2.5
0.90	1.067	1.067	1.067	1.111	0.889	1.067	1.067	1.111	1.037	1.067	1.067	1.111
Asym	0.926	0.968	1.017	1.064	0.916	0.962	1.013	1.062	0.911	0.959	1.011	1.060
0.95	1.067	1.067	1.111	1.111	1.037	1.067	1.111	1.111	1.037	1.067	1.111	1.111
Asym	0.910	0.932	0.963	0.999	0.900	0.926	0.960	0.996	0.895	0.923	0.958	0.995

Asym=asymptotic

Table 3b. Relative Efficiency Results for PBE

3S3P vs. 2S2P

	$\delta = 0.05$											
	$\sigma_{WR} = 0.15$ $\sigma_{BT}/\sigma_{WT} = \sigma_{BR}/\sigma_{WR}$				$\sigma_{WR} = 0.23$ $\sigma_{BT}/\sigma_{WT} = \sigma_{BR}/\sigma_{WR}$				$\sigma_{WR} = 0.5$ $\sigma_{BT}/\sigma_{WT} = \sigma_{BR}/\sigma_{WR}$			
ρ	1.0	1.5	2.0	2.5	1.0	1.5	2.0	2.5	1.0	1.5	2.0	2.5
0.90	1.250	1.250	1.406	1.286	1.313	1.250	1.406	1.286	1.227	1.250	1.406	1.286
Asym	1.327	1.380	1.434	1.483	1.345	1.391	1.440	1.488	1.357	1.398	1.444	1.490
0.95	1.406	1.406	1.286	1.50	1.125	1.250	1.286	1.50	1.227	1.250	1.286	1.50
Asym	1.313	1.350	1.386	1.423	1.332	1.361	1.394	1.428	1.344	1.368	1.398	1.430

Asym=asymptotic

Table 3c. Relative Efficiency Results for PBE

2S4P vs. 2S2P

	$\delta = 0.05$											
	$\sigma_{WR} = 0.15$ $\sigma_{BT}/\sigma_{WT} = \sigma_{BR}/\sigma_{WR}$				$\sigma_{WR} = 0.23$ $\sigma_{BT}/\sigma_{WT} = \sigma_{BR}/\sigma_{WR}$				$\sigma_{WR} = 0.5$ $\sigma_{BT}/\sigma_{WT} = \sigma_{BR}/\sigma_{WR}$			
ρ	1.0	1.5	2.0	2.5	1.0	1.5	2.0	2.5	1.0	1.5	2.0	2.5
0.90	1.333	1.333	1.50	1.429	1.167	1.333	1.50	1.429	1.273	1.333	1.50	1.429
Asym	1.161	1.261	1.376	1.489	1.163	1.263	1.377	1.490	1.165	1.263	1.378	1.490
0.95	1.50	1.50	1.429	1.667	1.167	1.333	1.429	1.667	1.273	1.333	1.429	1.667
Asym	1.130	1.188	1.261	1.341	1.132	1.189	1.262	1.341	1.134	1.190	1.262	1.342

Asym=asymptotic

Based on the empirical results, Table 3 suggests that there is very little to be gained by adding periods to a crossover study when the objective is assessing PBE, using Θ_{PBE}. Given the added problems that can occur with longer studies, and the fact that, assuming that there is no carryover effect, there is no additional information concerning Θ_{PBE} that one can derive from a longer study that is not available in the standard 2×2 crossover design, there does not seem to be any reason to use anything other than the 2×2 design for evaluating population bioequivalence. There does appear to be a direct relationship between κ and relative efficiency, which we feel is an important result, since we expect

between-subject variation to be larger than within-subject variation, as this is the chief rationale for using a crossover design. The role of ρ in determining relative efficiency is not as pronounced in the empirical results are it is in the asymptotic calculations. In general, the asymptotic results do not appear to predict the empirical efficiency as well as they did for IBE, with the 3S3P/2S2P comparison being a possible exception. The discreteness of the relative efficiency calculation seems to be more of a problem here than it was for IBE. This is not surprising, since the sample sizes are somewhat smaller for PBE, so that the 'jumps' in power are larger between successive values of n.

6.5 Discussion

Basing comparisons on the number of 'observations' (*i.e.*, *snp*), appeared to be the fairest way to compare 3 and 4-period designs. Another way of describing what we have done is that we have compared designs based on the total number of 'doses'. The subject-treatment-period notion of 'observation/dose' is merely an objective (although naive) way of quantifying a unit of effort/cost so that designs with different numbers of periods could be directly compared. A reasonable 4-period design, such as the 2S4P design, is naturally expected to be more efficient than a 3-period design with the same number of subjects. But it may be difficult to assess objectively the possible additional burden of the 4-period design, except to consider the additional 'observations' in Period 4 as somewhat equivalent to the additional subjects needed in the 3-period designs. Additional factors which are difficult to quantify, such as the increased risk of subject withdrawal in a longer study, may motivate investigators to choose a 3-period crossover design when the 3 and 4-period designs can be judged to be roughly 'equivalent' in some respect.

It is clear from the results that the 2S3P design, which is the 3-period design cited in the FDA guidance document, is inefficient compared to the sequence-balanced designs, 2S4P and 3S3P, in respect to the effort required (*i.e.*, the number of 'doses') to show individual bioequivalence. For compounds with a reasonable degree of bioequivalence, the large sample results suggest that the 2S3P design may require $65\% - 120\%$ more ob-

servations to estimate Θ_{IBE} with the same precision as the sequence-balanced designs. The only circumstance we found under which the 2S3P design is more efficient is for the low-variability situation (where the constant-scaled criterion is used), where the value of the parameter Θ_{IBE} is already outside of the acceptable range for individual bioequivalence, and that is only when it is compared to the other 3-period design. The 2S3P design never has better asymptotic efficiency than the 4-period design we considered, even though it already has 33.3% more subjects. Although the small sample results are not as dramatic as their large sample counterparts, especially when $r^2 = 0.5$, the difference in efficiency is still quite striking, with increases in the required number of doses in the range of $20 - 80\%$. The reason why the 2S3P design is so inefficient is that estimates of the within-subject variances, σ^2_{WT} and σ^2_{WR}, use only half of the subjects. We must also assume, under the null hypothesis of non-equivalence, that the variances of the estimate of δ are different for the 2 sequences. This leads to a loss in degrees of freedom (from a Satterthwaite correction) when the confidence interval for δ is computed, as we have done for the this portion of the Howe small-sample confidence interval method (*n.b.*, neither Hyslop and Iglewicz (2001) nor Chow *et al* (2002) use the corrected degrees of freedom for the confidence interval of $\delta = \mu_T - \mu_R$). Both of these things follow from the 2S3P design not being sequence-balanced.

On one level, the comparison of total number of treatments may not be so naive, since increasing the number of doses also increases the number of plasma samples that must be assayed. Suppose that a bioequivalence study with a 2S4P crossover design requires 15 subjects per sequence, or a total of 120 doses. For each dose there may be 10 to 12 (or more) time points at which plasma samples are taken, yielding a total of 1200 to 1440 samples that must be assayed. Now, the same study conducted with a 2S3P design is likely to require $50\% - 75\%$ more doses, which increases the number of samples to be assayed accordingly. That translates to 600 to 1080 more plasma samples that must be assayed. Depending on such factors as the per unit cost of supplies, doing the assay, the time required to do the laboratory work, the cost of transporting samples from the study

site to the laboratory, *etc.*, using the 2S3P design to study individual bioequivalence may considerably increase the cost of doing a bioequivalence study.

In the comparison of designs 2S4P and 3S3P there is not a clear winner, but design 3S3P does appear to be slightly more 'efficient' (in terms of 'doses'), both for reference-scaled tests overall and when for constant-scaled tests when the intra-subject variability of the treatment formulation is less than that of the reference. Here we have essentially the situation described above, where what may be interpreted as a very slight advantage, or even the lack of a distinct disadvantage (except when the reference formulation has minimal within-subject variance), in using the 3-period design may be considered as strong support for its use when other factors are taken into account. However, the choice may not be so clear-cut. The main drawback of the 3S3P design is that, since the design parameter c_1 equals 0, the parameters σ_D^2 and σ_{WT}^2 are not identifiable. The statistical FDA guidance document for bioequivalence does not state that all parameters be estimated explicitly (FDA, 2001), but the general guidance document on bioequivalence studies requests that a point estimate of the subject-formulation interaction σ_D be provided (FDA, 2000), and there also is a strong emphasis on the constituent parts of the individual bioequivalence criteria in the literature on the subject, especially with respect to σ_D^2 (Chen *et al.*, 2000). Individual parameter estimates are also useful in any post-mortem of study results, especially when non-equivalence has been concluded. The 3S3P design can give an estimate of $\sigma_D^2 + \sigma_{WT}^2$, which then provides an upper bound on σ_D, which may be sufficient for regulatory purposes, especially if the within-subject variances are small for a given compound. An estimate of σ_D^2 can also be obtained if one is willing to assume $\sigma_{WT}^2 = \sigma_{WR}^2$. We would go further and examine the function $\hat{\sigma}_D^2 = \hat{\sigma}_I^2 - (r^2 + 0.5)\hat{\sigma}_{WR}^2$, where $r^2 = \sigma_{WT}^2/\sigma_{WR}^2$ is allowed to vary over some reasonable range, say from 0.5 to 2. This sensitivity analysis was discussed in detail in Chapter 4, and may be acceptable to regulatory agencies as an adequate assessment of the subject-formulation interaction, and it also illustrates the trade-off between the variance parameters in the aggregate criteria.

There are two main results we derived when we applied our comparison methods to population bioequivalence. The first, based on the empirical results, is that there appears

to be little that one can gain by adding periods to crossover designs for bioequivalence studies. Therefore, the standard 2×2 crossover design that has a long history of use in bioequivalence studies is probably the best for investigators who want to evaluate Θ_{PBE}. The second result is that the asymptotic results we derived are probably not very useful for assessing the relative efficiency of designs for PBE. The latter result should not be surprising, since we were examining scenarios for PBE with very small numbers of subjects.

In conclusion, we recommend that the 2S3P design not be used for the assessment of individual bioequivalence, since the alternative 3 and 4-period designs are more efficient. We also recommend that the 2S2P design–the design that is already in wide use for assessing average bioequivalence–be used when the primary objective is to evaluate population bioequivalence. The 2S4P design appears to be reasonably efficient for IBE and provides useful information about secondary individual bioequivalence parameters, while the 3S3P (extra-reference) design is a viable alternative to the 2S4P design when time and subject withdrawal are primary concerns. An investigator who is about to initial a study to evaluate individual bioequivalence should carefully weigh the pros and cons of both designs when choosing between the 2S4P and 3S3P designs.

Chapter 7

GENERALIZED CONFIDENCE INTERVALS FOR THE FDA INDIVIDUAL AND POPULATION BIOEQUIVALENCE PARAMETERS

SUMMARY

A method for deriving interval estimates for the FDA individual and population bioequivalence criteria using *generalized pivotal quantities* is proposed. These *generalized confidence intervals* (GCI) are compared to *modified large sample* (MLS) intervals, both in an example and through simulation. We find that the GCI intervals have adequate probability of covering the true parameter values, and, because of their very close relationship to *generalized p-values*, the results of Chapter 3 suggest that testing procedures using the GCI will have superior power to those using MLS. Therefore, we recommend GCI intervals over their MLS counterparts for the individual and population bioequivalence parameters.

7.1 Introduction

Ever since Schuirmann (1987) introduced his *two one-sided tests* approach for average bioequivalence, confidence intervals have been the principal tool for evaluating bioequivalence. Schuirmann's procedure is equivalent to calculating a $100(1-2\alpha)\%$ confidence interval for the mean difference $\mu_T - \mu_R$; one concludes bioequivalence if this interval is completely contained in the interval (Δ_l, Δ_u), where the deltas represent the lower and upper bounds of a prespecified interval of equivalence. The FDA has followed this paradigm in the test procedures it recommends for IBE and PBE. The test procedure for IBE was derived by Hyslop, Hsuan, and Holder (2000), and their procedure has been adapted to PBE in Appendix F of the FDA guidance document. Alternative procedures for both IBE and PBE have been proposed by Quiroz *et al* (2002) based on the *modified large sample* (MLS) methodology of constructing confidence intervals for functions of variance components for random effects and mixed models. These authors showed that the MLS approach is slightly less powerful than the FDA test for IBE, and slightly more powerful for PBE. The Hyslop (FDA) and Quiroz (MLS) approaches differ in that the FDA interval is based on a linearized form of the appropriate bioequivalence criterion, whereas the MLS interval is in the same (unitless) scale as the criterion itself. Quiroz *et al* claim that this difference as one of the advantages of their methodology. A general overview of MLS methodology is contained in the book by Burdirk and Graybill (1992).

Previous chapters developed test procedures for IBE and PBE based on the *generalized p-value* (GPV) methodology introduced by Tsui and Weerahandi (1989). The GPV methodology for IBE and PBE is departure from the methods described above in that it is based not on confidence intervals, but on the calculation of the area of an extreme region based on a particular hypothesis, as is done in calculating classical p-values. This is done using a *generalized test variable* (GTV), which takes the role of the test statistic from classical decision theory. Since, by construction, the GTV does not depend on any unknown parameters, it is very similar to a pivotal quantity in classical statistical theory. Weerahandi (1993) used this fact to define the *generalized pivotal quantity* (GPQ), which

he used to derive *generalized confidence intervals* (GCI). Chiang (2001) later developed a methodology for constructing confidence intervals for functions of variance components that he called the *method of surrogate variables* (SV). The GCI and SV methodologies are identical, although Chiang's paper is more detailed and contains comparisons to MLS and Bayesian methodologies.

In this investigation, we derive GPQs and GCIs for Θ_{IBE} and Θ_{PBE}, *i.e.*, on the same scale as Quiroz *et al* for 4-period crossover study designs. Section 7.2 introduces the statistical tools that we will use. In Section 7.3 we derive generalized pivotal quantities for Θ_{IBE} and Θ_{PBE}, and we explain how these are used to compute generalized confidence intervals. This section also contains an example of the use of GCIs with bioequivalence data. In Section 7.4 we present simulation results, comparing the coverage of GCIs to the MLS intervals of Quiroz *et al* (2002). Section 7.5 contains a short discussion and our conclusions.

7.2 Preliminaries

7.2.1 Statistical Model and Summary Statistics

In this chapter, we will construct intervals for the 2-sequence/4-period design that was examined in Chapter 3. The statistical model and summary statistics are given in Chapter 3, Section 2.

7.2.2 The Generalized Pivotal Quantity and Generalized Confidence Intervals

Weerahandi (1993) defined $R = r(\boldsymbol{X}; \boldsymbol{x}, \boldsymbol{\xi})$ as a *generalized pivotal quantity* (GPQ) if it satisfies the following two conditions

Property A: R has a probability distribution free of unknown parameters.

Property B: The *observed pivotal*, defined as $r_{obs} = r(\boldsymbol{x}; \boldsymbol{x}, \boldsymbol{\xi})$, does not depend on the nuisance parameter ζ.

We then define a set C_γ such that

$$Pr[R \in C_\gamma] = \gamma.$$

In direct analogy to the pivotal quantity method, we obtain a subset $C(\boldsymbol{X})$ of the sample space such that

$$Pr[X \in C(\boldsymbol{x})] = \gamma.$$

We further define a subset Θ_c of the parameter space Θ of θ such that

$$\Theta_c(r) = \{\theta \in \Theta | r(\boldsymbol{x}; \boldsymbol{x}, \boldsymbol{\xi}) \in C_\gamma\}.$$

The interval estimate $\Theta_c(r)$ derived from the generalize pivotal quantity R is a generalized confidence interval of θ. If we are, for example, interested in finding a 95% upper bound on R, we would find U such that

$$Pr[R \leq U | R = r(\boldsymbol{x}; \boldsymbol{x}, \boldsymbol{\xi})] = 0.95.$$

Then the interval $(-\infty, U)$ is a 95% lower generalized confidence interval on R.

In Chapter 3 we introduced GTVs that we used to assess IBE and PBE. The GPQs we will use to construct GCIs for Θ_{IBE} and Θ_{PBE} are one-to-one transformations of those GTVs. These GPQs are given in the next section.

7.3 Generalized Pivotal Quantity and Generalized Confidence Intervals for Θ_{IBE} and Θ_{PBE}

In this section we construct the generalized pivotal quantities (GPQs) for deriving the generalize confidence intervals (GCIs) for IBE and PBE. This section ends with an example where we apply both the GCI and MLS methods to bioequivalence data.

7.3.1 Individual Bioequivalence

To construct the GPQ for IBE, we use the identity $\sigma_I^2 = \sigma_D^2 + 0.5\sigma_{WT}^2 + 0.5\sigma_{WR}^2$ to reparameterized Θ_{IBE}

$$\Theta_{IBE} = \frac{\delta^2 + \sigma_I^2 + 0.5\sigma_{WT}^2 - 1.5\sigma_{WR}^2}{max(\sigma_{WR}^2, \sigma_0^2)}$$

The GPQ for computing the GCI is

$$r(X;x,\xi) = \frac{\left(d - cZ_D\sqrt{\frac{ss_I}{U_I}}\right)^2 + \frac{ss_I}{U_I} + 0.5\frac{ss_{WT}}{U_{WR}} - 1.5\frac{ss_{WR}}{U_{WR}}}{max\left(\frac{ss_{WR}}{U_{WR}}, \sigma_0^2\right)}$$

where $\xi = (\Theta_{IBE}, \delta, \sigma_I^2, \sigma_{WT}^2, \sigma_{WR}^2)$ and $\sigma_0^2 = 0.04$. To calculate the GCI, we generate a large number of random vectors $\boldsymbol{X} = (Z_D, U_I, U_{WT}, U_{WR})$, and compute $r(X;x,\xi)$ for each set of random variables given the observed values $\boldsymbol{x} = (d, ss_I, ss_{WT}, ss_{WR})$. The upper $100(1-\alpha)\%$ bound is then the $100(1-\alpha)\%$ percentile of the empirical cumulative distribution function (CDF), of $r(X;x,\xi)$. Note that comparing this upper bound to the regulatory limit θ_I provides an approximate α-level test of IBE.

7.3.2 Population Bioequivalence

As with IBE, we need to reparameterized the PBE criteria so that we can construct a GPQ.

$$\Theta_{PBE} = \frac{\delta^2 + \sigma_T^2(1-\beta^2) + 0.5\sigma_{WT}^2 - \left(\sigma_{R|T}^2 + \frac{1}{2}\sigma_{WR}^2\right)}{max\left(\sigma_T^2\beta^2 + \sigma_{R|T}^2 + 0.5\sigma_{WR}^2, \sigma_0^2\right)}$$

This is the reparameterization that we used in Chapter 3 to construct their testing procedure for Θ_{PBE}. It differs from the reparameterization used by the FDA guidance document (2001) and Quiroz *et al* (2002) because it reflects the statistical dependence between the random variables SS_T and SS_R. The GPQ for computing the GCI for PBE is

$$r(X;x,\xi) = \frac{\left(d - cZ_D\sqrt{r_D}\right)^2 + \frac{ss_T}{U_T} + 0.5\frac{ss_{WT}}{U_{WT}} - r_R}{max(r_R, \sigma_0^2)}$$

$$r_D = \frac{ss_{R|T}}{U_{R|T}} + \frac{ss_T}{U_T}\left(1 - \left(b - Z_B\sqrt{\frac{ss_{R|T}}{ss_T U_{R|T}}}\right)\right)^2$$

$$r_R = \frac{ss_T}{U_T}\left(b - Z_B\sqrt{\frac{ss_{R|T}}{ss_T U_{R|T}}}\right)^2 + \frac{ss_{R|T}}{U_T} + 0.5\frac{ss_{WR}}{U_{WR}}$$

where $\xi = (\Theta_{IBE}, \delta, \sigma_I^2, \sigma_{R|T}^2, \sigma_{WT}^2, \sigma_{WR}^2, \beta)$ and $\sigma_0^2 = 0.04$. To calculate the GCI, we generate a large number of random vectors $\boldsymbol{X} = (Z_D, Z_B, U_T, U_{R|T}, U_{WT}, U_{WR})$, and

proceed as we did for IBE, calculating the desired upper bound from the empirical CDF of $r(X;x,\xi)$ for the given values $x = (d, b, ss_T, ss_{R|T}, ss_{WT}, ss_{WR})$. And as with the upper bound we derived for Θ_{IBE}, the upper bound for PBE can also be used for hypothesis testing.

7.3.3 Example: methylphenidate

Meyer *et al* (2000) conducted a study to characterize the intrasubject variability of methylphenidate. In this study 20 healthy male subjects were randomized to one of four 4-period sequences. Using SAS we computed the following estimates for area under the time-concentration curve (AUC) for the log-transformed data

$$d = 0.07796 \quad ss_{WR} = 0.54461 \quad ss_{WT} = 0.82991$$
$$ss_I = 0.42181 \quad b = 0.75669$$
$$ss_T = 2.07850 \quad ss_{R|T} = 0.29876$$

For IBE, the 95% upper bounds for Θ_{IBE} are 1.338 and 1.3564 for the MLS and GCI intervals, respectively. The upper bounds are quite similar, and both are well within the equivalence region for IBE. For PBE, the 95% upper bounds for Θ_{PBE} are 1.93625 and 1.1043 for the MLS and GCI intervals, respectively. Hypothesis tests based on these upper bounds would reach contrary conclusions, since the regulatory bound is $\theta_P = 1.7448$. This is not surprising in light of previous results, since we have shown that test procedures based on these MLS intervals, as well as the FDA test procedure, had much lower power and larger sample size requirements than the generalized p-value (the testing analogue of the GCI) test procedure for PBE.

7.4 Simulation Studies

Simulation studies were done to assess the performance of the GCI intervals and to compare them to the MLS intervals for IBE and PBE. The tables below show the proportion of times that either the GCI or MLS interval contained the actual value of the appropriate parameter (Θ_{IBE} or Θ_{PBE}). There were 5000 simulated data sets generated

for each set of parameters. Coverage of the GCI was calculated for each simulated data set using 10000 sets of Monte Carlo points. The nominal level of coverage for all intervals is 95%. All simulations were done using SAS® Version 8.2 for Windows.

7.4.1 Individual Bioequivalence

The performance of GCI and MLS intervals were evaluated for the following parameter values: $\delta = (0.05, 0.15), \sigma_D = (0.01, 0.1, 0.15, 0.2), \sigma_{WR} = (0.15, 0.2, 0.23, 0.3, 0.5)$, $\sigma_{WT}/\sigma_{WR} = (0.8, 1.0, 1.2)$, and $n = (8, 12, 16, 20)$. These parameter combinations represent a large range of values of Θ_{IBE} relevant to the experimental situation. The results are shown in Table 2.

Table 2a. Coverage of MLS and GCI intervals for IBE

	$\delta = 0.05, \sigma_{WT}/\sigma_{WR} = 0.8$									
	$\sigma_D = 0.01$									
	$\sigma_{WR} = 0.15$		$\sigma_{WR} = 0.2$		$\sigma_{WR} = 0.23$		$\sigma_{WR} = 0.3$		$\sigma_{WR} = 0.5$	
n	MLS	GCI	MLS	GCI	MLS	GCI	MLS	GCI	MLS	GCI
8	0.966	0.975	0.967	0.982	0.963	0.975	0.960	0.974	0.961	0.974
12	0.961	0.967	0.970	0.977	0.968	0.974	0.961	0.970	0.962	0.973
16	0.963	0.964	0.967	0.972	0.971	0.976	0.966	0.973	0.960	0.970
20	0.967	0.967	0.967	0.974	0.966	0.968	0.963	0.966	0.964	0.968
	$\sigma_D = 0.10$									
	$\sigma_{WR} = 0.15$		$\sigma_{WR} = 0.2$		$\sigma_{WR} = 0.23$		$\sigma_{WR} = 0.3$		$\sigma_{WR} = 0.5$	
n	MLS	GCI	MLS	GCI	MLS	GCI	MLS	GCI	MLS	GCI
8	0.967	0.972	0.961	0.973	0.963	0.974	0.968	0.978	0.969	0.982
12	0.970	0.971	0.966	0.973	0.969	0.974	0.965	0.973	0.962	0.972
16	0.964	0.962	0.963	0.968	0.971	0.974	0.967	0.971	0.968	0.977
20	0.971	0.967	0.962	0.965	0.965	0.967	0.964	0.967	0.967	0.972
	$\sigma_D = 0.15$									
	$\sigma_{WR} = 0.15$		$\sigma_{WR} = 0.2$		$\sigma_{WR} = 0.23$		$\sigma_{WR} = 0.3$		$\sigma_{WR} = 0.5$	
n	MLS	GCI	MLS	GCI	MLS	GCI	MLS	GCI	MLS	GCI
8	0.967	0.972	0.964	0.970	0.960	0.972	0.962	0.978	0.964	0.977
12	0.973	0.971	0.961	0.962	0.965	0.972	0.961	0.970	0.962	0.971
16	0.971	0.966	0.961	0.959	0.962	0.967	0.964	0.970	0.966	0.975
20	0.970	0.965	0.959	0.957	0.963	0.966	0.962	0.967	0.962	0.969
	$\sigma_D = 0.2$									
	$\sigma_{WR} = 0.15$		$\sigma_{WR} = 0.2$		$\sigma_{WR} = 0.23$		$\sigma_{WR} = 0.3$		$\sigma_{WR} = 0.5$	
n	MLS	GCI	MLS	GCI	MLS	GCI	MLS	GCI	MLS	GCI
8	0.969	0.971	0.956	0.958	0.958	0.969	0.961	0.975	0.964	0.976
12	0.973	0.971	0.953	0.952	0.965	0.968	0.961	0.970	0.960	0.969
16	0.969	0.968	0.962	0.952	0.962	0.966	0.965	0.968	0.967	0.975
20	0.972	0.966	0.957	0.949	0.963	0.965	0.963	0.967	0.964	0.968

Table 2a. (con't) Coverage of MLS and GCI intervals for IBE

	$\delta = 0.15, \sigma_{WT}/\sigma_{WR} = 0.8$									
	$\sigma_D = 0.01$									
	$\sigma_{WR} = 0.15$		$\sigma_{WR} = 0.2$		$\sigma_{WR} = 0.23$		$\sigma_{WR} = 0.3$		$\sigma_{WR} = 0.5$	
n	MLS	GCI	MLS	GCI	MLS	GCI	MLS	GCI	MLS	GCI
8	0.965	0.964	0.968	0.969	0.965	0.967	0.966	0.971	0.964	0.973
12	0.960	0.957	0.962	0.961	0.970	0.970	0.966	0.966	0.968	0.971
16	0.961	0.956	0.961	0.959	0.968	0.968	0.967	0.968	0.966	0.968
20	0.960	0.958	0.961	0.959	0.964	0.964	0.965	0.964	0.968	0.968
	$\sigma_D = 0.10$									
	$\sigma_{WR} = 0.15$		$\sigma_{WR} = 0.2$		$\sigma_{WR} = 0.23$		$\sigma_{WR} = 0.3$		$\sigma_{WR} = 0.5$	
n	MLS	GCI	MLS	GCI	MLS	GCI	MLS	GCI	MLS	GCI
8	0.973	0.966	0.962	0.962	0.963	0.968	0.970	0.973	0.969	0.980
12	0.970	0.963	0.965	0.959	0.971	0.970	0.969	0.969	0.967	0.970
16	0.962	0.954	0.960	0.953	0.971	0.969	0.971	0.970	0.973	0.976
20	0.966	0.958	0.958	0.952	0.966	0.964	0.964	0.963	0.972	0.972
	$\sigma_D = 0.15$									
	$\sigma_{WR} = 0.15$		$\sigma_{WR} = 0.2$		$\sigma_{WR} = 0.23$		$\sigma_{WR} = 0.3$		$\sigma_{WR} = 0.5$	
n	MLS	GCI	MLS	GCI	MLS	GCI	MLS	GCI	MLS	GCI
8	0.974	0.965	0.967	0.960	0.966	0.967	0.968	0.972	0.965	0.976
12	0.981	0.964	0.966	0.952	0.973	0.966	0.968	0.967	0.967	0.969
16	0.977	0.964	0.966	0.948	0.969	0.963	0.971	0.968	0.973	0.973
20	0.976	0.961	0.965	0.951	0.966	0.961	0.968	0.964	0.964	0.964
	$\sigma_D = 0.2$									
	$\sigma_{WR} = 0.15$		$\sigma_{WR} = 0.2$		$\sigma_{WR} = 0.23$		$\sigma_{WR} = 0.3$		$\sigma_{WR} = 0.5$	
n	MLS	GCI	MLS	GCI	MLS	GCI	MLS	GCI	MLS	GCI
8	0.974	0.965	0.965	0.952	0.963	0.965	0.965	0.971	0.964	0.975
12	0.980	0.966	0.964	0.946	0.970	0.965	0.967	0.966	0.967	0.969
16	0.980	0.966	0.968	0.948	0.968	0.961	0.970	0.967	0.972	0.973
20	0.981	0.963	0.964	0.942	0.969	0.961	0.969	0.964	0.964	0.965

Table 2b. Coverage of MLS and GCI intervals for IBE

	$\delta = 0.05, \sigma_{WT}/\sigma_{WR} = 1.0$									
	$\sigma_D = 0.01$									
	$\sigma_{WR} = 0.15$		$\sigma_{WR} = 0.2$		$\sigma_{WR} = 0.23$		$\sigma_{WR} = 0.3$		$\sigma_{WR} = 0.5$	
n	MLS	GCI	MLS	GCI	MLS	GCI	MLS	GCI	MLS	GCI
8	0.968	0.978	0.968	0.979	0.963	0.975	0.963	0.974	0.962	0.975
12	0.962	0.968	0.968	0.974	0.968	0.974	0.961	0.973	0.962	0.974
16	0.966	0.968	0.965	0.969	0.970	0.976	0.966	0.973	0.962	0.969
20	0.969	0.970	0.966	0.968	0.964	0.968	0.961	0.967	0.962	0.968
	$\sigma_D = 0.10$									
	$\sigma_{WR} = 0.15$		$\sigma_{WR} = 0.2$		$\sigma_{WR} = 0.23$		$\sigma_{WR} = 0.3$		$\sigma_{WR} = 0.5$	
n	MLS	GCI	MLS	GCI	MLS	GCI	MLS	GCI	MLS	GCI
8	0.968	0.974	0.962	0.970	0.963	0.975	0.968	0.979	0.970	0.982
12	0.971	0.973	0.965	0.970	0.969	0.974	0.968	0.975	0.963	0.973
16	0.967	0.964	0.961	0.964	0.969	0.975	0.965	0.972	0.968	0.977
20	0.971	0.969	0.960	0.961	0.964	0.968	0.962	0.967	0.967	0.971
	$\sigma_D = 0.15$									
	$\sigma_{WR} = 0.15$		$\sigma_{WR} = 0.2$		$\sigma_{WR} = 0.23$		$\sigma_{WR} = 0.3$		$\sigma_{WR} = 0.5$	
n	MLS	GCI	MLS	GCI	MLS	GCI	MLS	GCI	MLS	GCI
8	0.969	0.974	0.963	0.970	0.961	0.975	0.964	0.978	0.965	0.978
12	0.973	0.973	0.961	0.961	0.967	0.972	0.961	0.974	0.963	0.972
16	0.971	0.969	0.961	0.958	0.962	0.967	0.964	0.971	0.968	0.976
20	0.970	0.964	0.958	0.955	0.963	0.968	0.961	0.967	0.961	0.968
	$\sigma_D = 0.2$									
	$\sigma_{WR} = 0.15$		$\sigma_{WR} = 0.2$		$\sigma_{WR} = 0.23$		$\sigma_{WR} = 0.3$		$\sigma_{WR} = 0.5$	
n	MLS	GCI	MLS	GCI	MLS	GCI	MLS	GCI	MLS	GCI
8	0.968	0.973	0.955	0.959	0.960	0.970	0.962	0.978	0.965	0.977
12	0.973	0.972	0.953	0.952	0.965	0.970	0.961	0.971	0.962	0.972
16	0.969	0.968	0.961	0.953	0.961	0.966	0.962	0.971	0.966	0.975
20	0.971	0.968	0.957	0.949	0.961	0.965	0.960	0.967	0.963	0.969

Table 2b. (con't) Coverage of MLS and GCI intervals for IBE

	$\delta = 0.15, \sigma_{WT}/\sigma_{WR} = 1.0$									
	$\sigma_D = 0.01$									
	$\sigma_{WR} = 0.15$		$\sigma_{WR} = 0.2$		$\sigma_{WR} = 0.23$		$\sigma_{WR} = 0.3$		$\sigma_{WR} = 0.5$	
n	MLS	GCI	MLS	GCI	MLS	GCI	MLS	GCI	MLS	GCI
8	0.971	0.968	0.970	0.968	0.963	0.968	0.965	0.971	0.964	0.972
12	0.962	0.959	0.964	0.960	0.969	0.970	0.967	0.968	0.965	0.973
16	0.963	0.960	0.961	0.956	0.969	0.969	0.969	0.970	0.965	0.969
20	0.962	0.958	0.961	0.958	0.964	0.966	0.965	0.966	0.967	0.968
	$\sigma_D = 0.10$									
	$\sigma_{WR} = 0.15$		$\sigma_{WR} = 0.2$		$\sigma_{WR} = 0.23$		$\sigma_{WR} = 0.3$		$\sigma_{WR} = 0.5$	
n	MLS	GCI	MLS	GCI	MLS	GCI	MLS	GCI	MLS	GCI
8	0.973	0.968	0.964	0.962	0.966	0.969	0.967	0.975	0.969	0.981
12	0.972	0.963	0.964	0.960	0.970	0.970	0.971	0.972	0.968	0.972
16	0.965	0.957	0.960	0.951	0.970	0.969	0.972	0.971	0.974	0.976
20	0.968	0.961	0.959	0.949	0.966	0.966	0.965	0.965	0.969	0.972
	$\sigma_D = 0.15$									
	$\sigma_{WR} = 0.15$		$\sigma_{WR} = 0.2$		$\sigma_{WR} = 0.23$		$\sigma_{WR} = 0.3$		$\sigma_{WR} = 0.5$	
n	MLS	GCI	MLS	GCI	MLS	GCI	MLS	GCI	MLS	GCI
8	0.977	0.966	0.967	0.961	0.967	0.969	0.966	0.975	0.966	0.977
12	0.981	0.967	0.963	0.951	0.971	0.969	0.969	0.969	0.964	0.969
16	0.979	0.964	0.964	0.948	0.967	0.964	0.971	0.970	0.973	0.976
20	0.977	0.963	0.963	0.950	0.966	0.962	0.967	0.964	0.963	0.966
	$\sigma_D = 0.2$									
	$\sigma_{WR} = 0.15$		$\sigma_{WR} = 0.2$		$\sigma_{WR} = 0.23$		$\sigma_{WR} = 0.3$		$\sigma_{WR} = 0.5$	
n	MLS	GCI	MLS	GCI	MLS	GCI	MLS	GCI	MLS	GCI
8	0.975	0.967	0.965	0.954	0.964	0.966	0.966	0.975	0.965	0.976
12	0.981	0.967	0.963	0.945	0.971	0.966	0.966	0.967	0.964	0.969
16	0.980	0.967	0.966	0.948	0.968	0.963	0.968	0.969	0.973	0.975
20	0.981	0.963	0.963	0.941	0.967	0.961	0.968	0.965	0.963	0.966

Table 2c. Coverage of MLS and GCI intervals for IBE

	$\delta = 0.05, \sigma_{WT}/\sigma_{WR} = 1.2$									
	$\sigma_D = 0.01$									
	$\sigma_{WR} = 0.15$		$\sigma_{WR} = 0.2$		$\sigma_{WR} = 0.23$		$\sigma_{WR} = 0.3$		$\sigma_{WR} = 0.5$	
n	MLS	GCI	MLS	GCI	MLS	GCI	MLS	GCI	MLS	GCI
8	0.972	0.979	0.966	0.977	0.963	0.974	0.962	0.973	0.962	0.975
12	0.966	0.971	0.967	0.970	0.967	0.975	0.961	0.974	0.962	0.973
16	0.971	0.968	0.963	0.965	0.967	0.976	0.965	0.973	0.961	0.970
20	0.970	0.969	0.965	0.963	0.963	0.968	0.961	0.968	0.960	0.968
	$\sigma_D = 0.10$									
	$\sigma_{WR} = 0.15$		$\sigma_{WR} = 0.2$		$\sigma_{WR} = 0.23$		$\sigma_{WR} = 0.3$		$\sigma_{WR} = 0.5$	
n	MLS	GCI	MLS	GCI	MLS	GCI	MLS	GCI	MLS	GCI
8	0.970	0.974	0.962	0.969	0.966	0.976	0.967	0.980	0.969	0.981
12	0.973	0.972	0.964	0.968	0.963	0.972	0.967	0.974	0.963	0.973
16	0.969	0.968	0.958	0.959	0.965	0.973	0.965	0.973	0.967	0.977
20	0.974	0.970	0.959	0.958	0.961	0.965	0.962	0.967	0.966	0.972
	$\sigma_D = 0.15$									
	$\sigma_{WR} = 0.15$		$\sigma_{WR} = 0.2$		$\sigma_{WR} = 0.23$		$\sigma_{WR} = 0.3$		$\sigma_{WR} = 0.5$	
n	MLS	GCI	MLS	GCI	MLS	GCI	MLS	GCI	MLS	GCI
8	0.971	0.976	0.963	0.970	0.962	0.975	0.966	0.979	0.965	0.977
12	0.976	0.974	0.962	0.960	0.968	0.973	0.962	0.974	0.962	0.972
16	0.973	0.971	0.960	0.957	0.961	0.967	0.963	0.971	0.968	0.976
20	0.973	0.967	0.958	0.954	0.960	0.967	0.960	0.966	0.960	0.968
	$\sigma_D = 0.2$									
	$\sigma_{WR} = 0.15$		$\sigma_{WR} = 0.2$		$\sigma_{WR} = 0.23$		$\sigma_{WR} = 0.3$		$\sigma_{WR} = 0.5$	
n	MLS	GCI	MLS	GCI	MLS	GCI	MLS	GCI	MLS	GCI
8	0.965	0.966	0.957	0.960	0.961	0.971	0.966	0.976	0.960	0.975
12	0.971	0.971	0.953	0.953	0.966	0.971	0.962	0.971	0.959	0.972
16	0.972	0.969	0.960	0.951	0.965	0.972	0.959	0.968	0.957	0.970
20	0.973	0.967	0.957	0.948	0.967	0.973	0.959	0.966	0.958	0.966

Table 2c. (con't) Coverage of MLS and GCI intervals for IBE

	$\delta = 0.15, \sigma_{WT}/\sigma_{WR} = 1.2$									
	$\sigma_D = 0.01$									
	$\sigma_{WR} = 0.15$		$\sigma_{WR} = 0.2$		$\sigma_{WR} = 0.23$		$\sigma_{WR} = 0.3$		$\sigma_{WR} = 0.5$	
n	MLS	GCI	MLS	GCI	MLS	GCI	MLS	GCI	MLS	GCI
8	0.977	0.971	0.968	0.967	0.965	0.969	0.966	0.972	0.964	0.973
12	0.968	0.962	0.965	0.959	0.970	0.970	0.967	0.970	0.965	0.972
16	0.968	0.960	0.962	0.954	0.969	0.969	0.969	0.970	0.964	0.968
20	0.966	0.960	0.964	0.958	0.964	0.965	0.966	0.968	0.966	0.969
	$\sigma_D = 0.10$									
	$\sigma_{WR} = 0.15$		$\sigma_{WR} = 0.2$		$\sigma_{WR} = 0.23$		$\sigma_{WR} = 0.3$		$\sigma_{WR} = 0.5$	
n	MLS	GCI	MLS	GCI	MLS	GCI	MLS	GCI	MLS	GCI
8	0.976	0.970	0.962	0.959	0.965	0.969	0.968	0.976	0.970	0.981
12	0.974	0.967	0.964	0.960	0.971	0.970	0.971	0.974	0.966	0.973
16	0.970	0.959	0.961	0.948	0.970	0.969	0.971	0.973	0.971	0.976
20	0.970	0.963	0.958	0.946	0.965	0.964	0.966	0.963	0.968	0.972
	$\sigma_D = 0.15$									
	$\sigma_{WR} = 0.15$		$\sigma_{WR} = 0.2$		$\sigma_{WR} = 0.23$		$\sigma_{WR} = 0.3$		$\sigma_{WR} = 0.5$	
n	MLS	GCI	MLS	GCI	MLS	GCI	MLS	GCI	MLS	GCI
8	0.977	0.968	0.966	0.962	0.968	0.970	0.969	0.975	0.966	0.977
12	0.980	0.967	0.967	0.953	0.970	0.970	0.969	0.970	0.964	0.969
16	0.979	0.966	0.965	0.947	0.966	0.965	0.970	0.970	0.974	0.976
20	0.979	0.964	0.964	0.950	0.967	0.962	0.968	0.965	0.963	0.966
	$\sigma_D = 0.2$									
	$\sigma_{WR} = 0.15$		$\sigma_{WR} = 0.2$		$\sigma_{WR} = 0.23$		$\sigma_{WR} = 0.3$		$\sigma_{WR} = 0.5$	
n	MLS	GCI	MLS	GCI	MLS	GCI	MLS	GCI	MLS	GCI
8	0.975	0.968	0.964	0.956	0.964	0.969	0.967	0.975	0.966	0.977
12	0.981	0.968	0.962	0.944	0.969	0.968	0.968	0.969	0.964	0.969
16	0.983	0.967	0.967	0.948	0.967	0.964	0.970	0.971	0.975	0.976
20	0.980	0.964	0.962	0.941	0.966	0.962	0.967	0.966	0.963	0.965

The GCI intervals appear to do well at maintaining the nominal coverage, with only 6 out of 480 parameter combinations in Table 2 having observed coverage values less than 94.7%. These 6 cases all occur when $\sigma_{WR} = 0.2$, the changeover point from reference to constant-scaling. In Chapter 3, we observed that the GPV (the testing analogue of the GCI) was sometimes slightly anti-conservative for IBE when $\sigma_{WR} = 0.2$. The MLS intervals maintain the nominal coverage probability throughout Table 2. As a down side of their solid coverage, test procedures for IBE that use MLS intervals tend to have less power than the Hyslop/FDA and GPV procedures (Quiroz *et al.*, 2002).

7.4.2 Population Bioequivalence

Table 3 shows simulation results for GCI and MLS intervals for PBE for the following sets of parameters: $\delta = (0.05, 0.15), \rho = (0.7, 0.8, 0.9), \sigma_{WT} = \sigma_{WR} = (0.15, 0.23, 0.3, 0.5), \sigma_{BT}/\sigma_{WT} = \sigma_{BR}/\sigma_{WR} = (1, 2)$, and $n = (8, 12, 16, 20)$. Note that the value of Θ_{PBE} for these parameter combinations is less than 1, which puts them well within the region of equivalence.

Table 3a. Estimated coverage of procedures MLS and GPV for PBE

		$\delta = 0.05$							
		$\sigma_{BT} = \sigma_{BR} = 0.15$		$\sigma_{BT} = \sigma_{BR} = 0.3$		$\sigma_{BT} = \sigma_{BR} = 0.23$		$\sigma_{BT} = \sigma_{BR} = 0.46$	
		$\sigma_{WT} = \sigma_{WR} = 0.15$				$\sigma_{WT} = \sigma_{WR} = 0.23$			
n	ρ	MLS	GCI	MLS	GCI	MLS	GCI	MLS	GCI
8	0.7	0.9710	0.9680	0.9810	0.9632	0.9682	0.9622	0.9714	0.9512
	0.8	0.9746	0.9642	0.9878	0.9564	0.9664	0.9558	0.9874	0.9558
	0.9	0.9790	0.9620	0.9944	0.9512	0.9758	0.9622	0.9916	0.9572
12	0.7	0.9650	0.9650	0.9736	0.9474	0.9642	0.9594	0.9786	0.9544
	0.8	0.9696	0.9570	0.9896	0.9506	0.9730	0.9616	0.9886	0.9526
	0.9	0.9772	0.9572	0.9936	0.9482	0.9742	0.9576	0.9950	0.9572
16	0.7	0.9570	0.9490	0.9792	0.9508	0.9680	0.9602	0.9766	0.9528
	0.8	0.9674	0.9532	0.9888	0.9548	0.9686	0.9562	0.9890	0.9574
	0.9	0.9726	0.9542	0.9956	0.9530	0.9776	0.9612	0.9962	0.9534
20	0.7	0.9750	0.9620	0.9810	0.9532	0.9618	0.9520	0.9786	0.9504
	0.8	0.9704	0.9556	0.9882	0.9508	0.9702	0.9580	0.9900	0.9530
	0.9	0.9784	0.9590	0.9944	0.9510	0.9758	0.9564	0.9954	0.9570
		$\sigma_{BT} = \sigma_{BR} = 0.3$		$\sigma_{BT} = \sigma_{BR} = 0.6$		$\sigma_{BT} = \sigma_{BR} = 0.5$		$\sigma_{BT} = \sigma_{BR} = 1.0$	
		$\sigma_{WT} = \sigma_{WR} = 0.3$				$\sigma_{WT} = \sigma_{WR} = 0.5$			
n	ρ	MLS	GCI	MLS	GCI	MLS	GCI	MLS	GCI
8	0.7	0.9688	0.9628	0.9774	0.9534	0.9656	0.9614	0.9790	0.9532
	0.8	0.9704	0.9590	0.9840	0.9520	0.9732	0.9648	0.9828	0.9502
	0.9	0.9744	0.9570	0.9946	0.9534	0.9766	0.9654	0.9936	0.9522
12	0.7	0.9688	0.9614	0.9774	0.9526	0.9686	0.9636	0.9766	0.9548
	0.8	0.9716	0.9600	0.9882	0.9540	0.9724	0.9618	0.9832	0.9528
	0.9	0.9772	0.9614	0.9934	0.9526	0.9794	0.9606	0.9948	0.9528
16	0.7	0.9698	0.9586	0.9826	0.9592	0.9654	0.9570	0.9786	0.9496
	0.8	0.9742	0.9584	0.9876	0.9544	0.9722	0.9576	0.9876	0.9576
	0.9	0.9764	0.9568	0.9934	0.9512	0.9816	0.9644	0.9958	0.9492
20	0.7	0.9686	0.9606	0.9786	0.9524	0.9664	0.9572	0.9802	0.9484
	0.8	0.9688	0.9552	0.9856	0.9476	0.9704	0.9556	0.9842	0.9442
	0.9	0.9732	0.9540	0.9948	0.9508	0.9772	0.9592	0.9970	0.9490

Table 3b. Estimated coverage of procedures MLS and GPV for PBE

		$\delta = 0.15$							
		$\sigma_{BT} = \sigma_{BR} = 0.15$		$\sigma_{BT} = \sigma_{BR} = 0.3$		$\sigma_{BT} = \sigma_{BR} = 0.23$		$\sigma_{BT} = \sigma_{BR} = 0.46$	
		$\sigma_{WT} = \sigma_{WR} = 0.15$				$\sigma_{WT} = \sigma_{WR} = 0.23$			
n	ρ	MLS	GCI	MLS	GCI	MLS	GCI	MLS	GCI
8	0.7	0.9640	0.9490	0.9810	0.9566	0.9664	0.9556	0.9716	0.9478
	0.8	0.9668	0.9432	0.9852	0.9514	0.9670	0.9538	0.9872	0.9556
	0.9	0.9752	0.9496	0.9942	0.9492	0.9752	0.9544	0.9938	0.9582
12	0.7	0.9700	0.9510	0.9744	0.9414	0.9632	0.9534	0.9776	0.9548
	0.8	0.9708	0.9410	0.9890	0.9464	0.9682	0.9518	0.9892	0.9524
	0.9	0.9738	0.9404	0.9928	0.9502	0.9742	0.9500	0.9952	0.9554
16	0.7	0.9540	0.9340	0.9784	0.9470	0.9656	0.9516	0.9762	0.9506
	0.8	0.9644	0.9388	0.9880	0.9534	0.9684	0.9482	0.9888	0.9584
	0.9	0.9708	0.9428	0.9956	0.9516	0.9746	0.9518	0.9956	0.9526
20	0.7	0.9680	0.9450	0.9790	0.9492	0.9606	0.9462	0.9776	0.9486
	0.8	0.9692	0.9430	0.9850	0.9526	0.9706	0.9536	0.9884	0.9518
	0.9	0.9784	0.9456	0.9936	0.9516	0.9730	0.9496	0.9954	0.9554
		$\sigma_{BT} = \sigma_{BR} = 0.3$		$\sigma_{BT} = \sigma_{BR} = 0.6$		$\sigma_{BT} = \sigma_{BR} = 0.5$		$\sigma_{BT} = \sigma_{BR} = 1.0$	
		$\sigma_{WT} = \sigma_{WR} = 0.3$				$\sigma_{WT} = \sigma_{WR} = 0.5$			
n	ρ	MLS	GCI	MLS	GCI	MLS	GCI	MLS	GCI
8	0.7	0.9652	0.9582	0.9792	0.9544	0.9626	0.9588	0.9788	0.9552
	0.8	0.9712	0.9576	0.9852	0.9512	0.9744	0.9638	0.9832	0.9512
	0.9	0.9722	0.9554	0.9956	0.9528	0.9758	0.9630	0.9936	0.9518
12	0.7	0.9668	0.9540	0.9780	0.9524	0.9664	0.9604	0.9768	0.9528
	0.8	0.9714	0.9558	0.9890	0.9542	0.9694	0.9600	0.9838	0.9530
	0.9	0.9762	0.9584	0.9928	0.9518	0.9796	0.9604	0.9940	0.9516
16	0.7	0.9672	0.9546	0.9826	0.9566	0.9628	0.9554	0.9770	0.9498
	0.8	0.9700	0.9558	0.9876	0.9548	0.9712	0.9572	0.9874	0.9566
	0.9	0.9776	0.9552	0.9944	0.9502	0.9808	0.9628	0.9964	0.9504
20	0.7	0.9652	0.9528	0.9802	0.9492	0.9630	0.9556	0.9788	0.9494
	0.8	0.9680	0.9520	0.9864	0.9474	0.9700	0.9552	0.9856	0.9452
	0.9	0.9744	0.9532	0.9956	0.9478	0.9752	0.9574	0.9970	0.9502

When $\delta = 0.05$, the GCI intervals appear to do fairly well in maintaining the nominal coverage probability, with only 1 out of 96 coverage values less than 94.7%. When δ increases to 0.15, there are more cases where the coverage probabilities of the GCI are slightly lower than the nominal level. With 12 out of 96 values less than 94.7%, and 2 values less than 94%. On the other hand, the coverage of the MLS intervals is quite high, sometimes exceeding 99.5%. Neither result is surprising. Chapter 3 showed that the GPV, the testing analogue of the GCI, produced slightly anti-conservative tests of PBE, especially as the absolute value of δ increased. They also showed that MLS intervals for PBE produced very conservative test procedures, which had a detrimental effect on the power of these procedures.

Table 3 addresses the coverage of intervals for Θ_{PBE} when the hypothesis of equivalence, which here is the alternative hypothesis, is true. Since we are also interested in the performance of these intervals when the null hypothesis of non-equivalence is true, we present Table 4 for completeness, which shows simulation results for parameter combinations that produce values of Θ_{PBE} greater than 2, well outside the region of equivalence.

Table 4. Estimated coverage of procedures MLS and GPV for PBE

		$\delta = 0.15$							
		$\sigma_{BT} = 0.3$ $\sigma_{BR} = 0.15$ $\sigma_{WT} = 0.15$ $\sigma_{WR} = 0.1$		$\sigma_{BT} = 0.46$ $\sigma_{BR} = 0.23$ $\sigma_{WT} = 0.23$ $\sigma_{WR} = 0.2$		$\sigma_{BT} = 0.6$ $\sigma_{BR} = 0.3$ $\sigma_{WT} = 0.3$ $\sigma_{WR} = 0.2$		$\sigma_{BT} = 1.0$ $\sigma_{BR} = 0.5$ $\sigma_{WT} = 0.5$ $\sigma_{WR} = 0.4$	
n	ρ	MLS	GCI	MLS	GCI	MLS	GCI	MLS	GCI
8	0.7	0.9528	0.9406	0.9638	0.9366	0.9756	0.9424	0.9728	0.9412
	0.8	0.9534	0.9304	0.9790	0.9444	0.9824	0.9440	0.9790	0.9464
	0.9	0.9572	0.9326	0.9832	0.9428	0.9894	0.9446	0.9874	0.9468
12	0.7	0.9524	0.9306	0.9724	0.9470	0.9758	0.9438	0.9708	0.9432
	0.8	0.9636	0.9406	0.9816	0.9470	0.9854	0.9474	0.9798	0.9482
	0.9	0.9610	0.9356	0.9870	0.9522	0.9896	0.9496	0.9872	0.9452
16	0.7	0.9590	0.9446	0.9708	0.9470	0.9804	0.9502	0.9746	0.9456
	0.8	0.9570	0.9400	0.9834	0.9508	0.9852	0.9476	0.9810	0.9534
	0.9	0.9662	0.9416	0.9870	0.9468	0.9912	0.9480	0.9900	0.9440
20	0.7	0.9572	0.9454	0.9706	0.9466	0.9770	0.9472	0.9748	0.9412
	0.8	0.9610	0.9406	0.9796	0.9450	0.9838	0.9430	0.9774	0.9430
	0.9	0.9584	0.9352	0.9878	0.9502	0.9910	0.9468	0.9894	0.9484

The coverage of the GCI in Table 4 is slightly lower than the nominal coverage probability, with most values less than 95%, but only 5 out of 48 less than 94%, and none less than 93%. On the other hand, the MLS intervals all maintain the nominal coverage, although some coverage probabilities are very high, sometimes in excess of 99%. Again, neither result is surprising, in light of the results of Chapter 3. We will discuss these result further in the next section.

7.5 Discussion

The generalized confidence intervals (GCI) for IBE appear to do quite well in maintaining the nominal coverage probability, and to be comparable to the corresponding MLS intervals of Quiroz *et al.* (2002) with respect to this coverage. The only circumstance where the coverage of the GCI intervals dips very slightly below the nominal level is when

$\sigma_{WR} = 0.2$, the changeover point from constant to reference scaling. GCI intervals for PBE can tend to have coverage slightly below the nominal level, especially as δ starts to deviate from 0. We saw in Table 3 that the coverage of the deteriorates slightly when δ goes from 0.05 to 0.15. However, it seems that values of δ greater than 0.1 are rare in bioequivalence studies (Henney, 1999), at least for generic products that have gained regulatory approval in the past 2 decades. Other than the obvious selection bias, this observation suggests that the mean is a relatively easy mark to hit for a competent chemist at a generic manufacturer, even though he does not know the the explicit details of how the innovator (*i.e.*, reference) product was compounded. Taking this and the fact that some sort of average bioequivalence is needed in order to conclude PBE (FDA, 2001), the practical significance of the GCIs for PBE falling slightly below the nominal level may be attenuated. On the other hand, the MLS intervals are very solid when it comes to maintaining nominal coverage. But, as we will see, their application in testing procedures for bioequivalence can tend to be overly conservative, especially with respect to PBE.

Having primarily examined interval coverage in this paper, the question of comparative interval length comes to mind. We address this question by referring to 2 other sources. The first is Chiang (2001), whose method of surrogate variables (SV) is exactly the same as the procedure we have outlined for GCIs. Table 8 in Chiang shows ratios of interval lengths for equal-tailed SV (GCI) and MLS intervals for a parameter that is estimated by taking the difference of means squares, similar to the constant-scaled criteria for IBE and PBE. This table shows that the 2 methodologies produce intervals for this parameter that are almost identical in length. Table 10 in Chiang shows ratios of intervals lengths for a ratio parameter that is estimated by a difference of mean squares is the numerator and has a mean square in the denominator, similar to the reference-scaled criteria for IBE and PBE. Here there appears to be a tendency for the equal-tailed SV (GCI) interval to be slightly shorter than its MLS counterpart. These results suggest that the GCI intervals for bioequivalence should at least be no longer than the MLS intervals. Even more suggestive are the power results we presented in Chapter 3 (note that the generalized p-value (GPV) is the hypothesis testing analog of the GCI). There we showed that the

GPV had slightly better power than the MLS procedure for IBE, and that the GPV was significantly more powerful than the MLS for PBE. The large power difference between the GPV and both the MLS and FDA tests for PBE appears to be due to the latter 2 tests ignoring the statistical dependence between the sufficient statistics in this problem. This leads to intervals that are excessively long for PBE, as we saw in the example given above. Since the generalized test variable (GTV) of the GPV and the generalized pivotal quantity of the GCI are closely related (they are almost always either identical or one a one-to-one function of the other), we expect GCI intervals to shorter than MLS intervals in critical cases for IBE, and in general for PBE.

In our experience with GPVs in the previous chapters, we have found that conclusions drawn from examining 4 period study designs tend to generalize to 3 period study designs and even to the evaluation of PBE in a 2×2 study design. Therefore, we expect our results for GCIs also to generalize to other study designs. This assertion also seems to be confirmed by some special cases we have looked at for other designs used to assess bioequivalence.

In this chapter, we have presented GPQs and GCIs for IBE and PBE are in the same scale, respectively, as Θ_{IBE} and Θ_{PBE}. We can as easily generate GPQs and GCIs for the linearized criteria of the FDA guidance document, or for any arbitrary scale that we choose. As noted above, Quiroz *et al* argued that an advantage of the MLS confidence intervals over the intervals derived by Hyslop *et al.* is that their MLS intervals, as our GCIs, are in the same scale as Θ_{IBE} and Θ_{PBE}. We would temper this assertion with the following observations. The scale of the linearized criteria is the same as the scale of the variance of the mean difference of the log-transformed data. This may lend some intuitive interpretation to intervals based on the linearized criteria–often the variance of the mean on a log-transformed scale is interpreted as a squared coefficient of variation (CV^2)–although it should be noted that the length of these intervals is a function of the treatment variances, which is probably why the FDA chose to implement scaled criteria for IBE and PBE. The GCI and MLS intervals, on the other hand, are on a unitless scale. Quiroz *et al* appear to take for granted that this scale has 'more meaningful interpretations

(pg. 1840).' The individual and population bioequivalence criteria are constructed by taking a distance metric (Schall and Luus, 1992) and dividing by a scale factor based on the variation of the reference formulation. Consider the statement, 'the distance between formulations is less than or equal to 1.5 times the reference variance with 95% certainty.' Because of the novelty of the the parameters in question, we are not sure what the meaning of this statement would be to a pharmacologist, other than whether 1.5 is greater or less than the regulatory upper bound. Therefore, it is not necessarily given that one scale is preferable to any other. However, as investigators and regulatory agencies gain greater experience with IBE and PBE, it is possible that the intervals that both we and Quiroz *et al.* have proposed will also gain intuitive meaning, potentially making them useful tool for assessing IBE and PBE. It should also be noted that, since the GPQ for the linearized criteria is a one-to-one function of the GPQ that we present, we can generate GCIs with the exact same coverage probabilities as those we computed above on the same scale as the Hyslop/FDA intervals.

In conclusion, we would recommend the GCI over MLS intervals on two counts: the coverage of the GCI intervals appears to be adequate, and the GCI intervals are analogous to testing procedures, namely the GPV tests, that are more powerful than tests derived from the MLS intervals. We feel that the slight anti-conservativeness of the GCI intervals for PBE is preferable to the overly conservative MLS intervals, given the significant power advantage of the former over the latter.

Chapter 8

CONCLUSION

8.1 Summary of Primary Conclusions

In this dissertation, we have derived new tests for the FDA individual and population bioequivalence parameters using the generalized p-value methodology, we have used generalized confidence intervals to construct interval estimates for these same parameters, and we have shown that these methods produce viable ways to evaluate these same parameters. These methods have been derived for the most likely experimental designs for evaluating bioequivalence, and have been rigorously assessed for a large part of the likely parameter space. In the course of these investigations, we have addressed the following important statistical and practical issues:

1. We have derived a test that is not discontinuous when one switches from constant to reference scaling in the bioequivalence criteria.

2. We have derived a test that is more powerful than the test recommended by the FDA for population bioequivalence. This test was constructed by taking into account the dependence structure of the summary statistics, going beyond the assumption that most researchers in field have made: that the statistics used to estimate the variance component functions are independent.

3. We have examined the efficiency, both asymptotically and empirically, of the most likely bioequivalence study designs. We have also presented a strong case for using the sequence-balanced 3-period crossover design instead of the 3-period design in the FDA guidance.

4. We have suggested strategies for addressing some of the issues related to particular designs, *e.g.*, the non-identifiability of σ_D^2 and σ_{WT}^2 in the 3S3P design.

A somewhat unusual characteristic of the FDA bioequivalence parameters is that they are functions of both fixed effects and variance components. From the success of the generalized p-value and the generalized confidence interval in assessing these parameters, it appears that these methods should be useful for assessing general functions of fixed and variance parameters from linear mixed models. We believe that this dissertation represents significant contributions to a very timely and important topic.

8.2 Topics for Future Study

8.2.1 Theoretical properties of GPVs and GCIs

Our investigations have concentrated on the properties of generalized p-values and generalized confidence intervals for specific functions of the parameters of a linear mixed model for relatively small samples. Other investigations into similar problems have been conducted by other researchers, such as Zhou and Mathew (1994) and Chiang (2001). Zhou and Mathew (1996) stated that 'the test based on the generalized p value [sic] does have theoretical justifications, and many theoretical properties are known.' However, proofs for the results that they cite are not available in the published litterature. Weerahandi gives other properties in his book (1995) where he relies on the results that Meng (1994) derived for posterior predictive p-values, *e.g.*, that the size of a GPV test is bounded by 2α, where α is the nominal size of the test. The results of Meng appear only to be applicable when there is a Bayesian test that corresponds to the GPV test, as with the Behren-Fisher problem.

As we have noted, GPV tests have many similarities to both fiducial and Bayesian methods. The link to fiducial appears to be very strong, but the lack of a general theory of fiducial inference–fiducial arguments for a specific inferential approach are usually made on a case by case basis–weakens the usefulness of this connection. A stronger connection

to Bayesian methods would endow the GPV and GCI with many desirable properties, but at present this connection is somewhat tenuous. Establishing a stronger, more general link between the GPV and the PPV, *i.e.*, whether and when one can derive a discrepancy variable and specify a prior distribution so that the 2 methods give identical answers, would be extremely useful in documenting the theoretical properties of the GPV and GCI.

8.2.2 Analytic approximations

There is little or no literature on the actual sampling (or even approximate) distributions of the generalized test variable and generalized pivotal quantity. With the availability of relatively inexpensive computing power, one can calculate GPVs and GCIs for a particular set of data practically instantaneously. But the study of statistical properties requires simulation studies that use much more time and computing power. Therefore, analytic methods such as saddlepoint approximations would be very useful. While good second-order (or higher) approximations would not be likely to replace the Monte Carlo methods for computing GPVs and GCIs for individual data sets, they could prove indispensable for studying properties such as statistical power and the median length of GCIs. Such approximations could also be very useful in studying how variations of underlying parameter values and sufficient statistics effect the properties of GPV tests, *e.g.*, why does the size of test increase when σ_{WR} is close to $\sigma_0 = 0.2$. They could also be useful for exploring the theoretical properties of the GPV and GCI, assisting in the investigations in the previous subsection.

8.2.3 Interval estimates for the subject-by-formulation interaction

One of the problems with the disaggregate criteria approach for individual is that the distribution of the subject-by-formulation interaction, σ_D^2, is generally intractable (Chow and Liu, 2000). Fortunately, all of the methods that we have discussed–the Howe method (1974), the modified large sample method (Burdirk and Graybill, 1992), and the GCI–can be used to compute approximate intervals for this parameter. Even the

methodology of Tukey (1951) and Williams (1962) can be adapted to this problem. Given the importance of this parameter to the FDA, a study of the relative merits of each of the above methodologies with respect to the interval estimation of σ_D^2 is in order. And since the variability of estimates of σ_D^2 appears to be a function of the within-subject variabilites (Endreyi, Taback, and Tothfalusi , 2000), study of an appropriately scaled version of the subject-by-formulation interaction might also be merited. One intriguing candidate for a scaled subject-by-formulation interaction is

$$\frac{\sigma_D^2}{\sigma_{WT}^2 + \sigma_{WR}^2}$$

a function for which an 'exact' interval estimate exists.

8.2.4 Population equivalence

Placebo-controlled clinical trials have historically been the standard route for evaluating the efficacy of novel therapies. But the proliferation of *me too* products (*e.g.*, a novel calcium-channel blocker currently in clinical development for the treatment of hypertension is a *me too* drug, since there are several calcium-channel blockers that have already been approved and marketed for this indication) has led to concerns, especially in the international community, concerning the standard of care in and appropriateness of some placebo-controlled clinical trials. More and more, developers of *me too* drugs are being encouraged to use active controls instead of placebos in their clinical trials. In an active-control trial, the goal is to show the *non-inferiority* of the novel product with respect to the appropriate active control. The hypothesis of non-inferiority is the union of the hypotheses of equivalence and superiority. Many of the issues of non-inferiority trials are discussed by Hauck and Anderson (1999). One of the issues they briefly discuss is *population equivalence*. This type of equivalence could be assessed using the parameter

$$\frac{\delta^2 + \sigma_T^2 - \sigma_R^2}{max(\sigma_R^2, \sigma_0^2)},$$

where δ is the mean treatment difference, R is the active comparator, and T is the novel therapy. We could conclude population equivalence if we can show that with some specified

assurance that this parameter is less than some upper bound that would probably have to be determined for each case individually. Note that, in all likelihood, the data will come from a study with a parallel-group design: in this case, the dependency between estimates of σ_T^2 and σ_R^2 is no longer an issue. In this approach, slight superiority of the active comparator (*i.e.*, $\delta < 0$) could be offset if the novel therapy has less patient-to-patient variability. The obvious disadvantage of this approach is that we want to allow (and even encourage) larger positive values of δ, *i.e.*, allow the novel therapy to be superior. An alternative parameter might be

$$\frac{sgn(-\delta)\delta^2 + \sigma_T^2 - \sigma_R^2}{max(\sigma_R^2, \sigma_0^2)}.$$

The emphasis on the tradeoff between component parameters is even stronger here than it was for individual and population bioequivalence. However, by manipulating the sign of δ^2 we lose the interpretation of the numerator as a distance metric. There is also a question of whether this parameter provides any useful information beyond what can be extracted by assessing estimates of the mean difference parameter δ.

8.2.5 More general dependency structures

Both Zhou and Mathew (1994) and Chiang (2001) only considered problems in which the sufficient statistics were independently distributed. To derive our test for population bioequivalence, we needed to consider dependent sufficient statistics, and the multivariate generalization of the χ^2 distribution, the Wishart. McNally (2002, see Appendix C) discussed a similar generalization of the methodology of Wolfinger and Kass (2000) also applied to PBE. The dependency structure in this problem is often called *exchangable* or *compound symmetry* in mixed models literature. Extending our methods to more general dependency structure, such as *first-order autoregressive* or *unstructured* could be very useful in the analysis of longitudinal data, especially in studies with relatively small numbers of subjects.

8.2.6 Other topics

Here are some other topics for which the GPV methodology may prove useful:

1. Multiple comparison and testing situations, especially when the results of tests are dependent on each other.

2. Meta-analysis: reaching a single conclusion based on summary results from two or more investigations.

3. Multiple imputation for missing data, along the lines discussed in Meng (1994).

Appendix A

REFERENCES

Anderson S, Hauck WW (1990). Consideration of individual bioequivalence. *Journal of Pharmacokinetics and Biopharmaceutics*; **18**:259–273.

Billingsley P (1995). *Probability and Measure*, 3rd ed., John Wiley & Sons, New York.

Brown LD, Hwang JTG, Munk A (1997). An unbiased test for the bioequivalence problem. *The Annals of Statistics*, **26**:2345–2367.

Bulmer MG (1957). Approximate confidence intervals for components of variance. *Biometrika*, **44**:159–167.

Burdick RK, Graybill FA (1992). *Confidence intervals on variance components*. Marcel Dekker, New York, 1992.

Chen ML, Patnaik R, Hauck WW, Schuirmann DJ, Hyslop T, Williams R (2000). An individual bioequivalence criterion: regulatory considerations. *Statistic in Medicine*. **19**:2821–2842.

Chiang AKL (2001). A simple general method for constructing confidence intervals for functions of variance components. *Technometrics*, **43**:356–367.

Chinchilli VM, Esinhart JD (1996). Design and analysis of intra-subject variability in cross-over experiments. *Statistics in Medicine*; **15**:1619–1634.

Chow SC (1999). Individual bioequivalence–a review of the FDA draft Guidance. *Drug Information Journal*, **33**:435–444.

Chow SC, Liu JP. *Design and Analysis of Bioavailability and Bioequivalence Studies*, Marcel Dekker, New York, 2000.

Chow SC, Shao J, Wang H (2002). Individual bioequivalence testing under 2 × 3 designs. *Statistics in Medicine*, **21**:629–648.

Efron B, Tibshirani RJ. *An Introduction to the Bootstrap*, Chapman & Hall, New York, 1993.

Ekbohm G, Melander H (1989). The subject-by-formulation interaction as a criterion for interchangeability of drugs. Biometrics; **45**:1249–1254.

Endrenyi L, Taback N, Tothfalusi L (2000). Properties of the estimated variance component for subject-by-formulation interaction in studies of individual bioequivalence. *Statistics in Medicine*, **19**:2867–2878.

Gould AL (2000). A practical approach for evaluating population and individual bioequivalence. *Statistics in Medicine*, **19**:2721-2740.

Graybill FA, Wang CM (1980). Confidence intervals on nonnegative linear combinations of variances. *Journal of the American Statistical Association*; **75**: 869–873.

Hauck WW (1996). Mean difference vs. variability reduction: tradeoffs in aggregate measures for individual bioequivalence. *Int. J. Clin. Pharmacol. Ther. Toxicol.*; **34**:535-541.

Hauck WW, Anderson S (1991). Individual bioequivalence: what matters to the patient. *Statistics in Medicine*, **10**:959–959.

Hauck WW, Anderson S (1999). Some issues in the design and analysis of equivalence trials. *Drug Information Journal*, **33**:109–118.

Haynes JD (1981). Statistical simulation study of new proposed uniformity requirement for bioequivalencey studies, *Journal of Pharmaceutical Sciences*, **70**:673–675.

Henney JE (1999). Review of Generic Bioequivalence Studies, *Journal of the American Medical Association*, **282**:21

Howe WG (1974). Approximate confidence limits on the mean of $X + Y$ where X and Y are two tabled independent random variables, *Journal of the American Statistical Association*, **69**:789–794.

Hsu JC, Hwang JTG, Liu H-K, Ruberg SJ (1994). Confidence intervals associated with tests for bioequivalence, *Biometrika*, **81**:103–114.

Hsuan F (2000). Some statistical considerations on the FDA draft guidance for individual bioequivalence. *Statistics in Medicine*, **19**:2879–2884.

Hyslop T, Hsuan F, Holder DJ (2000). A small sample confidence interval approach to assess individual bioequivalence, *Statistics in Medicine*; **19**:2885–2897.

Hyslop T, Iglewicz B (2001). Alternative cross-over designs for individual bioequivalence. *Proceedings of the Biopharmaceutical Section, American Statistical Association, in press.*

Jeffreys H (1967). *Theory of Probability*, 3rd ed. Oxford University Press.

Jones B, Kenward MG (1989). *Design and analysis of cross-over trials.* Chapman and Hall: London, 1989.

Lehmann EL, Casella G (1998). *Theory of Point Estimation*, Springer-Verlag, New York.

Lu T-FC, Graybill FA, Burdirk RK (1989). Confidence intervals of the ratio of expected mean squares $(\theta_1 - d\theta_2)/\theta_3$. *Journal of Statistical Planning and Inference*, **21**:179–190.

Mathew T, Webb D (2002). Mixed model inference for Army test and evaluation. *Submitted for publication.*

McNally RJ (2002). Bayesian analysis of population bioequivalence using the independence chain algorithm in PROC MIXED. Annual Conference of the Pharmaceutical Industry SAS® Users Group (PharmaSUG), **20**:397–401.

Meng XL (1994). Posterior predictive p-values. *The Annals of Statistics*, **22**:1142–1160.

Meyer MC, Straughn AB, Jarvi EJ, Patrick KS, Pelsor FR, Williams RL, Patnaik R, Chen ML, Shah VP (2000). Bioequivalence of methylphenidate immediate-release tablets using a replicated study design to characterize intrasubject variability. *Pharmaceutical Research*, **17**:381–384.

Moriguti S (1954). Confidence limits for a variance components. *Rep. Stat. Appl. Res., JUSE* **3**:7–19.

Pharmaceutical Sciences Advisory Committee (1999). Transcript of the open meeting

of the Pharmaceutical Sciences Advisory Committee, 23 September 1999. *http://www.fda.gov/ohrms/dockets/ac/cder99t.htm*

Quiroz J, Ting N, Wei GCG, Burdick RK (2002). Alternative Confidence Intervals for the Assessment of Bioequivalence in Four-Period Crossover Designs, *Statistics in Medicine*, **21**:1825–1847.

Patnaik PB (1949). The non-central χ^2- and F-distributions and their applications. *Biometrika*, **36**:215–227.

Rubin DB (1984). Bayesianly justifiable and relevant frequency calculations for the applied statistician. *Annals of Statistics*, **12**:1151–1172.

Satterthwaite FE (1946). An approximate distribution of estimates of variance component. *Biometrics Bulletin*, **2**:110–114.

Schall R (1995). Assessment of individual and population bioequivalence using the probability that bioavailabilities are similar, *Biometrics*; **51**:615–626.

Schall R. Luus HG (1993). On population and individual bioequivalence, *Statistics in Medicine*; **12**:1109–1124.

Schuirmann DJ (1987). 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* 1990; **15**:657–680.

Sheiner LB (1992). Bioequivalence revisited. *Statistics in Medicine*; **11**:1777–1788.

Shumaker RC, Metzler CM (1998). The phenytoin trial is a case study of 'individual' bioequivalence. *Drug Information Journal*; **32**:1063-1072.

Steel RGD, Torrie JH, Dickey D (1997). *Principles and Procedures of Statistics: A Biometrical Approach*, 3rd ed., McGraw-Hill, New York.

Schall R, Luus HG (1993). On population and individual bioequivalence, *Statistics in Medicine*, **12**:1109–1124.

Ting N, Burdirk RK, Graybill FA, Jeyaratnam S, Lu T-FC (1990). Confidence intervals on linear combinations of variance components that are unrestricted in sign. *Journal of Statistical Computation and Simulation*, **35**:135–143.

Tsui KW, Weerahandi S (1989). Generalized p-values in significance testing of

hypotheses in the presence of nuisance parameters, *Journal of the American Statistical Association*; **84**:602–607.

Tukey JW (1951). Components in regression. *Biometrics* **7**:33–69.

US Food and Drug Administration (1992). Guidance on statistical procedures for bioequivalence using a standard two-treatment crossover design. Center for Drug Evaluation and Research, Rockville, MD.

US Food and Drug Administration (2000). Bioavailability and bioequivalence studies for orally administered drug products–general considerations. Center for Drug Evaluation and Research, Rockville, MD.

US Food and Drug Administration (2001). In Vivo bioequivalence studies based on population and individual bioequivalence approaches. Center for Drug Evaluation and Research, Rockville, MD.

Wang W (1999). On testing of individual bioequivalence. *Journal of the American Statistical Association*, **94**:880–887.

Weerahandi S (1991). Testing Variance Components in Mixed Models with Generalized p-Values. *Journal of the American Statistical Association.* **86**:151–153.

Weerahandi S (1993). Generalized confidence intervals. *Journal of the American Statistical Association.* **88**:899–905.

Weerahandi S (1995). *Exact Statistical Methods in Data Analysis*, Springer-Verlag, New York.

Williams JS (1962). A confidence interval for variance components. *Biometrika* **49**:278–281.

Wolfinger RD, Kass RE (2000). Non-conjugate Bayesian analysis of variance component models. *Biometrics*, **56**:768–774.

Zhou L, Mathew T (1994). Some tests for variance components using generalized p values. *Technometrics*, **36**:394–402.

Zhou L, Mathew T (1996). Response to RK Burdirk. *Technometrics*, **38**:93.

Appendix B

SAS CODE FOR EXAMPLES AND SIMULATIONS

```
/*****************************************
VER_MI3.SAS
This program does HSCI, MLS, and GPV procedures for the IBE/2S4P
of verapamil (Esinhart and Chichilli, 1994).
This is Example 1 in Chapter 3, Section 4.1.
*****************************************/
options ps=64 ls=120 nodate nonumber;
title 'Verapamil IBE';
libname mi3 'j:\mi3';
run;
*input data;
data hw5;
do sub = 10 to 32;
input seq @@;
do period = 1 to 4;
input trt $ auc @@;
lnauc = log(auc);
output;
end;
end;
datalines;
1 T 165.45 R 270.44 T 318.43 R 185.37
2 R 299.59 T 318.44 R 594.41 T 233.51
3 T 319.01 R 210.33 R 332.87 T 212.61
4 R 418.94 T 548.33 T 502.86 R 655.63
1 T 273.24 R 629.80 T 393.62 R 482.12
2 R 279.47 T 154.97 R 263.47 T 176.62
3 T 237.68 R 323.68 R 301.66 T 303.26
4 R 318.14 T 600.59 T 522.57 R 314.21
1 T 197.89 R 251.37 T 140.54 R 115.24
2 R 390.48 T 292.85 R 275.97 T 144.20
3 T 612.90 R 448.88 R 883.06 T 514.90
4 R 156.62 T 240.80 T 343.65 R 191.39
1 T 217.36 R 185.30 T 185.01 R 272.20
2 R 407.60 T 469.05 R 482.65 T 297.56
3 T 100.18 R 113.60 R 120.23 T 137.65
4 R  66.86 T  54.19 T 151.00 R  82.30
1 T 665.87 R 442.46 T 333.79 R 277.27
2 R 351.88 T 281.96 R 293.22 T 178.56
3 T 277.50 R 167.04 R 235.89 T 191.85
4 R 240.68 T 511.24 T 447.99 R 303.57
1 T 611.83 R 711.83 T 255.66 R 396.34
2 R 125.98 T 203.73 R 294.89 T 284.17
3 T 849.39 R 763.20 R 517.79 T 292.56
;;;;
run;
```

```
*compute est. of model parameters;
proc mixed data=hw5 method=reml maxiter=500 noitprint noclprint;
class seq sub period trt;
model lnauc=seq period trt seq*period(trt)/ ddfm=satterth;
random trt/subject=sub(seq) type=un g;
repeated/group=trt;
lsmeans trt;
estimate 'F' trt -1 1 / cl alpha=0.1;
ods output g=gm;
ods output covparms=cv;
ods output estimates=df;
run;
quit;

data cv; set cv(where=(covparm='Residual')) end=eof;
if group='trt R' then swr=estimate;
if group='trt T' then swt=estimate;
retain swr;
keep swr swt;
if eof then output;
run;

data gm; set gm end=eof;
if row=1 then s1=col1;
if eof then do; s2=col2; sd=s1+s2-2*col1; output; end;
retain s1;
keep sd s1 s2;
run;
*compute HSCI;
data ball; merge cv gm df(keep=estimate);
mi1=(estimate*estimate+sd+swt-swr);
mi3=mi1/max(swr,0.04);
e_i=sd+(swr+swt)/2;
e_d=estimate*estimate;
e_t=swt/2;
e_r=-3.9948*swr;
t=tinv(0.95,19);
h_d=(abs(estimate)+t*sqrt(0.7*e_i/16))**2;
qu=cinv(0.95,19);
ql=cinv(0.05,19);
h_i=19*e_i/ql;
h_t=19*e_t/ql;
h_r=19*e_r/qu;
u_d=(h_d-e_d)**2;
u_i=(h_i-e_i)**2;
u_t=(h_t-e_t)**2;
u_r=(h_r-e_r)**2;
```

```
h_mi3=(e_d+e_i+e_t+e_r)+sqrt(u_d+u_i+u_t+u_r);
run;

proc print data=ball;
run;
*compute MLS;
data ting; merge cv gm df(keep=estimate);
s_0=0.04;
eta0=((log(1.25))**2 + 0.05)/s_0;
sn=23;
nu=19;
d=estimate;
m_i=sd+(swr+swt)/2;
m_t=swt;
m_r=swr;
ssdrug=sn*d*d;
****This stuff is direct from Burdick simulation program****;
lambhat=ssdrug/(2*M_i);
nstar=int((1+2*lambhat)**2/(1+4*lambhat))+1;
fl5=cinv(0.05,nu)/nu;
fu2=cinv(0.95,nu)/nu;
fl6=finv(0.05,nstar,nu);
fl7=finv(0.05,nu,nu);
hd=nstar/cinv(0.05,nstar)-1;
hi=1/fl5-1;
ht=hi;
gr=1-1/fu2;
hdr=((1-fl6)**2-(hd*fl6)**2-gr**2)/fl6;
hir=((1-fl7)**2-(hi*fl7)**2-gr**2)/fl7;
htr=hir;
k1=ssdrug/(sn)+(1-1/(sn))*m_i+.5*m_t;
k2=m_r;
rhohtibe=k1/k2;
k6=(hd*ssdrug/(sn))**2+(hi*(1-1/(sn))*m_i)**2+(ht*.5*m_t)**2;
k7=hdr/(sn)*ssdrug*m_r+ hir*(1-1/(sn))*m_i*m_r+htr*.5*m_t*m_r;
k8=(gr*m_r)**2;
if (m_r gt s_0) then do;
vuiber=(2+k7/(k1*k2))**2-4*(1-k8/k2**2)*(1-k6/k1**2);
umlsi=rhohtibe*((2+k7/(k1*k2)+sqrt(vuiber))/(2* (1-k8/k2**2)))-1.5;
end;
else do;
vustar=k6+(1.5)**2*k8+1.5*k7;
umlsi=(k1-1.5*k2+sqrt(vustar))/s_0;
end;
run;

proc print data=ting;
```

```
run;
*compute GPV;
data call; merge cv gm df(keep=estimate);
mi1=(estimate*estimate+sd+swt-swr);
mi3=mi1/max(swr,0.04);
s_i=sd+(swr+swt)/2;
nu=19;
c=sqrt(0.7);
d=estimate;
ssu=nu*s_i;
ssv=nu*swt;
ssw=nu*swr;
eta0=2.4948;
s0=0.04;
do i=1 to 50000;
z=normal(0);
u=2*rangam(0,nu/2);
v=2*rangam(0,nu/2);
w=2*rangam(0,nu/2);
e_d=d-(c*z*sqrt(ssu/16))/sqrt(u);
e_i=ssu/u;
e_t=ssv/v;
e_r=ssw/w;
t=(((1.5*ssw)/w)+eta0*max(ssw/w,s0))/((e_d*e_d)+e_i+(e_t/2));
piv=((e_d*e_d)+e_i+(e_t/2)-((1.5*ssw)/w))/(max(ssw/w,s0));
psi=(t lt 1);
output;
end;

proc univariate;
var psi piv;
run;

/************************************************
MPH_IBE.SAS
This program does HSCI, MLS, GPV procedures,
and GCIs, for the IBE/2S4P
of methylphenidate (Meyer et al, 2000).
This is Example 2 in Chapter 3, Section 4.2.
It is also the example in Chapter 7, Section 3.3.
************************************************/
options ps=64 ls=120 nodate nonumber;
title 'Ritalin data/IBE';
libname mi3 'j:\mi3';
run;
*input data from Excel spreadsheet provided by Drs. Meyer and Straugh;
PROC IMPORT OUT= WORK.mph
```

```
            DATAFILE= "J:\mi3\MPH.xls"
            DBMS=EXCEL2000 REPLACE;
     GETNAMES=YES;
RUN;
*do log transform;
data mph; set mph;
lnauc=log(auc);  *change input variable to cmax to do other analysis;
run;
*compute est. of model parameters;
proc mixed data=mph method=reml maxiter=500 noitprint noclprint;
class sequence subject week treatment;
model lnauc=sequence week treatment sequence*week(treatment)/ ddfm=satterth;
random treatment/subject=subject(sequence) type=un g;
repeated/group=treatment;
lsmeans treatment;
estimate 'F' treatment -1 1 / cl alpha=0.1;
ods output g=gm;
ods output covparms=cv;
ods output estimates=df;
run;

data cv; set cv(where=(covparm='Residual')) end=eof;
if group='Treatment 1' then swr=estimate;
if group='Treatment 2' then swt=estimate;
retain swr;
keep swr swt;
if eof then output;
run;

data gm; set gm end=eof;
if row=1 then s1=col1;
if eof then do; s2=col2; sd=s1+s2-2*col1; output; end;
retain s1;
keep sd s1 s2;
run;
*compute HSCI;
data ball; merge cv gm df(keep=estimate);
mi1=(estimate*estimate+sd+swt-swr);
mi3=mi1/max(swr,0.04);
eta0=2.4948;
e_i=sd+(swr+swt)/2;
e_d=estimate*estimate;
e_t=swt/2;
e_r=(-1.5-eta0*(swr gt 0.04))*swr;
t=tinv(0.95,16);
h_d=(abs(estimate)+t*sqrt(0.8*e_i/16))**2;
qu=cinv(0.95,16);
```

```
ql=cinv(0.05,16);
h_i=16*e_i/ql;
h_t=16*e_t/ql;
h_r=16*e_r/qu;
u_d=(h_d-e_d)**2;
u_i=(h_i-e_i)**2;
u_t=(h_t-e_t)**2;
u_r=(h_r-e_r)**2;
h_mi3=(e_d+e_i+e_t+e_r-0.04*eta0*(swr le 0.04));
u_mi3=h_mi3+sqrt(u_d+u_i+u_t+u_r);
run;

proc print data=ball;
run;
*compute MLS;
data ting; merge cv gm df(keep=estimate);
s_0=0.04;
eta0=((log(1.25))**2 + 0.05)/s_0;
sn=20;
nu=16;
d=estimate;
m_i=sd+(swr+swt)/2;
m_t=swt;
m_r=swr;
ssdrug=sn*d*d;
****This stuff is direct from Burdick simulation program****;
lambhat=ssdrug/(2*M_i);
nstar=int((1+2*lambhat)**2/(1+4*lambhat))+1;
fl5=cinv(0.05,nu)/nu;
fu2=cinv(0.95,nu)/nu;
fl6=finv(0.05,nstar,nu);
fl7=finv(0.05,nu,nu);
hd=nstar/cinv(0.05,nstar)-1;
hi=1/fl5-1;
ht=hi;
gr=1-1/fu2;
hdr=((1-fl6)**2-(hd*fl6)**2-gr**2)/fl6;
hir=((1-fl7)**2-(hi*fl7)**2-gr**2)/fl7;
htr=hir;
k1=ssdrug/(sn)+(1-1/(sn))*m_i+.5*m_t;
k2=m_r;
rhohtibe=k1/k2;
k6=(hd*ssdrug/(sn))**2+(hi*(1-1/(sn))*m_i)**2+(ht*.5*m_t)**2;
k7=hdr/(sn)*ssdrug*m_r+ hir*(1-1/(sn))*m_i*m_r+htr*.5*m_t*m_r;
k8=(gr*m_r)**2;
if (m_r gt s_0) then do;
vuiber=(2+k7/(k1*k2))**2-4*(1-k8/k2**2)*(1-k6/k1**2);
```

```
umlsi=rhohtibe*((2+k7/(k1*k2)+sqrt(vuiber))/(2* (1-k8/k2**2)))-1.5;
end;
else do;
vustar=k6+(1.5)**2*k8+1.5*k7;
umlsi=(k1-1.5*k2+sqrt(vustar))/s_0;
end;
run;

proc print data=ting;
run;
*compute GPV;
data call; merge cv gm df(keep=estimate);
mi1=(estimate*estimate+sd+swt-swr);
mi3=mi1/max(swr,0.04);
s_i=sd+(swr+swt)/2;
nu=16;
c=sqrt(0.8);
d=estimate;
ssu=nu*s_i;
ssv=nu*swt;
ssw=nu*swr;
eta0=2.4948;
s0=0.04;
do i=1 to 50000;
z=normal(0);
u=2*rangam(0,nu/2);
v=2*rangam(0,nu/2);
w=2*rangam(0,nu/2);
e_i=ssu/u;
e_d=d-c*z*sqrt(e_i/16);
e_t=ssv/v;
e_r=ssw/w;
t=(1.5*e_r+eta0*max(e_r,s0))/((e_d*e_d)+e_i+(e_t/2));
piv=((e_d*e_d)+e_i+(e_t/2)-1.5*e_r)/(max(e_r,s0));
psi=(t lt 1);
phi=(piv gt eta0);
output;
end;

proc univariate;
var psi phi piv;
run;

/*****************************************
VER_MP3.SAS
This program does HSCI, MLS, and GPV procedures for the PBE/2S4P
of verapamil (Esinhart and Chichilli, 1994).
```

```
This is Example 1 in Chapter 3, Section 5.1.
*****************************************/
options ps=64 ls=120 nodate nonumber;
title 'Verapamil PBE';
libname st740 'j:\st740ph';
run;

*input, transform data and compute statistics;
data hw5;
input seq t1 $ c1 t2 $ c2 t3 $ c3 t4 $ c4;
la1 = log(c1);
la2 = log(c2);
la3 = log(c3);
la4 = log(c4);
if seq=1 then do;
   ut=(la1+la3)/2;
   ur=(la2+la4)/2;
   vt=(la1-la3)/sqrt(2);
   vr=(la2-la4)/sqrt(2);
   end;
if seq=2 then do;
   ur=(la1+la3)/2;
   ut=(la2+la4)/2;
   vr=(la1-la3)/sqrt(2);
   vt=(la2-la4)/sqrt(2);
   end;
if seq=3 then do;
   ut=(la1+la4)/2;
   ur=(la2+la3)/2;
   vt=(la1-la4)/sqrt(2);
   vr=(la2-la3)/sqrt(2);
   end;
if seq=4 then do;
   ut=(la2+la3)/2;
   ur=(la1+la4)/2;
   vt=(la2-la3)/sqrt(2);
   vr=(la1-la4)/sqrt(2);
   end;
   i=ut-ur;
datalines;
1 T 165.45 R 270.44 T 318.43 R 185.37
2 R 299.59 T 318.44 R 594.41 T 233.51
3 T 319.01 R 210.33 R 332.87 T 212.61
4 R 418.94 T 548.33 T 502.86 R 655.63
1 T 273.24 R 629.80 T 393.62 R 482.12
2 R 279.47 T 154.97 R 263.47 T 176.62
3 T 237.68 R 323.68 R 301.66 T 303.26
```

```
4 R 318.14 T 600.59 T 522.57 R 314.21
1 T 197.89 R 251.37 T 140.54 R 115.24
2 R 390.48 T 292.85 R 275.97 T 144.20
3 T 612.90 R 448.88 R 883.06 T 514.90
4 R 156.62 T 240.80 T 343.65 R 191.39
1 T 217.36 R 185.30 T 185.01 R 272.20
2 R 407.60 T 469.05 R 482.65 T 297.56
3 T 100.18 R 113.60 R 120.23 T 137.65
4 R  66.86 T  54.19 T 151.00 R  82.30
1 T 665.87 R 442.46 T 333.79 R 277.27
2 R 351.88 T 281.96 R 293.22 T 178.56
3 T 277.50 R 167.04 R 235.89 T 191.85
4 R 240.68 T 511.24 T 447.99 R 303.57
1 T 611.83 R 711.83 T 255.66 R 396.34
2 R 125.98 T 203.73 R 294.89 T 284.17
3 T 849.39 R 763.20 R 517.79 T 292.56
;;;;
run;
*compute sigma_2|1;
proc glm data=hw5 outstat=tx(where=(_type_='ERROR'));
class seq;
model ur=seq ut / solution;
run;
quit;

proc sort data=hw5;
by seq;
*compute statistics for each sequence;
proc univariate data=hw5 noprint;
var ut ur vt vr i;
output out=st mean=i1-i5 css=s1-s5;
by seq;
run;
*compute cross product for each seq.;
data xp; merge hw5(in=a) st;
by seq;
if a;
if first.seq then sxp=0;
xp=(ut-i1)*(ur-i2);
sxp=sxp+xp;
retain sxp;
if last.seq then output;
run;
*finish computing summary statistics;
proc univariate noprint;
var i5 s1-s5 sxp;
output out=st2 sum=i ss1 ss2 sswt sswr ssi ss12;
```

```
run;

data st2; set st2;
nu=19;
sig2i=ssi/nu;
d=i/4;
run;

*compute HSCI;
data ball; set st2;
s_0=0.04;
eta0=((log(1.25))**2 + 0.02)/s_0;
e_d=d*d;
e_1=ss1/nu;
e_2=sswt/(2*nu);
e_3=-(1+eta0*(sswr/nu gt s_0))*ss2/nu;
e_4=-(1+eta0*(sswr/nu gt s_0))*sswr/(2*nu);
t=tinv(0.95,nu);
h_d=(abs(d)+t*sqrt(0.7*sig2i/16))**2;
qu=cinv(0.95,nu);
ql=cinv(0.05,nu);
h_1=nu*e_1/ql;
h_2=nu*e_2/ql;
h_3=nu*e_3/qu;
h_4=nu*e_4/qu;
u_d=(h_d-e_d)**2;
u_1=(h_1-e_1)**2;
u_2=(h_2-e_2)**2;
u_3=(h_3-e_3)**2;
u_4=(h_4-e_4)**2;
h_mi3=(e_d+e_1+e_2+e_3+e_4-eta0*s_0*(sswr/nu le s_0))
+sqrt(u_d+u_1+u_2+u_3+u_4);
run;

proc print data=ball;
run;
*compute MLS;
data ting; set st2;
s_0=0.04;
eta0=((log(1.25))**2 + 0.02)/s_0;
stt=2*ss1/nu;
srr=2*ss2/nu;
st=sswt/nu;
sr=sswr/nu;
sn=23;
ssd=sn*d*d;
lambda=ssd/(2*sig2i);
```

```
nstar=ceil((1+2*lambda)*(1+2*lambda)/(1+4*lambda));
sdt=2*ssd/sn+stt+st;
ndt=ceil(sdt*sdt/(4*ssd*ssd/(sn*sn*nstar)+(stt*stt/nu)+(st*st/nu)));
if (srr+sr gt 2*s_0) then do;
src=srr+sr;
nrc=ceil(src*src/((srr*srr/nu)+(sr*sr/nu)));
f1=finv(0.05,ndt,nu);
f2=finv(0.95,nu,ndt);
f3=finv(0.05,nrc,nu);
f4=finv(0.95,nrc,ndt);
c0=1-2/nrc;
c2=(c0-f3)**2;
c1=((c0*(f2-1))**2)-c2*f2*f2;
c3=(f4-c0)**2;
c4=((c0*(1-f1))**2)-c3*f1*f1;
rhohat=(sdt-2*sig2i/sn)/src;
vu1=c1*sdt*sdt+4*c2*sig2i*sig2i/(sn*sn);
vu2=c3*sdt*sdt+4*c4*sig2i*sig2i/(sn*sn);
dt=sn*sdt/(2*sig2i);
if (dt le f1) then u_pber=c0*rhohat+sqrt(vu1)/src-1;
else if (dt gt f1) then u_pber=c0*rhohat+sqrt(vu2)/src-1;
end;
else do;
sir=srr+sr+2*sig2i/sn;
nir=ceil(sir*sir/((4*sig2i*sig2i/(sn*sn*nu))+(srr*srr/nu)+(sr*sr/nu)));
h1=(ndt/cinv(0.05,ndt))-1;
g2=1-nir/cinv(0.95,nir);
c5=((1-finv(0.05,ndt,nir))**2-finv(0.05,ndt,nir)
*finv(0.05,ndt,nir)*h1*h1-g2*g2)/finv(0.05,ndt,nir);
vm3=sdt*sdt*h1*h1+sir*sir*g2*g2+c5*sdt*sir;
u_pbec=(sdt-sir+sqrt(vm3))/(2*s_0);
end;
run;

proc print;
run;
*compute GPV;
data call; merge st2 tx;
k=sqrt(0.7/16);
ss2on1=ss;
b=ss12/ss1;
s_0=0.04;
eta0=((log(1.25))**2 + 0.02)/s_0;
do i=1 to 100000;
z=normal(0);
z12=normal(0);
u1=2*rangam(0,nu/2);
```

```
u2=2*rangam(0,(nu-1)/2);
ut=2*rangam(0,nu/2);
ur=2*rangam(0,nu/2);
sig2wt=sswt/ut;
sig2wr=sswr/ur;
sigsq1=ss1/u1;
lambdasq2=ss2on1/u2;
beta=b-z12*sqrt(lambdasq2/ss1);
sigi=sqrt(sigsq1*(beta-1)**2 + lambdasq2);
delta=d-k*sigi*z;
sigsq2=lambdasq2+beta*beta*sigsq1;
gpvmp3=(delta*delta+sigsq1+sig2wt/2-sigsq2-sig2wr/2)/max(sigsq2+sig2wr/2,s_0);
psi=(gpvmp3 gt eta0);
output;
end;
run;

proc univariate;
var psi gpvmp3 delta;
run;

/***********************************************
MPH_PBE.SAS
This program does HSCI, MLS, GPV procedures,
and GCIs, for the PBE/2S4P
of methylphenidate (Meyer et al, 2000).
This is Example 2 in Chapter 3, Section 5.2.
It is also the example in Chapter 7, Section 3.3.
***********************************************/
options ps=64 ls=120 nodate nonumber;
title 'Ritalin data/PBE';
run;
*input data from Excel spreadsheet provided by Drs. Meyer and Straugh;
PROC IMPORT OUT= WORK.mph
            DATAFILE= "j:\mi3\mph.xls"
            DBMS=EXCEL2000 REPLACE;
     GETNAMES=YES;
RUN;
*transpose data and do preliminary calculations;
proc transpose data=mph out=tm prefix=c;
var auc;  *change variable name to cmax to get other analysis;
id week;
by sequence subject;
run;
/*
proc print;
run;
```

```
*/
data hw5; set tm;
la1 = log(c1);
la2 = log(c2);
la3 = log(c3);
la4 = log(c4);
if sequence=3 then do;
   ut=(la1+la2)/2;
   ur=(la3+la4)/2;
   vt=(la1-la2)/sqrt(2);
   vr=(la3-la4)/sqrt(2);
   end;
if sequence=1 then do;
   ur=(la1+la2)/2;
   ut=(la3+la4)/2;
   vr=(la1-la2)/sqrt(2);
   vt=(la3-la4)/sqrt(2);
   end;
if sequence=4 then do;
   ut=(la1+la4)/2;
   ur=(la2+la3)/2;
   vt=(la1-la4)/sqrt(2);
   vr=(la2-la3)/sqrt(2);
   end;
if sequence=2 then do;
   ut=(la2+la3)/2;
   ur=(la1+la4)/2;
   vt=(la2-la3)/sqrt(2);
   vr=(la1-la4)/sqrt(2);
   end;
   i=ut-ur;
run;
*compute summary statistics;
proc glm data=hw5 outstat=tx(where=(_type_='ERROR'));
class sequence;
model ur=sequence ut / solution;
run;
quit;

proc sort data=hw5;
by sequence;

proc univariate data=hw5 noprint;
var ut ur vt vr i;
output out=st mean=i1-i5 css=s1-s5;
by sequence;
run;
```

```
data xp; merge hw5(in=a) st;
by sequence;
if a;
if first.sequence then sxp=0;
xp=(ut-i1)*(ur-i2);
sxp=sxp+xp;
retain sxp;
if last.sequence then output;
run;

proc univariate noprint;
var i5 s1-s5 sxp;
output out=st2 sum=i ss1 ss2 sswt sswr ssi ss12;
run;

data st2; set st2;
nu=16;
mut=ss1/nu;
mur=ss2/nu;
mwt=sswt/nu;
mwr=sswr/nu;
sig2i=ssi/nu;
d=i/4;
run;

*proc print;
run;
*compute HSCI;
data ball; set st2;
s_0=0.04;
eta0=((log(1.25))**2 + 0.02)/s_0;
e_d=d*d;
e_1=mut;
e_2=mwt/2;
e_3=-(1+eta0)*mur;
e_4=-(1+eta0)*mwr/2;
t=tinv(0.95,nu);
h_d=(abs(d)+t*sqrt(0.8*sig2i)/4)**2;
qu=cinv(0.95,nu);
ql=cinv(0.05,nu);
h_1=nu*e_1/ql;
h_2=nu*e_2/ql;
h_3=nu*e_3/qu;
h_4=nu*e_4/qu;
u_d=(h_d-e_d)**2;
u_1=(h_1-e_1)**2;
```

```
u_2=(h_2-e_2)**2;
u_3=(h_3-e_3)**2;
u_4=(h_4-e_4)**2;
mp3=(e_d+e_1+e_2-mur-mwr/2)/(mur+mwr/2);
lmp3=e_d+e_1+e_2+e_3+e_4;
h_mp3=lmp3+sqrt(u_d+u_1+u_2+u_3+u_4);
run;

proc print data=ball;
run;
*compute MLS;
data ting; set st2;
s_0=0.04;
eta0=((log(1.25))**2 + 0.02)/s_0;
stt=2*ss1/nu;
srr=2*ss2/nu;
st=sswt/nu;
sr=sswr/nu;
sn=20;
ssd=sn*d*d;
lambda=ssd/(2*sig2i);
nstar=ceil((1+2*lambda)*(1+2*lambda)/(1+4*lambda));
sdt=2*ssd/sn+stt+st;
ndt=ceil(sdt*sdt/(4*ssd*ssd/(sn*sn*nstar)+(stt*stt/nu)+(st*st/nu)));
if (srr+sr gt 2*s_0) then do;
src=srr+sr;
nrc=ceil(src*src/((srr*srr/nu)+(sr*sr/nu)));
f1=finv(0.05,ndt,nu);
f2=finv(0.95,nu,ndt);
f3=finv(0.05,nrc,nu);
f4=finv(0.95,nrc,ndt);
c0=1-2/nrc;
c2=(c0-f3)**2;
c1=((c0*(f2-1))**2)-c2*f2*f2;
c3=(f4-c0)**2;
c4=((c0*(1-f1))**2)-c3*f1*f1;
rhohat=(sdt-2*sig2i/sn)/src;
vu1=c1*sdt*sdt+4*c2*sig2i*sig2i/(sn*sn);
vu2=c3*sdt*sdt+4*c4*sig2i*sig2i/(sn*sn);
dt=sn*sdt/(2*sig2i);
if (dt le f1) then u_pber=c0*rhohat+sqrt(vu1)/src-1;
else if (dt gt f1) then u_pber=c0*rhohat+sqrt(vu2)/src-1;
end;
else do;
sir=srr+sr+2*sig2i/sn;
nir=ceil(sir*sir/((4*sig2i*sig2i/(sn*sn*nu))+(srr*srr/nu)+(sr*sr/nu)));
h1=(ndt/cinv(0.05,ndt))-1;
```

```
g2=1-nir/cinv(0.95,nir);
c5=((1-finv(0.05,ndt,nir))**2-finv(0.05,ndt,nir)
*finv(0.05,ndt,nir)*h1*h1-g2*g2)/finv(0.05,ndt,nir);
vm3=sdt*sdt*h1*h1+sir*sir*g2*g2+c5*sdt*sir;
u_pbec=(sdt-sir+sqrt(vm3))/(2*s_0);
end;
run;

proc print;
run;

proc print;
run;
*compute GPV;
data call; merge st2 tx;
nu=16;
k=sqrt(0.8/16);
ss2on1=ss;
b=ss12/ss1;
s_0=0.04;
eta0=((log(1.25))**2 + 0.02)/s_0;
do j=1 to 50000;
z=normal(0);
z12=normal(0);
u1=2*rangam(0,nu/2);
u2=2*rangam(0,(nu-1)/2);
ut=2*rangam(0,nu/2);
ur=2*rangam(0,nu/2);
sig2wt=sswt/ut;
sig2wr=sswr/ur;
sigsq1=ss1/u1;
lambdasq2=ss2on1/u2;
beta=b-z12*sqrt(lambdasq2/ss1);
sigi=sqrt(sigsq1*(beta-1)**2 + lambdasq2);
delta=d-k*sigi*z;
sigsq2=lambdasq2+beta*beta*sigsq1;
gpvmp3=(delta*delta+sigsq1+sig2wt/2-sigsq2-sig2wr/2)/max(sigsq2+sig2wr/2,s_0);
psi=(gpvmp3 gt eta0);
output;
end;
run;

proc univariate;
var psi gpvmp3 delta;
run;

/*****************************************
```

```
SIMIBE.SAS
This program does the test size simulations
for IBE/2S4P.
Chapter 3, Section 6.1, Table 2.
*******************************************/
options ps=66 ls=120 nodate nonumber mprint;
libname mi3 'c:\My Documents\My SAS Files\V8';
*ibname mi3 'j:\mi3';
title 'Simulations for IBE';
run;

%macro simibe(del,sd,swt,swr,nsim,ngp,seed,ds);

title2 "delta=&del, sigma_d=&sd sigma_wt=&swt sigma_wr=&swr";
*Create simulation data sets;
data sim;
****Input model parameter values****;
sigd=&sd;
sigwt=&swt;
sigwr=&swr;
delta=&del;
****Calculate needed parameter values****;
sig2wt=sigwt*sigwt;
sig2wr=sigwr*sigwr;
sig2i=sigd*sigd+sig2wt/2+sig2wr/2;
sigi=sqrt(sig2i);
mi3=(delta*delta+sig2i+0.5*sig2wt-1.5*sig2wr)/max(sig2wr,0.04);
call symput('MI3',mi3);
****Compute simulated values****;
junk=int(100000*ranuni(&seed));
do n=6,12,18,24;
nu=2*(n-1);
k=1/sqrt(2*n);
do i=1 to &nsim;
z=normal(0);
u=2*rangam(0,nu/2);
v=2*rangam(0,nu/2);
w=2*rangam(0,nu/2);
d=delta+k*sigi*z;
sswt=sig2wt*u;
sswr=sig2wr*v;
ssi=sig2i*w;
keep n i nu d ssi sswt sswr k mi3;
output;
end;
end;
run;
```

```
****Compute HSCI****;
data ball; set sim;
s_0=0.04;
eta0=((log(1.25))**2 + 0.05)/s_0;
s2wr=sswr/nu;
e_d=d*d;
e_i=ssi/nu;
e_t=sswt/(2*nu);
e_r=(-3.9948*(s2wr gt s_0)-1.5*(s2wr le s_0))*s2wr;
t=tinv(0.95,nu);
h_d=(abs(d)+t*k*sqrt(e_i))**2;
qu=cinv(0.95,nu);
ql=cinv(0.05,nu);
h_i=ssi/ql;
h_t=nu*e_t/ql;
h_r=nu*e_r/qu;
u_d=(h_d-e_d)**2;
u_i=(h_i-e_i)**2;
u_t=(h_t-e_t)**2;
u_r=(h_r-e_r)**2;
h_mi3=(e_d+e_i+e_t+e_r-eta0*s_0*(s2wr le s_0))+sqrt(u_d+u_i+u_t+u_r);
phi=(h_mi3 lt 0);
keep n i phi h_mi3;
run;

proc univariate data=ball noprint;
var phi;
by n;
output out=h3 mean=hsci;
run;

****Compute Ting-Burdick****;
data ting; set sim;
s_0=0.04;
eta0=((log(1.25))**2 + 0.05)/s_0;
sn=2*n;
m_i=ssi/nu;
m_t=sswt/nu;
m_r=sswr/nu;
ssdrug=sn*d*d;
****This stuff is direct from Burdick simulation program****;
lambhat=ssdrug/(2*M_i);
nstar=int((1+2*lambhat)**2/(1+4*lambhat))+1;
fl5=cinv(0.05,nu)/nu;
fu2=cinv(0.95,nu)/nu;
fl6=finv(0.05,nstar,nu);
```

```
fl7=finv(0.05,nu,nu);
hd=nstar/cinv(0.05,nstar)-1;
hi=1/fl5-1;
ht=hi;
gr=1-1/fu2;
hdr=((1-fl6)**2-(hd*fl6)**2-gr**2)/fl6;
hir=((1-fl7)**2-(hi*fl7)**2-gr**2)/fl7;
htr=hir;
k1=ssdrug/(sn)+(1-1/(sn))*m_i+.5*m_t;
k2=m_r;
rhohtibe=k1/k2;
k6=(hd*ssdrug/(sn))**2+(hi*(1-1/(sn))*m_i)**2+(ht*.5*m_t)**2;
k7=hdr/(sn)*ssdrug*m_r+ hir*(1-1/(sn))*m_i*m_r+htr*.5*m_t*m_r;
k8=(gr*m_r)**2;
if (m_r gt s_0) then do;
vuiber=(2+k7/(k1*k2))**2-4*(1-k8/k2**2)*(1-k6/k1**2);
umlsi=rhohtibe*((2+k7/(k1*k2)+sqrt(vuiber))/(2* (1-k8/k2**2)))-1.5;
end;
else do;
vustar=k6+(1.5)**2*k8+1.5*k7;
umlsi=(k1-1.5*k2+sqrt(vustar))/s_0;
end;
chi=(umlsi lt eta0);
run;

proc univariate data=ting noprint;
var chi;
output out=mls mean=mls;
by n;
run;

****Compute GPV****;
data call; set sim;
s_0=0.04;
eta0=((log(1.25))**2 + 0.05)/s_0;
xi=0;
xi2=0;
do j=1 to &ngp;
z=normal(0);
u=2*rangam(0,nu/2);
v=2*rangam(0,nu/2);
w=2*rangam(0,nu/2);
sig2i=ssi/w;
delta=d-k*z*sqrt(sig2i);
sig2wt=sswt/u;
sig2wr=sswr/v;
gpvmi3=(delta*delta+sig2i+0.5*sig2wt-1.5*sig2wr)/max(sig2wr,s_0);
```

```
xi=xi+(gpvmi3 gt eta0);
xi2=xi2+(gpvmi3 gt mi3);
end;
p=xi/&ngp;
c=xi2/&ngp;
psi05=(p lt 0.05);
psi025=(p lt 0.025);
cover=(c gt 0.05);
run;

proc univariate data=call noprint;
var psi05 psi025 cover;
output out=gpv mean=gpv5 gpv25 cover95;
by n;
run;
*merge results;
data tests; merge h3 mls gpv;
length test $ 5;
by n;
delta=&del;
sigma_d=&sd;
sigma_wt=&swt;
sigma_wr=&swr;
test="&ds";
mi3=&mi3;
run;

proc print data=tests;
run;

proc append base=mi3.testsi data=tests;
run;

%mend;

%simibe(0.315742,0.01,0.15,0.15,5000,10000,84431,i1a);
%simibe(0.299655,0.1,0.15,0.15,5000,10000,79087,i1b);
%simibe(0.278016,0.15,0.15,0.15,5000,10000,71089,i1c);
%simibe(0.315742,0.01,0.20,0.20,5000,10000,38693,i2a);
%simibe(0.299655,0.1,0.20,0.20,5000,10000,90617,i2b);
%simibe(0.278016,0.15,0.20,0.20,5000,10000,26627,i2c);
%simibe(0.363148,0.01,0.23,0.23,5000,10000,19157,i3a);
%simibe(0.349251,0.1,0.23,0.23,5000,10000,35393,i3b);
%simibe(0.330872,0.15,0.23,0.23,5000,10000,53507,i3c);
%simibe(0.4737450264747,0.01,0.3,0.3,5000,10000,56401,i4a);
%simibe(0.4631785294133,0.1,0.3,0.3,5000,10000,97931,i4b);
%simibe(0.4494823134557,0.15,0.3,0.3,5000,10000,90053,i4c);
```

```
%simibe(0.7896876142387,0.01,0.5,0.5,5000,10000,88211,i5a);
%simibe(0.7833942354153,0.1,0.5,0.5,5000,10000,31123,i5b);
%simibe(0.775375088639,0.15,0.5,0.5,5000,10000,36931,i5c);

proc print data=mi3.testsi;
run;

/*****************************************
SIMIBEP.SAS
This program does the power/sample size simulations
for IBE/2S4P.
Chapter 3, Section 6.1, Table 3.
*****************************************/
options ps=64 ls=120 nodate nonumber mprint;
title 'Simulation data';
libname mi3 'c:\My Documents\My SAS Files\V8';
*ibname mi3 'j:\mi3';
title 'Simulation for PBE';
run;

%macro simibe(del,sd,swt,swr,nsim,ngp,n1,n2,seed,ds);

title2 "delta=&del, sigma_d=&sd sigma_wt=&swt sigma_wr=&swr";
*create simulation data sets;
data sim;
****Input model parameter values****;
sigd=&sd;
sigwt=&swt;
sigwr=&swr;
delta=&del;
****Calculate needed parameter values****;
sig2wt=sigwt*sigwt;
sig2wr=sigwr*sigwr;
sig2i=sigd*sigd+sig2wt/2+sig2wr/2;
sigi=sqrt(sig2i);
mi3=(delta*delta+sig2i+0.5*sig2wt-1.5*sig2wr)/max(sig2wr,0.04);
call symput('MI3',mi3);
****Compute simulated values****;
junk=int(100000*ranuni(&seed));
do n=&n1 to &n2;
*do n=5 to 35 by 3;
nu=2*(n-1);
k=1/sqrt(2*n);
do i=1 to &nsim;
z=normal(0);
u=2*rangam(0,nu/2);
v=2*rangam(0,nu/2);
```

```
w=2*rangam(0,nu/2);
d=delta+k*sigi*z;
sswt=sig2wt*u;
sswr=sig2wr*v;
ssi=sig2i*w;
keep n i nu d ssi sswt sswr k mi3;
output;
end;
end;
run;

****Compute HSCI****;
data ball; set sim;
s_0=0.04;
eta0=((log(1.25))**2 + 0.05)/s_0;
s2wr=sswr/nu;
e_d=d*d;
e_i=ssi/nu;
e_t=sswt/(2*nu);
e_r=(-3.9948*(s2wr gt s_0)-1.5*(s2wr le s_0))*s2wr;
t=tinv(0.95,nu);
h_d=(abs(d)+t*k*sqrt(e_i))**2;
qu=cinv(0.95,nu);
ql=cinv(0.05,nu);
h_i=ssi/ql;
h_t=nu*e_t/ql;
h_r=nu*e_r/qu;
u_d=(h_d-e_d)**2;
u_i=(h_i-e_i)**2;
u_t=(h_t-e_t)**2;
u_r=(h_r-e_r)**2;
h_mi3=(e_d+e_i+e_t+e_r-eta0*s_0*(s2wr le s_0))+sqrt(u_d+u_i+u_t+u_r);
phi=(h_mi3 lt 0);
keep n i phi h_mi3;
run;

proc univariate data=ball noprint;
var phi;
by n;
output out=h3 mean=hsci;
run;

****Compute Ting-Burdick****;
data ting; set sim;
s_0=0.04;
eta0=((log(1.25))**2 + 0.05)/s_0;
sn=2*n;
```

```
m_i=ssi/nu;
m_t=sswt/nu;
m_r=sswr/nu;
ssdrug=sn*d*d;
****This stuff is direct from Burdick simulation program****;
lambhat=ssdrug/(2*M_i);
nstar=int((1+2*lambhat)**2/(1+4*lambhat))+1;
fl5=cinv(0.05,nu)/nu;
fu2=cinv(0.95,nu)/nu;
fl6=finv(0.05,nstar,nu);
fl7=finv(0.05,nu,nu);
hd=nstar/cinv(0.05,nstar)-1;
hi=1/fl5-1;
ht=hi;
gr=1-1/fu2;
hdr=((1-fl6)**2-(hd*fl6)**2-gr**2)/fl6;
hir=((1-fl7)**2-(hi*fl7)**2-gr**2)/fl7;
htr=hir;
k1=ssdrug/(sn)+(1-1/(sn))*m_i+.5*m_t;
k2=m_r;
rhohtibe=k1/k2;
k6=(hd*ssdrug/(sn))**2+(hi*(1-1/(sn))*m_i)**2+(ht*.5*m_t)**2;
k7=hdr/(sn)*ssdrug*m_r+ hir*(1-1/(sn))*m_i*m_r+htr*.5*m_t*m_r;
k8=(gr*m_r)**2;
if (m_r gt s_0) then do;
vuiber=(2+k7/(k1*k2))**2-4*(1-k8/k2**2)*(1-k6/k1**2);
umlsi=rhohtibe*((2+k7/(k1*k2)+sqrt(vuiber))/(2* (1-k8/k2**2)))-1.5;
end;
else do;
vustar=k6+(1.5)**2*k8+1.5*k7;
umlsi=(k1-1.5*k2+sqrt(vustar))/s_0;
end;
chi=(umlsi lt eta0);
run;

proc univariate data=ting noprint;
var chi;
output out=mls mean=mls;
by n;
run;

****Compute GPV****;
data call; set sim;
s_0=0.04;
eta0=((log(1.25))**2 + 0.05)/s_0;
xi=0;
xi2=0;
```

```
do j=1 to &ngp;
z=normal(0);
u=2*rangam(0,nu/2);
v=2*rangam(0,nu/2);
w=2*rangam(0,nu/2);
sig2i=ssi/w;
delta=d-k*z*sqrt(sig2i);
sig2wt=sswt/u;
sig2wr=sswr/v;
gpvmi3=(delta*delta+sig2i+0.5*sig2wt-1.5*sig2wr)/max(sig2wr,s_0);
xi=xi+(gpvmi3 gt eta0);
xi2=xi2+(gpvmi3 gt mi3);
end;
p=xi/&ngp;
c=xi2/&ngp;
psi05=(p lt 0.05);
cover=(c gt 0.05);
run;

proc univariate data=call noprint;
var psi05 cover;
output out=gpv mean=gpv5 cover95;
by n;
run;
*merge results;
data tests; merge h3 mls gpv;
length test $ 5;
by n;
delta=&del;
sigma_d=&sd;
sigma_wt=&swt;
sigma_wr=&swr;
test="&ds";
mi3=&mi3;
run;

proc print data=tests;
run;

proc append base=mi3.poweri data=tests;
run;

%mend;

%simibe(0.05,0.01,0.15,0.15,5000,10000,5,6,65479,p1a);
%simibe(0.05,0.1,0.15,0.15,5000,10000,6,9,47041,p1b);
%simibe(0.05,0.15,0.15,0.15,5000,10000,10,13,45823,p1c);
```

```
%simibe(0.05,0.01,0.20,0.20,5000,10000,8,10,54001,p2a);
%simibe(0.05,0.1,0.20,0.20,5000,10000,11,14,28031,p2b);
%simibe(0.05,0.15,0.20,0.20,5000,10000,17,21,18503,p2c);
%simibe(0.05,0.01,0.23,0.23,5000,10000,11,14,54577,p3a);
%simibe(0.05,0.1,0.23,0.23,5000,10000,14,17,33893,p3b);
%simibe(0.05,0.15,0.23,0.23,5000,10000,20,24,79087,p3c);
%simibe(0.05,0.01,0.3,0.3,5000,10000,13,15,47939,p4a);
%simibe(0.05,0.1,0.3,0.3,5000,10000,14,16,41243,p4b);
%simibe(0.05,0.15,0.3,0.3,5000,10000,18,20,86453,p4c);
%simibe(0.05,0.01,0.5,0.5,5000,10000,12,15,26701,p5a);
%simibe(0.05,0.1,0.5,0.5,5000,10000,13,15,23227,p5b);
%simibe(0.05,0.15,0.5,0.5,5000,10000,14,16,53551,p5c);

proc print data=mi3.poweri;
run;

/*******************************************
SIMPBE.SAS
This program does the test size simulations
for PBE/2S4P.
Chapter 3, Section 6.2, Table 4.
*******************************************/
options ps=64 ls=120 nodate nonumber;*print;
title 'Simulation data';
libname mi3 'c:\My Documents\My SAS Files\V8';
*ibname mi3 'j:\mi3';
title 'Simulation for PBE';
run;

%macro simpbe(del,r,sbt,sbr,swt,swr,nsim,ngp,seed,ds);

title2 "delta=&del, rho=&r, sigma_bt=&sbt
sigma_br=&sbr sigma_wt=&swt sigma_wr=&swr";
run;
*create simulation data sets;
data sim;
****Input model parameter values****;
sigbt=&sbt;
sigbr=&sbr;
sigwt=&swt;
sigwr=&swr;
delta=&del;
rho=&r;
****Calculate needed parameter values****;
sig2bt=sigbt*sigbt;
sig2br=sigbr*sigbr;
sig2wt=sigwt*sigwt;
```

```
sig2wr=sigwr*sigwr;
sigsq1=sig2bt+sig2wt/2;
sigsq2=sig2br+sig2wr/2;
sig12=rho*sigbt*sigbr;
beta=sig12/sigsq1;
sigi=sqrt(sigsq1+sigsq2-2*sig12);
lambda1=sqrt(sigsq1);
lambdasq2=sigsq2-sig12*sig12/sigsq1;
lambda2=sqrt(lambdasq2);
mp3=(delta*delta+sig2bt+sig2wt-sig2br-sig2wr)/max(sig2br+sig2wr,0.04);
call symput('MP3',mp3);
****Compute simulated values****;
junk=int(100000*ranuni(&seed));
do n=12,18,24;
nu=2*(n-1);
k=1/sqrt(2*n);
do i=1 to &nsim;
z=normal(0);
u1=2*rangam(0,nu/2);
u2=2*rangam(0,(nu-1)/2);
ut=2*rangam(0,nu/2);
ur=2*rangam(0,nu/2);
z12=normal(0);
d=delta+k*sigi*z;
sswt=sig2wt*ut;
sswr=sig2wr*ur;
ss1=sigsq1*u1;
ss2on1=lambdasq2*u2;
b=beta+z12*lambda2/sqrt(ss1);
ss2=b*b*ss1+ss2on1;
sig2i=(ss2+ss1*(1-2*b))/nu;
keep n i nu d ss1 ss2 ss2on1 sig2i sswt sswr b k mp3;
output;
end;
end;
run;

****Compute HSCI****;
data ball; set sim;
s_0=0.04;
eta0=((log(1.25))**2 + 0.02)/s_0;
e_d=d*d;
e_1=ss1/nu;
e_2=sswt/(2*nu);
t=tinv(0.95,nu);
h_d=(abs(d)+k*t*sqrt(sig2i))**2;
qu=cinv(0.95,nu);
```

```
ql=cinv(0.05,nu);
h_1=nu*e_1/ql;
h_2=nu*e_2/ql;
u_d=(h_d-e_d)**2;
u_1=(h_1-e_1)**2;
u_2=(h_2-e_2)**2;
if ((ss2/nu+sswr/(2*nu)) gt s_0) then do;
e_3=-(1+eta0)*ss2/nu;
e_4=-(1+eta0)*sswr/(2*nu);
h_3=nu*e_3/qu;
h_4=nu*e_4/qu;
u_3=(h_3-e_3)**2;
u_4=(h_4-e_4)**2;
lmp3=e_d+e_1+e_2+e_3+e_4;
h_mp3=lmp3+sqrt(u_d+u_1+u_2+u_3+u_4);
end;
else do;
e_3=-1*ss2/nu;
e_4=-1*sswr/(2*nu);
h_3=nu*e_3/qu;
h_4=nu*e_4/qu;
u_3=(h_3-e_3)**2;
u_4=(h_4-e_4)**2;
lmp3=e_d+e_1+e_2+e_3+e_4-s_0*eta0;
h_mp3=lmp3+sqrt(u_d+u_1+u_2+u_3+u_4);
end;
phi=(h_mp3 lt 0);
keep n i lmp3 h_mp3 phi;
run;

proc univariate data=ball noprint;
var phi;
by n;
output out=h3 mean=hsci;
run;

****Compute Ting-Burdick****;
data ting; set sim;
s_0=0.04;
eta0=((log(1.25))**2 + 0.02)/s_0;
stt=2*ss1/nu;
srr=2*ss2/nu;
st=sswt/nu;
sr=sswr/nu;
sn=2*n;
ssd=sn*d*d;
lambda=ssd/(2*sig2i);
```

```
nstar=ceil((1+2*lambda)*(1+2*lambda)/(1+4*lambda));
sdt=2*ssd/sn+stt+st;
ndt=ceil(sdt*sdt/(4*ssd*ssd/(sn*sn*nstar)+(stt*stt/nu)+(st*st/nu)));
if (srr+sr gt 2*s_0) then do;
src=srr+sr;
nrc=ceil(src*src/((srr*srr/nu)+(sr*sr/nu)));
f1=finv(0.05,ndt,nu);
f2=finv(0.95,nu,ndt);
f3=finv(0.05,nrc,nu);
f4=finv(0.95,nrc,ndt);
c0=1-2/nrc;
c2=(c0-f3)**2;
c1=((c0*(f2-1))**2)-c2*f2*f2;
c3=(f4-c0)**2;
c4=((c0*(1-f1))**2)-c3*f1*f1;
rhohat=(sdt-2*sig2i/sn)/src;
vu1=c1*sdt*sdt+4*c2*sig2i*sig2i/(sn*sn);
vu2=c3*sdt*sdt+4*c4*sig2i*sig2i/(sn*sn);
dt=sn*sdt/(2*sig2i);
if (dt le f1) then u_pber=c0*rhohat+sqrt(vu1)/src-1;
else if (dt gt f1) then u_pber=c0*rhohat+sqrt(vu2)/src-1;
chi=(u_pber lt eta0);
end;
else do;
sir=srr+sr+2*sig2i/sn;
nir=ceil(sir*sir/((4*sig2i*sig2i/(sn*sn*nu))+(srr*srr/nu)+(sr*sr/nu)));
h1=(ndt/cinv(0.05,ndt))-1;
g2=1-nir/cinv(0.95,nir);
c5=((1-finv(0.05,ndt,nir))**2-finv(0.05,ndt,nir)
*finv(0.05,ndt,nir)*h1*h1-g2*g2)/finv(0.05,ndt,nir);
vm3=sdt*sdt*h1*h1+sir*sir*g2*g2+c5*sdt*sir;
u_pbec=(sdt-sir+sqrt(vm3))/(2*s_0);
chi=(u_pbec lt eta0);
end;
run;

proc univariate data=ting noprint;
var chi;
output out=mls mean=mls;
by n;
run;

****Compute GPV****;
data call; set sim;
s_0=0.04;
eta0=((log(1.25))**2 + 0.02)/s_0;
xi=0;
```

```
xi2=0;
do j=1 to &ngp;
z=normal(0);
z12=normal(0);
u1=2*rangam(0,nu/2);
u2=2*rangam(0,(nu-1)/2);
ut=2*rangam(0,nu/2);
ur=2*rangam(0,nu/2);
sig2wt=sswt/ut;
sig2wr=sswr/ur;
sigsq1=ss1/u1;
lambdasq2=ss2on1/u2;
beta=b-z12*sqrt(lambdasq2/ss1);
sigi=sqrt(sigsq1*(beta-1)**2 + lambdasq2);
delta=d-k*sigi*z;
sigsq2=lambdasq2+beta*beta*sigsq1;
gpvmp3=(delta*delta+sigsq1+sig2wt/2-sigsq2-sig2wr/2)/max(sigsq2+sig2wr/2,s_0);
xi=xi+(gpvmp3 gt eta0);
xi2=xi2+(gpvmp3 gt mp3);
end;
p=xi/&ngp;
c=xi2/&ngp;
psi05=(p lt 0.05);
psi025=(p lt 0.025);
cover=(c gt 0.05);
run;

proc univariate data=call noprint;
var psi05 psi025 cover;
output out=gpv mean=gpv5 gpv25 cover95;
by n;
run;
*merge results;
data tests; merge h3 mls gpv;
length test $ 5;
by n;
delta=&del;
r=&r;
sigma_bt=&sbt;
sigma_br=&sbr;
sigma_wt=&swt;
sigma_wr=&swr;
test="&ds";
mp3=&mp3;
run;run;

proc print data=tests;
```

```
run;

proc append base=mi3.tests2f data=tests;
run;

%mend;


%simpbe(0.2641837,0.8,0.1,0.1,0.1,0.1,5000,10000,71363,d1a);
%simpbe(0.2641837,0.9,0.1,0.1,0.1,0.1,5000,10000,88241,d1b);
%simpbe(0.295366,0.8,0.2,0.2,0.1,0.1,5000,10000,54559,d1c);
%simpbe(0.295366,0.9,0.2,0.2,0.1,0.1,5000,10000,32917,d1d);
%simpbe(0.2641837,0.8,0.2,0.2,0.2,0.2,5000,10000,46867,d2a);
%simpbe(0.2641837,0.9,0.2,0.2,0.2,0.2,5000,10000,19237,d2b);
%simpbe(0.590733,0.8,0.4,0.4,0.2,0.2,5000,10000,31699,d2c);
%simpbe(0.590733,0.9,0.4,0.4,0.2,0.2,5000,10000,61211,d2d);
%simpbe(0.560418,0.8,0.3,0.3,0.3,0.3,5000,10000,85103,d3a);
%simpbe(0.560418,0.9,0.3,0.3,0.3,0.3,5000,10000,65267,d3b);
%simpbe(0.886099,0.8,0.6,0.6,0.3,0.3,5000,10000,55061,d3c);
%simpbe(0.886099,0.9,0.6,0.6,0.3,0.3,5000,10000,72277,d3d);
%simpbe(0.295366,0.8,0.2,0.173205,0.141412,0.1,5000,10000,20857,e1a);
%simpbe(0.295366,0.9,0.2,0.173205,0.141412,0.1,5000,10000,32143,e1b);
%simpbe(0.295366,0.8,0.173205,0.2,0.1,0.141421,5000,10000,14197,e1c);
%simpbe(0.295366,0.9,0.173205,0.2,0.1,0.141421,5000,10000,33343,e1d);
%simpbe(0.590733,0.8,0.387298,0.4,0.223607,0.2,5000,10000,98321,e2a);
%simpbe(0.590733,0.9,0.387298,0.4,0.223607,0.2,5000,10000,54517,e2b;
%simpbe(0.590733,0.8,0.4,0.387298,0.2,0.223607,5000,10000,69497,e2c);
%simpbe(0.590733,0.9,0.4,0.387298,0.2,0.223607,5000,10000,80917,e2d);
%simpbe(0.886099,0.8,0.6,0.5,0.3,0.4472136,5000,10000,55061,e3a);
%simpbe(0.886099,0.9,0.6,0.5,0.3,0.4472136,5000,10000,72277,e3a);
%simpbe(0.886099,0.8,0.5,0.6,0.4472136,0.3,5000,10000,55061,e3c);
%simpbe(0.886099,0.9,0.5,0.6,0.4472136,0.3,5000,10000,72277,e3d);
%simpbe(0,0.8,0.282476626,0.1,0.1,0.1,5000,10000,75577,f1a);
%simpbe(0,0.9,0.282476626,0.1,0.1,0.1,5000,10000,64621,f1b);
%simpbe(0,0.8,0.356708993,0.2,0.1,0.1,5000,10000,79043,f1c);
%simpbe(0,0.9,0.356708993,0.2,0.1,0.1,5000,10000,86923,f1d);
%simpbe(0,0.8,0.423775989,0.2,0.2,0.2,5000,10000,46867,f2a);
%simpbe(0,0.9,0.423775989,0.2,0.2,0.2,5000,10000,19237,f2b);
%simpbe(0,0.8,0.713417986,0.4,0.2,0.2,5000,10000,31699,f2c);
%simpbe(0,0.9,0.713417986,0.4,0.2,0.2,5000,10000,61211,f2d);
%simpbe(0,0.8,0.635663984,0.3,0.3,0.3,5000,10000,85103,f3a);
%simpbe(0,0.9,0.635663984,0.3,0.3,0.3,5000,10000,65267,f3b);
%simpbe(0,0.8,1.070126979,0.6,0.3,0.3,5000,10000,55061,f3c);
%simpbe(0,0.9,1.070126979,0.6,0.3,0.3,5000,10000,44897,f3d);

proc print;
run;
```

```
/*****************************************
SIMPBE_P.SAS
This program does the power/sample size simulations
for PBE/2S4P.
Chapter 3, Section 6.2, Table 5.
*****************************************/
options ps=66 ls=120 nodate nonumber;*print;
title 'Simulation data';
libname mi3 'c:\My Documents\My SAS Files\V8';
*ibname mi3 'j:\mi3';
title 'Simulation for PBE';
run;

%macro simpbe(del,r,sbt,sbr,swt,swr,nsim,ngp,seed,n1,n2,n3,n4,ds);

title2 "delta=&del, rho=&r, sigma_bt=&sbt
sigma_br=&sbr sigma_wt=&swt sigma_wr=&swr";
run;
*create simulation data sets;
data sim;
****Input model parameter values****;
sigbt=&sbt;
sigbr=&sbr;
sigwt=&swt;
sigwr=&swr;
delta=&del;
rho=&r;
****Calculate needed parameter values****;
sig2bt=sigbt*sigbt;
sig2br=sigbr*sigbr;
sig2wt=sigwt*sigwt;
sig2wr=sigwr*sigwr;
sigsq1=sig2bt+sig2wt/2;
sigsq2=sig2br+sig2wr/2;
sig12=rho*sigbt*sigbr;
beta=sig12/sigsq1;
sigi=sqrt(sigsq1+sigsq2-2*sig12);
lambda1=sqrt(sigsq1);
lambdasq2=sigsq2-sig12*sig12/sigsq1;
lambda2=sqrt(lambdasq2);
mp3=(delta*delta+sig2bt+sig2wt-sig2br-sig2wr)/max(sig2br+sig2wr,0.04);
****Compute simulated values****;
junk=int(100000*ranuni(&seed));
do n=&n1,&n2,&n3,&n4;
*o n=5 to 15 by 2;
nu=2*(n-1);
```

```
k=1/sqrt(2*n);
do i=1 to &nsim;
z=normal(0);
u1=2*rangam(0,nu/2);
u2=2*rangam(0,(nu-1)/2);
ut=2*rangam(0,nu/2);
ur=2*rangam(0,nu/2);
z12=normal(0);
d=delta+k*sigi*z;
sswt=sig2wt*ut;
sswr=sig2wr*ur;
ss1=sigsq1*u1;
ss2on1=lambdasq2*u2;
b=beta+z12*lambda2/sqrt(ss1);
ss2=b*b*ss1+ss2on1;
sig2i=(ss2+ss1*(1-2*b))/nu;
keep n i nu d ss1 ss2 ss2on1 sig2i sswt sswr b k mp3;
output;
end;
end;
run;

****Compute HSCI****;
data ball; set sim;
s_0=0.04;
eta0=((log(1.25))**2 + 0.02)/s_0;
e_d=d*d;
e_1=ss1/nu;
e_2=sswt/(2*nu);
t=tinv(0.95,nu);
h_d=(abs(d)+k*t*sqrt(sig2i))**2;
qu=cinv(0.95,nu);
ql=cinv(0.05,nu);
h_1=nu*e_1/ql;
h_2=nu*e_2/ql;
u_d=(h_d-e_d)**2;
u_1=(h_1-e_1)**2;
u_2=(h_2-e_2)**2;
if ((ss2/nu+sswr/(2*nu)) gt s_0) then do;
e_3=-(1+eta0)*ss2/nu;
e_4=-(1+eta0)*sswr/(2*nu);
h_3=nu*e_3/qu;
h_4=nu*e_4/qu;
u_3=(h_3-e_3)**2;
u_4=(h_4-e_4)**2;
lmp3=e_d+e_1+e_2+e_3+e_4;
h_mp3=lmp3+sqrt(u_d+u_1+u_2+u_3+u_4);
```

```
end;
else do;
e_3=-1*ss2/nu;
e_4=-1*sswr/(2*nu);
h_3=nu*e_3/qu;
h_4=nu*e_4/qu;
u_3=(h_3-e_3)**2;
u_4=(h_4-e_4)**2;
lmp3=e_d+e_1+e_2+e_3+e_4-s_0*eta0;
h_mp3=lmp3+sqrt(u_d+u_1+u_2+u_3+u_4);
end;
phi=(h_mp3 lt 0);
keep n i lmp3 h_mp3 phi;
run;

proc univariate data=ball noprint;
var phi;
by n;
output out=h3 mean=hsci;
run;

****Compute Ting-Burdick****;
data ting; set sim;
s_0=0.04;
eta0=((log(1.25))**2 + 0.02)/s_0;
stt=2*ss1/nu;
srr=2*ss2/nu;
st=sswt/nu;
sr=sswr/nu;
sn=2*n;
ssd=sn*d*d;
lambda=ssd/(2*sig2i);
nstar=ceil((1+2*lambda)*(1+2*lambda)/(1+4*lambda));
sdt=2*ssd/sn+stt+st;
ndt=ceil(sdt*sdt/(4*ssd*ssd/(sn*sn*nstar)+(stt*stt/nu)+(st*st/nu)));
if (srr+sr gt 2*s_0) then do;
src=srr+sr;
nrc=ceil(src*src/((srr*srr/nu)+(sr*sr/nu)));
f1=finv(0.05,ndt,nu);
f2=finv(0.95,nu,ndt);
f3=finv(0.05,nrc,nu);
f4=finv(0.95,nrc,ndt);
c0=1-2/nrc;
c2=(c0-f3)**2;
c1=((c0*(f2-1))**2)-c2*f2*f2;
c3=(f4-c0)**2;
c4=((c0*(1-f1))**2)-c3*f1*f1;
```

```
rhohat=(sdt-2*sig2i/sn)/src;
vu1=c1*sdt*sdt+4*c2*sig2i*sig2i/(sn*sn);
vu2=c3*sdt*sdt+4*c4*sig2i*sig2i/(sn*sn);
dt=sn*sdt/(2*sig2i);
if (dt le f1) then u_pber=c0*rhohat+sqrt(vu1)/src-1;
else if (dt gt f1) then u_pber=c0*rhohat+sqrt(vu2)/src-1;
chi=(u_pber lt eta0);
end;
else do;
sir=srr+sr+2*sig2i/sn;
nir=ceil(sir*sir/((4*sig2i*sig2i/(sn*sn*nu))+(srr*srr/nu)+(sr*sr/nu)));
h1=(ndt/cinv(0.05,ndt))-1;
g2=1-nir/cinv(0.95,nir);
c5=((1-finv(0.05,ndt,nir))**2-finv(0.05,ndt,nir)
*finv(0.05,ndt,nir)*h1*h1-g2*g2)/finv(0.05,ndt,nir);
vm3=sdt*sdt*h1*h1+sir*sir*g2*g2+c5*sdt*sir;
u_pbec=(sdt-sir+sqrt(vm3))/(2*s_0);
chi=(u_pbec lt eta0);
end;
run;

proc univariate data=ting noprint;
var chi;
output out=mls mean=mls;
by n;
run;

****Compute GPV****;
data call; set sim;
s_0=0.04;
eta0=((log(1.25))**2 + 0.02)/s_0;
xi=0;
do j=1 to &ngp;
z=normal(0);
z12=normal(0);
u1=2*rangam(0,nu/2);
u2=2*rangam(0,(nu-1)/2);
ut=2*rangam(0,nu/2);
ur=2*rangam(0,nu/2);
sig2wt=sswt/ut;
sig2wr=sswr/ur;
sigsq1=ss1/u1;
lambdasq2=ss2on1/u2;
beta=b-z12*sqrt(lambdasq2/ss1);
sigi=sqrt(sigsq1*(beta-1)**2 + lambdasq2);
delta=d-k*sigi*z;
sigsq2=lambdasq2+beta*beta*sigsq1;
```

```
gpvmp3=(delta*delta+sigsq1+sig2wt/2-sigsq2-sig2wr/2)/max(sigsq2+sig2wr/2,s_0);
xi=xi+(gpvmp3 gt eta0);
end;
p=xi/&ngp;
psi05=(p lt 0.05);
psi025=(p lt 0.025);
run;

proc univariate data=call noprint;
var psi05 psi025;
output out=gpv mean=gpv5 gpv25;
by n;
run;
*merge results;
data tests; merge h3 mls gpv;
length test $ 5;
by n;
delta=&del;
r=&r;
sigma_bt=&sbt;
sigma_br=&sbr;
sigma_wt=&swt;
sigma_wr=&swr;
test="&ds";
run;

proc print data=tests;
run;

proc append base=mi3.power data=tests;
run;

%mend;

%simpbe(0.05,0.90,0.15,0.15,0.15,0.15,5000,10000,20297,6,7,8,9,p1a);
%simpbe(0.05,0.95,0.15,0.15,0.15,0.15,5000,10000,41659,6,7,8,9,p1b);
%simpbe(0.05,0.90,0.30,0.30,0.15,0.15,5000,10000,99487,5,6,9,10,p1c);
%simpbe(0.05,0.95,0.30,0.30,0.15,0.15,5000,10000,56821,5,8,9,10,p1d);
%simpbe(0.05,0.90,0.23,0.23,0.23,0.23,5000,10000,81847,7,8,9,10,p2a);
%simpbe(0.05,0.95,0.23,0.23,0.23,0.23,5000,10000,11597,7,8,9,10,p2b);
%simpbe(0.05,0.90,0.46,0.46,0.23,0.23,5000,10000,29059,6,8,9,10,p2c);
%simpbe(0.05,0.95,0.46,0.46,0.23,0.23,5000,10000,18593,5,6,9,10,p2d);
%simpbe(0.05,0.90,0.3,0.3,0.3,0.3,5000,10000,60737,7,8,9,10,p3a);
%simpbe(0.05,0.95,0.3,0.3,0.3,0.3,5000,10000,88069,7,8,9,10,p3b);
%simpbe(0.05,0.90,0.6,0.6,0.3,0.3,5000,10000,60077,6,8,9,10,p3c);
%simpbe(0.05,0.95,0.6,0.6,0.3,0.3,5000,10000,47629,5,8,9,10,p3d);
%simpbe(0.05,0.90,0.5,0.5,0.5,0.5,5000,10000,62299,7,8,9,10,p4a);
```

```
%simpbe(0.05,0.95,0.5,0.5,0.5,0.5,5000,10000,50503,7,8,9,10,p4b);
%simpbe(0.05,0.90,1.0,1.0,0.5,0.5,5000,10000,77081,6,7,9,10,p4c);
%simpbe(0.05,0.95,1.0,1.0,0.5,0.5,5000,10000,48673,5,6,8,9,p4d);

proc print;
run;

/************************************************
DRUG15A_IBE.SAS
This program does HSCI and GPV procedures
for the IBE/2S3P of Drug 15a (FDA bioequivalence website).
This is the example in Chapter 4, Section 3.4.
************************************************/
title 'Drug 15a/IBE';

proc format;
value $tr 'A'= 'Test'
'B'= 'Reference';
run;
*input and transform data;
data drg15a;
input SUB SEQ PERiod TRT $ AUC AUCINF Cmax;
la=log(auc);
cards;
1 2 1 A 1106.94 1202.9 59.9
4 3 1 A 347.98 422.54 30.4
5 3 1 A 698.06 . 17.1
8 2 1 A 1018.47 1313.37 61.2
9 2 1 A 306.72 469.78 20.06
12 3 1 A 599.21 . 34.4
14 3 1 A 1407.79 1572.95 64.5
15 2 1 A 1280.25 1361.5 55
17 2 1 A 616.23 728.25 21.7
20 3 1 A 552.77 691.04 28.6
22 3 1 A 2097.65 2260.71 83.9
23 2 1 A 799.25 911.75 40.8
25 2 1 A 924.2 1169.2 35.9
28 3 1 A 972.39 1115.8 34.1
29 3 1 A 419.84 . 19.4
32 2 1 A 771.03 945.15 38.2
33 2 1 A 1325.16 1482.57 38.9
36 3 1 A 762.03 897.19 39.7
37 2 1 A 639.65 752.84 23.5
40 3 1 A 1354.86 1408.89 86.8
2 1 2 A 885.09 1008.28 49.7
3 4 2 A 424.88 . 37.9
6 4 2 A 535.15 602.41 36.7
```

```
7 1 2 A 940.62 1100.05 36.3
10 1 2 A 780.57 900.29 26.3
11 4 2 A 90.57 142.87 14.4
13 4 2 A 1547.94 1664.23 73.9
16 1 2 A 974.45 1051.02 66
18 1 2 A 484.45 . 39.5
19 4 2 A 455.77 . 34.9
21 4 2 A 1493.88 1634.19 76.5
24 1 2 A 366.45 507.58 32.5
26 1 2 A 586.84 703.22 36.1
27 4 2 A 594.31 684.04 19.6
30 4 2 A 387.06 474.18 26.8
31 1 2 A 265.83 340.09 18
34 1 2 A 347.75 413.17 29.7
35 4 2 A 874 987.19 45.2
38 1 2 A 1222.18 . 61.3
39 4 2 A 1115.72 1267.88 28.8
2 1 3 A 886.29 1000.06 51.5
4 3 3 A 287.53 390.71 21.3
5 3 3 A 412.79 734.64 15.9
7 1 3 A 697.91 783.85 26
10 1 3 A 595.93 733.4 18
12 3 3 A 487.32 605.44 25.8
14 3 3 A 1534.55 1652.35 57.1
16 1 3 A 630.64 689.1 53
18 1 3 A 509.89 . 33.3
20 3 3 A 498.62 797.48 22.7
22 3 3 A 2317.14 2428.52 90.7
24 1 3 A 392.21 464.99 22
26 1 3 A 748.53 . 34.1
28 3 3 A 598.77 703.39 32
29 3 3 A 516.08 646.47 25.4
31 1 3 A 296.91 423.58 21.8
34 1 3 A 493.64 575.14 42
36 3 3 A 644.77 757.58 24.3
38 1 3 A 1090.89 1206.01 35.5
40 3 3 A 1242.51 1373.88 64
2 1 1 B 1059.96 1139.74 69
3 4 1 B 850.25 1052.56 45.3
6 4 1 B 465.81 545.11 27.6
7 1 1 B 653.84 804.39 32.4
10 1 1 B 804.72 1139.48 34.9
11 4 1 B 192.56 . 13.1
13 4 1 B 1508.07 1674.98 88.4
16 1 1 B 759.39 823.6 63.4
18 1 1 B 372.3 436.25 30.4
19 4 1 B 580.41 669.93 31.9
```

```
21 4 1 B 1640.85 1821.26 51.2
24 1 1 B 131.75 167.44 26.2
26 1 1 B 480.88 605.3 28.5
27 4 1 B 627.77 744.34 22.7
30 4 1 B 383.13 430.21 29.8
31 1 1 B 369.67 . 27.2
34 1 1 B 424.16 463.96 54.4
35 4 1 B 710.32 893.89 32.8
38 1 1 B 1138.46 1272.11 80.3
39 4 1 B 1171.83 1331.54 46.7
1 2 2 B 1024.15 1263.21 47.4
4 3 2 B 254.24 362.69 16.1
5 3 2 B 1114.16 1191.76 61.6
8 2 2 B 1484.06 1822.64 50.7
9 2 2 B 569.68 649.36 31.66
12 3 2 B 719.55 889.94 45.9
14 3 2 B 1459.8 1547.43 88.2
15 2 2 B 1128.02 1304.21 55.2
17 2 2 B 666.95 821.53 36.2
20 3 2 B 535.32 805.95 24.1
22 3 2 B 2485.93 2620.67 97.3
23 2 2 B 883.48 979.3 48.6
25 2 2 B 842.02 1005.35 29.9
28 3 2 B 1092.78 1369.75 50.8
29 3 2 B 422.41 573.86 22.9
32 2 2 B 795.97 971.09 37.4
33 2 2 B 1628.5 1762.04 71.8
36 3 2 B 737.82 884.44 54.4
37 2 2 B 644.97 . 31.4
40 3 2 B 1114.65 1239.1 67.4
1 2 3 B 247.03 . 11.9
3 4 3 B 1061.37 1165.1 57.8
6 4 3 B 424.78 . 37.6
8 2 3 B 1371.62 1489.25 54.3
9 2 3 B 548.73 849.72 28.19
11 4 3 B 253.92 . 20.7
13 4 3 B 1461.61 1576.15 77
15 2 3 B 1325.94 1439.3 76.3
17 2 3 B 473.69 633.78 27.6
19 4 3 B 410.86 574.26 24.7
21 4 3 B 1336.93 1484.17 52.3
23 2 3 B 727.74 913.9 42.3
25 2 3 B 762.28 1011.55 26.8
27 4 3 B 259.8 . 17.6
30 4 3 B 318.12 441.32 17.6
32 2 3 B 737.7 859.5 37
33 2 3 B 1174.89 1351.98 50.4
```

```
35 4 3 B 851.68 942 51
37 2 3 B 526.54 . 24.4
39 4 3 B 1120.5 1240.08 42.2
;;;;
run;

proc sort;
by seq sub;
run;
*compute summary statistics;
proc transpose data=drg15a out=t15a prefix=c;
var la;
id period;
by seq sub;
run;

data hw; set t15a;
if seq=1 then do;
v1=(c2+c3)/2 - c1;
u1=(c2-c3)/sqrt(2);
end;
if seq=3 then do;
v1=(c1+c3)/2 - c2;
u1=(c1-c3)/sqrt(2);
end;
if seq=2 then do;
v2=c1-(c2+c3)/2;
u2=(c2-c3)/sqrt(2);
end;
if seq=4 then do;
v2=c2-(c1+c3)/2;
u2=(c1-c3)/sqrt(2);
end;
run;

proc univariate noprint;
var v1 v2 u1 u2;
by seq;
output out=st mean=i1-i4 css=ss1-ss4;
run;

proc univariate noprint;
var i1 i2 ss1-ss4;
output out=st2 sum=i1 i2 ssi1 ssi2 sswt sswr;
run;

data tt; set st2;
```

```
nu=18;
k=sqrt(1/40);
d=(i1+i2)/4;
sig2i1=ssi1/nu;
sig2i2=ssi2/nu;
swr=sswr/nu;
nui=(sig2i1/(nu+2) + sig2i2/(nu+2))**2/((sig2i1/(nu+2))**2/nu
+(sig2i2/(nu+2))**2/nu);
run;
*compute HSCI;
data ball; set tt;
mi1=d*d+(sig2i1+sig2i2)/2+0.25*sswt/nu-1.75*sswr/nu;
mi3=mi1/max(swr,0.04);
s0=0.04;
eta0=(log(1.25)**2+0.05)/s0;
e_i1=sig2i1/2;
e_i2=sig2i2/2;
e_d=d*d;
e_t=sswt/(4*nu);
e_r=(-1.75-eta0*(swr gt 0.04))*sswr/nu;
t=tinv(0.95,nui);
h_d=(abs(d)+t*k*sqrt((sig2i1+sig2i2)/2))**2;
qu=cinv(0.95,nu);
ql=cinv(0.05,nu);
h_i1=nu*e_i1/ql;
h_i2=nu*e_i2/ql;
h_t=nu*e_t/ql;
h_r=nu*e_r/qu;
u_d=(h_d-e_d)**2;
u_i1=(h_i1-e_i1)**2;
u_i2=(h_i2-e_i2)**2;
u_t=(h_t-e_t)**2;
u_r=(h_r-e_r)**2;
h_mi3=e_d+e_i1+e_i2+e_t+e_r-0.04*eta0*(swr le 0.04);
l_mi3=h_mi3+sqrt(u_d+u_i1+u_i2+u_t+u_r);
run;

proc print data=ball;
run;
*compute GPV;
data call; set tt;
s0=0.04;
eta0=(log(1.25)**2+0.05)/s0;
do i=1 to 50000;
z=normal(0);
u1=2*rangam(0,nu/2);
u2=2*rangam(0,nu/2);
```

```
v=2*rangam(0,nu/2);
w=2*rangam(0,nu/2);
e_i=(ssi1/u1 + ssi2/u2)/2;
e_d=d-k*z*sqrt(e_i);
e_t=sswt/v;
e_r=sswr/w;
piv=((e_d*e_d)+e_i+0.25*e_t-1.75*e_r)/(max(e_r,s0));
piva=((e_d*e_d)+ssi1/u1-1.5*e_r)/(max(e_r,s0));
pivb=((e_d*e_d)+ssi2/u2+0.5*e_t-2*e_r)/(max(e_r,s0));
piv2=(e_d*e_d)+e_i+0.25*e_t-1.75*e_r-eta0*max(e_r,s0);
pivd=e_i-0.75*e_t-0.75*e_r;
phi=(piv gt eta0);
phi2=(piva gt eta0);
phi3=(pivb gt eta0);
psi=(pivd gt 0.04);
output;
end;

proc univariate;
var phi piv;
run;

/************************************************
DRUG15A_PBE.SAS
This program does HSCI and GPV procedures
for the PBE/2S3P of Drug 15a (FDA bioequivalence website).
This is the example in Chapter 4, Section 3.4.
************************************************/
title;
title 'Drug 15a/PBE';

proc format;
value $tr 'A'= 'Test'
'B'= 'Reference';
run;
*input and transform data;
data drg15a;
input SUB SEQ PERiod TRT $ AUC AUCINF Cmax;
la=log(auc);
cards;
1 2 1 A 1106.94 1202.9 59.9
4 3 1 A 347.98 422.54 30.4
5 3 1 A 698.06 . 17.1
8 2 1 A 1018.47 1313.37 61.2
9 2 1 A 306.72 469.78 20.06
12 3 1 A 599.21 . 34.4
14 3 1 A 1407.79 1572.95 64.5
```

```
15 2 1 A 1280.25 1361.5 55
17 2 1 A 616.23 728.25 21.7
20 3 1 A 552.77 691.04 28.6
22 3 1 A 2097.65 2260.71 83.9
23 2 1 A 799.25 911.75 40.8
25 2 1 A 924.2 1169.2 35.9
28 3 1 A 972.39 1115.8 34.1
29 3 1 A 419.84 . 19.4
32 2 1 A 771.03 945.15 38.2
33 2 1 A 1325.16 1482.57 38.9
36 3 1 A 762.03 897.19 39.7
37 2 1 A 639.65 752.84 23.5
40 3 1 A 1354.86 1408.89 86.8
2 1 2 A 885.09 1008.28 49.7
3 4 2 A 424.88 . 37.9
6 4 2 A 535.15 602.41 36.7
7 1 2 A 940.62 1100.05 36.3
10 1 2 A 780.57 900.29 26.3
11 4 2 A 90.57 142.87 14.4
13 4 2 A 1547.94 1664.23 73.9
16 1 2 A 974.45 1051.02 66
18 1 2 A 484.45 . 39.5
19 4 2 A 455.77 . 34.9
21 4 2 A 1493.88 1634.19 76.5
24 1 2 A 366.45 507.58 32.5
26 1 2 A 586.84 703.22 36.1
27 4 2 A 594.31 684.04 19.6
30 4 2 A 387.06 474.18 26.8
31 1 2 A 265.83 340.09 18
34 1 2 A 347.75 413.17 29.7
35 4 2 A 874 987.19 45.2
38 1 2 A 1222.18 . 61.3
39 4 2 A 1115.72 1267.88 28.8
2 1 3 A 886.29 1000.06 51.5
4 3 3 A 287.53 390.71 21.3
5 3 3 A 412.79 734.64 15.9
7 1 3 A 697.91 783.85 26
10 1 3 A 595.93 733.4 18
12 3 3 A 487.32 605.44 25.8
14 3 3 A 1534.55 1652.35 57.1
16 1 3 A 630.64 689.1 53
18 1 3 A 509.89 . 33.3
20 3 3 A 498.62 797.48 22.7
22 3 3 A 2317.14 2428.52 90.7
24 1 3 A 392.21 464.99 22
26 1 3 A 748.53 . 34.1
28 3 3 A 598.77 703.39 32
```

```
29 3 3 A 516.08 646.47 25.4
31 1 3 A 296.91 423.58 21.8
34 1 3 A 493.64 575.14 42
36 3 3 A 644.77 757.58 24.3
38 1 3 A 1090.89 1206.01 35.5
40 3 3 A 1242.51 1373.88 64
2 1 1 B 1059.96 1139.74 69
3 4 1 B 850.25 1052.56 45.3
6 4 1 B 465.81 545.11 27.6
7 1 1 B 653.84 804.39 32.4
10 1 1 B 804.72 1139.48 34.9
11 4 1 B 192.56 . 13.1
13 4 1 B 1508.07 1674.98 88.4
16 1 1 B 759.39 823.6 63.4
18 1 1 B 372.3 436.25 30.4
19 4 1 B 580.41 669.93 31.9
21 4 1 B 1640.85 1821.26 51.2
24 1 1 B 131.75 167.44 26.2
26 1 1 B 480.88 605.3 28.5
27 4 1 B 627.77 744.34 22.7
30 4 1 B 383.13 430.21 29.8
31 1 1 B 369.67 . 27.2
34 1 1 B 424.16 463.96 54.4
35 4 1 B 710.32 893.89 32.8
38 1 1 B 1138.46 1272.11 80.3
39 4 1 B 1171.83 1331.54 46.7
1 2 2 B 1024.15 1263.21 47.4
4 3 2 B 254.24 362.69 16.1
5 3 2 B 1114.16 1191.76 61.6
8 2 2 B 1484.06 1822.64 50.7
9 2 2 B 569.68 649.36 31.66
12 3 2 B 719.55 889.94 45.9
14 3 2 B 1459.8 1547.43 88.2
15 2 2 B 1128.02 1304.21 55.2
17 2 2 B 666.95 821.53 36.2
20 3 2 B 535.32 805.95 24.1
22 3 2 B 2485.93 2620.67 97.3
23 2 2 B 883.48 979.3 48.6
25 2 2 B 842.02 1005.35 29.9
28 3 2 B 1092.78 1369.75 50.8
29 3 2 B 422.41 573.86 22.9
32 2 2 B 795.97 971.09 37.4
33 2 2 B 1628.5 1762.04 71.8
36 3 2 B 737.82 884.44 54.4
37 2 2 B 644.97 . 31.4
40 3 2 B 1114.65 1239.1 67.4
1 2 3 B 247.03 . 11.9
```

```
3 4 3 B 1061.37 1165.1 57.8
6 4 3 B 424.78 . 37.6
8 2 3 B 1371.62 1489.25 54.3
9 2 3 B 548.73 849.72 28.19
11 4 3 B 253.92 . 20.7
13 4 3 B 1461.61 1576.15 77
15 2 3 B 1325.94 1439.3 76.3
17 2 3 B 473.69 633.78 27.6
19 4 3 B 410.86 574.26 24.7
21 4 3 B 1336.93 1484.17 52.3
23 2 3 B 727.74 913.9 42.3
25 2 3 B 762.28 1011.55 26.8
27 4 3 B 259.8 . 17.6
30 4 3 B 318.12 441.32 17.6
32 2 3 B 737.7 859.5 37
33 2 3 B 1174.89 1351.98 50.4
35 4 3 B 851.68 942 51
37 2 3 B 526.54 . 24.4
39 4 3 B 1120.5 1240.08 42.2
;;;;
run;

proc sort;
by seq sub;
run;
*compute summary statistics;
proc transpose data=drg15a out=t15a prefix=c;
var la;
id period;
by seq sub;
run;

data hw; set t15a;
if seq=1 then do;
vt1=(c2+c3)/2;
vr1=c1;
uwt=(c2-c3)/sqrt(2);
end;
if seq=3 then do;
vt1=(c1+c3)/2;
vr1=c2;
uwt=(c1-c3)/sqrt(2);
end;
if seq=2 then do;
vt2=c1;
vr2=(c2+c3)/2;
uwr=(c2-c3)/sqrt(2);
```

```
end;
if seq=4 then do;
vt2=c2;
vr2=(c1+c3)/2;
uwr=(c1-c3)/sqrt(2);
end;
i1=vt1-vr1;
i2=vt2-vr2;
run;

proc univariate noprint;
var i1 i2 vt1 vr1 vt2 vr2 uwt uwr;
by seq;
output out=st mean=i1-i8 css=ss1-ss8;
run;

data xp1; merge hw(in=a) st;
where seq in (1,3);
by seq;
if a;
if first.seq then sxp1=0;
xp=(vt1-i3)*(vr1-i4);
sxp1=sxp1+xp;
retain sxp1;
if last.seq then output;
run;

data xp2; merge hw(in=a) st;
where seq in (2,4);
by seq;
if a;
if first.seq then sxp2=0;
xp=(vt2-i5)*(vr2-i6);
sxp2=sxp2+xp;
retain sxp2;
if last.seq then output;
run;

data xp; merge xp1 xp2;
by seq;
run;

proc univariate noprint;
var i1 i2 ss1-ss8 sxp1 sxp2;
output out=st2 sum=i1 i2 ssi1 ssi2 sst1 ssr1 sst2 ssr2 sswt sswr sstr1 sstr2;
run;
```

```
data tt; set st2;
nu=18;
k=sqrt(1/40);
d=(i1+i2)/4;
sig2i1=ssi1/nu;
sig2i2=ssi2/nu;
b1=sstr1/sst1;
b2=sstr2/sst2;
ssrt1=ssr1-sstr1*sstr1/sst1;
ssrt2=ssr2-sstr2*sstr2/sst2;
sr=0.5*(ssr1/nu+ssr2/nu+sswr/(2*nu));
nui=(sig2i1/(nu+2) + sig2i2/(nu+2))**2/((sig2i1/(nu+2))**2/nu
+(sig2i2/(nu+2))**2/nu);
run;
*compute HSCI;
data ball; set tt;
mi1=d*d+0.5*(sst1/nu + sst2/nu + 0.5*sswt/nu)-sr;
mi3=mi1/max(sr,0.04);
s0=0.04;
eta0=(log(1.25)**2+0.02)/s0;
e_d=d*d;
e_t1=sst1/(2*nu);
e_t2=sst2/(2*nu);
e_wt=sswt/(4*nu);
e_r1=-(0.5+eta0)*ssr1/nu;
e_r2=-(0.5+eta0)*ssr2/nu;
e_wr=-(0.25+eta0)*sswr/nu;
t=tinv(0.95,nui);
h_d=(abs(d)+t*k*sqrt((sig2i1+sig2i2)/2))**2;
qu=cinv(0.95,nu);
ql=cinv(0.05,nu);
h_t1=nu*e_t1/ql;
h_t2=nu*e_t2/ql;
h_wt=nu*e_wt/ql;
h_r1=nu*e_r1/qu;
h_r2=nu*e_r2/qu;
h_wr=nu*e_wr/qu;
u_d=(h_d-e_d)**2;
u_t1=(h_t1-e_t1)**2;
u_t2=(h_t2-e_t2)**2;
u_wt=(h_wt-e_wt)**2;
u_r1=(h_r1-e_r1)**2;
u_r2=(h_r2-e_r2)**2;
u_wr=(h_wr-e_wr)**2;
h_mi3=e_d+e_t1+e_t2+e_wt+e_r1+e_r2+e_wr;
l_mi3=h_mi3+sqrt(u_d+u_t1+u_t2+u_wt+u_r1+u_r2+u_wr);
run;
```

```
proc print data=ball;
run;
*compute GPV;
data call; set tt;
s0=0.04;
eta0=(log(1.25)**2+0.02)/s0;
do i=1 to 250000;
z=normal(0);
zb1=normal(0);
zb2=normal(0);
ut1=2*rangam(0,nu/2);
ut2=2*rangam(0,nu/2);
ur1=2*rangam(0,(nu-1)/2);
ur2=2*rangam(0,(nu-1)/2);
wt=2*rangam(0,nu/2);
wr=2*rangam(0,nu/2);
be1=b1-zb1*sqrt(ssrt1/(ur1*sst1));
be2=b2-zb2*sqrt(ssrt2/(ur2*sst2));
t_d=(sst1*(1-be1)**2 + ssrt1/ur1 + sst2*(1-be2)**2 + ssrt2/ur2)/20;
e_d=d-k*z*sqrt(t_d);
e_t=sst1/ut1 + sst2/ut2 + sswt/(2*wt);
e_r=(sst1/ut1)*be1*be1 + ssrt1/ur1 + (sst2/ut2)*be2*be2 + ssrt2/ur2
+sswr/(2*wr);
piv=((e_d*e_d)+0.5*e_t-0.5*e_r)/(max(0.5*e_r,s0));
phi=(piv gt eta0);
output;
end;

proc univariate;
var phi piv;
run;

/***********************************************
MPH_IBE3.SAS
This program does HSCI and GPV procedures
for the IBE/3S3P of methylphenidate meta-data
(adapted from Meyer et al 2000).
This is the example in Chapter 4, Section 3.8.
***********************************************/
options ps=66 ls=120 nodate nonumber;
title 'Ritalin data/PBE';
*libname mi3 'j:\mi3';
run;
*input data from Excel file provided by Meyer and Straughn;
PROC IMPORT OUT= WORK.mph
            DATAFILE= "J:\mi3\MPH.xls"
```

```
          DBMS=EXCEL2000 REPLACE;
     GETNAMES=YES;
RUN;
*transform and redefine the meta-data;
proc transpose data=mph out=tm prefix=c;
var auc;
id week;
by sequence subject;
run;
*compute preliminary and summary statistics;
data hw5; set tm;
la1 = log(c1);
la2 = log(c2);
la3 = log(c3);
la4 = log(c4);
if sequence=3 then delete;
if sequence=1 then do;
   i=la3-(la1+la2)/2;
   vr=(la1-la2)/sqrt(2);
   end;
if sequence=4 then do;
   i=la1-(la2+la3)/2;
   vr=(la2-la3)/sqrt(2);
   end;
if sequence=2 then do;
   i=la2-(la1+la4)/2;
   vr=(la1-la4)/sqrt(2);
   end;
run;

proc sort data=hw5;
by sequence;

proc univariate data=hw5 noprint;
var i vr;
output out=st mean=i1-i2 css=s1-s2;
by sequence;
run;

proc univariate noprint;
var i1 s1-s2;
output out=st2 sum=i ssi sswr;
run;

data st2; set st2;
mwr=sswr/12;
mi=ssi/12;
```

```
d=i/3;
run;
*compute HSCI;
data ball; set st2;
s_0=0.04;
eta0=((log(1.25))**2 + 0.05)/s_0;
e_d=d*d;
e_1=mi;
e_3=-3*mwr/2;
e_3r=-(1.5+eta0)*mwr;
t=tinv(0.95,12);
h_d=(abs(d)+t*sqrt(0.6*mi)/3)**2;
qu=cinv(0.95,12);
ql=cinv(0.05,12);
h_1=12*e_1/ql;
h_3=12*e_3/qu;
h_3r=12*e_3r/qu;
u_d=(h_d-e_d)**2;
u_1=(h_1-e_1)**2;
u_3=(h_3-e_3)**2;
u_3r=(h_3r-e_3r)**2;
mi3=(e_d+e_1+e_3)/s_0;
mi3b=(e_d+e_1+e_3)/mwr;
lmi3=e_d+e_1+e_3-eta0*s_0;
h_mi3c=lmi3+sqrt(u_d+u_1+u_3);
mi3r=(e_d+e_1)/mwr-1.5;
lmi3r=e_d+e_1+e_3r;
h_mi3r=lmi3r+sqrt(u_d+u_1+u_3r);
run;

proc print data=ball;
run;
*compute GPV;
data call; set st2;
nu=12;
k=sqrt(0.6/9);
s_0=0.04;
eta0=((log(1.25))**2 + 0.05)/s_0;
do j=1 to 50000;
z=normal(0);
ui=2*rangam(0,nu/2);
ur=2*rangam(0,nu/2);
sig2wr=sswr/ur;
sigi=sqrt(ssi/ui);
delta=d-k*sigi*z;
gpvmi3=(delta*delta+sigi*sigi-3*sig2wr/2)/max(sig2wr,s_0);
psi=(gpvmi3 gt eta0);
```

```
output;
end;
run;

proc univariate;
var psi gpvmi3;
run;

/****************************************
MPH_PBE3.SAS
This program does HSCI and GPV procedures for the PBE/3S3P
of men (Meyer et al 2000).
This is Example 1 in Chapter 4, Section 5.1.
****************************************/
options ps=64 ls=120 nodate nonumber;
title 'Ritalin data/PBE';
run;
*input Excel data file provided by Meyer and Straughn;
PROC IMPORT OUT= WORK.mph
            DATAFILE= "c:\my documents\my sas files\mph.xls"
            DBMS=EXCEL2000 REPLACE;
     GETNAMES=YES;
RUN;

proc transpose data=mph out=tm prefix=c;
var auc;            ****select parameter AUC/Cmax;
id week;
by sequence subject;
run;
*create meta-data set and preliminary statistics;
data hw5; set tm;
la1 = log(c1);
la2 = log(c2);
la3 = log(c3);
la4 = log(c4);
if sequence=3 then delete;
if sequence=1 then do;
   ur=(la1+la2)/2;
   ut=la3;
   vr=(la1-la2)/sqrt(2);
   end;
if sequence=4 then do;
   ut=la1;
   ur=(la2+la3)/2;
   vr=(la2-la3)/sqrt(2);
   end;
if sequence=2 then do;
```

```
   ut=la2;
   ur=(la1+la4)/2;
   vr=(la1-la4)/sqrt(2);
   end;
   i=ut-ur;
run;
*compute summary statistics;
proc glm data=hw5 outstat=tx(where=(_type_='ERROR'));
class sequence;
model ur=sequence ut / solution;
run;
quit;

proc sort data=hw5;
by sequence;

proc univariate data=hw5 noprint;
var ut ur vr i;
output out=st mean=i1-i4 css=s1-s4;
by sequence;
run;

data xp; merge hw5(in=a) st;
by sequence;
if a;
if first.sequence then sxp=0;
xp=(ut-i1)*(ur-i2);
sxp=sxp+xp;
retain sxp;
if last.sequence then output;
run;

proc univariate noprint;
var i4 s1-s4 sxp;
output out=st2 sum=i sst ssr sswr ssi ss12;
run;

data st2; set st2;
mut=sst/12;
mur=ssr/12;
mwr=sswr/12;
mi=ssi/12;
b=ss12/sst;
d=i/3;
run;
*Compute HSCI;
data ball; set st2;
```

```
s_0=0.04;
eta0=((log(1.25))**2 + 0.02)/s_0;
e_d=d*d;
e_1=mut;
e_3=-(1+eta0)*mur;
e_4=-(0.5+eta0)*mwr;
t=tinv(0.95,12);
h_d=(abs(d)+t*sqrt(0.6*mi)/3)**2;
qu=cinv(0.95,12);
ql=cinv(0.05,12);
h_1=12*e_1/ql;
h_3=12*e_3/qu;
h_4=12*e_4/qu;
u_d=(h_d-e_d)**2;
u_1=(h_1-e_1)**2;
u_3=(h_3-e_3)**2;
u_4=(h_4-e_4)**2;
mp3=(e_d+e_1-mur-mwr/2)/(mur+mwr/2);
lmp3=e_d+e_1+e_3+e_4;
h_mp3=lmp3+sqrt(u_d+u_1+u_3+u_4);
run;

proc print data=ball;
run;
*compute GPV;
data call; merge st2 tx;
nu=12;
k=sqrt(0.6/12);
ss2on1=ss;
s_0=0.04;
eta0=((log(1.25))**2 + 0.02)/s_0;
do j=1 to 100000;
z=normal(0);
z12=normal(0);
u1=2*rangam(0,nu/2);
u2=2*rangam(0,(nu-1)/2);
ur=2*rangam(0,nu/2);
sig2wr=sswr/ur;
sigsq1=sst/u1;
lambdasq2=ss2on1/u2;
beta=b-z12*sqrt(lambdasq2/sst);
sigi=sqrt(sigsq1*(beta-1)**2 + lambdasq2);
delta=d-k*sigi*z;
sigsq2=lambdasq2+beta*beta*sigsq1;
gpvmp3=(delta*delta+sigsq1-sigsq2-sig2wr/2)/max(sigsq2+sig2wr/2,s_0);
psi=(gpvmp3 gt eta0);
output;
```

```
end;
run;

proc univariate;
var psi gpvmp3;
run;

/*****************************************
SIMIBED3.SAS
This program does the test size simulations
for IBE/2S3P.
Chapter 4, Section 4.1, Table 2.
*****************************************/
options ps=64 ls=120 nodate nonumber nomprint;
libname mi3 'c:\My Documents\My SAS Files\V8';
*ibname mi3 'j:\mi3';
title 'Simulation for PBE/3Pd';
run;

%macro simibe(del,sd,swt,swr,nsim,ngp,seed,ds);

title2 "delta=&del, sigma_d=&sd sigma_wt=&swt sigma_wr=&swr";
*create simulation data sets;
data sim;
****Input model parameter values****;
sigd=&sd;
sigwt=&swt;
sigwr=&swr;
delta=&del;
****Calculate needed parameter values****;
sig2wt=sigwt*sigwt;
sig2wr=sigwr*sigwr;
sig2i1=sigd*sigd+sig2wt+sig2wr/2;
sig2i2=sigd*sigd+sig2wt/2+sig2wr;
mi3=(delta*delta+sigd*sigd+sig2wt-sig2wr)/max(sig2wr,0.04);
call symput('MI3',mi3);
****Compute simulated values****;
junk=int(100000*ranuni(&seed));
do n=12,18,24;
nu=n-1;
k=1/sqrt(2*n);
do j=1 to &nsim;
z=normal(0);
i1=2*rangam(0,nu/2);
i2=2*rangam(0,nu/2);
v=2*rangam(0,nu/2);
w=2*rangam(0,nu/2);
```

```
d=delta+k*sqrt(sig2i1+sig2i2)*z;
sswt=sig2wt*v;
sswr=sig2wr*w;
ssi1=sig2i1*i1;
ssi2=sig2i2*i2;
keep n j nu d ssi1 ssi2 sswt sswr k mi3;
output;
end;
end;
run;
/*
proc univariate;
var d ssi sswt sswr;
by n;
run;
*/
****Compute HSCI****;
data ball; set sim;
s_0=0.04;
eta0=((log(1.25))**2 + 0.05)/s_0;
s2wr=sswr/nu;
m_i1=ssi1/nu;
m_i2=ssi2/nu;
e_d=d*d;
e_1=m_i1/2;
e_2=m_i2/2;
e_t=sswt/(4*nu);
e_r=-(1.75+eta0*(s2wr gt s_0))*s2wr;
nua=(m_i1/n+m_i2/n)**2/(((m_i1/n)**2/nu)+((m_i2/n)**2/nu));
t=tinv(0.95,nua);
h_d=(abs(d)+t*k*sqrt(m_i1+m_i2))**2;
qu=cinv(0.95,nu);
ql=cinv(0.05,nu);
h_i1=nu*e_1/ql;
h_i2=nu*e_2/ql;
h_t=nu*e_t/ql;
h_r=nu*e_r/qu;
u_d=(h_d-e_d)**2;
u_i1=(h_i1-e_1)**2;
u_i2=(h_i2-e_2)**2;
u_t=(h_t-e_t)**2;
u_r=(h_r-e_r)**2;
h_mi3=(e_d+e_1+e_2+e_t+e_r-eta0*s_0*(s2wr le s_0))
+sqrt(u_d+u_i1+u_i2+u_t+u_r);
phi=(h_mi3 lt 0);
*keep n j phi h_mi3 nua t;
run;
```

```
proc univariate data=ball noprint;
var phi nua;
by n;
output out=h3 mean=hsci;
run;
****Compute GPV****;
data call; set sim;
s_0=0.04;
eta0=((log(1.25))**2 + 0.05)/s_0;
xi=0;
do j=1 to &ngp;
z=normal(0);
u1=2*rangam(0,nu/2);
u2=2*rangam(0,nu/2);
v=2*rangam(0,nu/2);
w=2*rangam(0,nu/2);
sig2wt=sswt/v;
sig2wr=sswr/w;
sig2i1=ssi1/u1;
sig2i2=ssi2/u2;
delta=d-k*sqrt(sig2i1+sig2i2)*z;
gpvmi3=(delta*delta+(sig2i1+sig2i2)/2+sig2wt/4-1.75*sig2wr)/max(sig2wr,s_0);
xi=xi+(gpvmi3 gt eta0);
end;
p=xi/&ngp;
psi05=(p lt 0.05);
run;

proc univariate data=call noprint;
var psi05;
output out=gpv mean=gpv;
by n;
run;
*merge results;
data tests; merge h3 gpv;
length test $ 5;
by n;
delta=&del;
sigma_d=&sd;
sigma_wt=&swt;
sigma_wr=&swr;
test="&ds";
mi3=&mi3;
run;

proc print data=tests;
```

```
run;

proc append base=mi3.testsid3 data=tests;
run;

%mend;

%simibe(0.315742,0.01,0.15,0.15,5000,10000,84431,i1a);
%simibe(0.299655,0.1,0.15,0.15,5000,10000,79087,i1b);
%simibe(0.278016,0.15,0.15,0.15,5000,10000,71089,i1c);
%simibe(0.315742,0.01,0.20,0.20,5000,10000,38693,i2a);
%simibe(0.299655,0.1,0.20,0.20,5000,10000,90617,i2b);
%simibe(0.278016,0.15,0.20,0.20,5000,10000,26627,i2c);
%simibe(0.363148,0.01,0.23,0.23,5000,10000,19157,i3a);
%simibe(0.349251,0.1,0.23,0.23,5000,10000,35393,i3b);
%simibe(0.330872,0.15,0.23,0.23,5000,10000,53507,i3c);
%simibe(0.4737450264747,0.01,0.3,0.3,5000,10000,56401,i4a);
%simibe(0.4631785294133,0.1,0.3,0.3,5000,10000,97931,i4b);
%simibe(0.4494823134557,0.15,0.3,0.3,5000,10000,90053,i4c);
%simibe(0.7896876142387,0.01,0.5,0.5,5000,10000,88211,i5a);
%simibe(0.7833942354153,0.1,0.5,0.5,5000,10000,31123,i5b);
%simibe(0.775375088639,0.15,0.5,0.5,5000,10000,36931,i5c);
%simibe(0.05,0.01,0.345966826,0.15,5000,10000,49787,i6a);
%simibe(0.05,0.1,0.331350335,0.15,5000,10000,50909,i6b);
%simibe(0.05,0.15,0.31191833,0.15,5000,10000,51973,i6c);
*simibe(0.05,0.2,0.2824766264545,0.15,5000,10000,34141,i6d);
%simibe(0.05,0.01,0.370395794,0.20,5000,10000,53117,i7a);
%simibe(0.05,0.1,0.356781508,0.20,5000,10000,54277,i7b);
%simibe(0.05,0.15,0.338811223,0.20,5000,10000,55373,i7c);
*simibe(0.05,0.2,0.31191833,0.20,5000,10000,36299,i7d);
%simibe(0.05,0.01,0.426938288,0.23,5000,10000,56479,i8a);
%simibe(0.05,0.1,0.415182251,0.23,5000,10000,57503,i8b);
%simibe(0.05,0.15,0.399845347,0.23,5000,10000,58601,i8c);
*simibe(0.05,0.2,0.3773278433,0.23,5000,10000,38501,i8d);
%simibe(0.05,0.01,0.558510833,0.3,5000,10000,59671,i9a);
%simibe(0.05,0.1,0.549576519,0.3,5000,10000,60887,i9b);
%simibe(0.05,0.15,0.538083962,0.3,5000,10000,62039,i9c);
*simibe(0.05,0.2,0.52156912304,0.3,5000,10000,43577,i9d);
%simibe(0.05,0.01,0.933330878,0.5,5000,10000,65449,i10a);
%simibe(0.05,0.1,0.928012138,0.5,5000,10000,66587,i10b);
%simibe(0.05,0.15,0.921252695,0.5,5000,10000,68711,i10c);
*simibe(0.05,0.2,0.9117052857596,0.5,5000,10000,48131,i10d);

proc print data=mi3.testsid3;
title2;
run;
```

```
/*******************************************
SIMIBED_P.SAS
This program does the power/sample size simulations
for IBE/2S3P.
Chapter 4, Section 4.2, Table 3.
*******************************************/
options ps=64 ls=120 nodate nonumber nomprint;
libname mi3 'c:\My Documents\My SAS Files\V8';
*ibname mi3 'j:\mi3';
title 'Simulation for PBE/3Pd';
run;

%macro simibe(del,sd,swt,swr,nsim,ngp,n1,n2,seed,ds);

title2 "delta=&del, sigma_d=&sd sigma_wt=&swt sigma_wr=&swr";
*create simulation data sets;
data sim;
****Input model parameter values****;
sigd=&sd;
sigwt=&swt;
sigwr=&swr;
delta=&del;
****Calculate needed parameter values****;
sig2wt=sigwt*sigwt;
sig2wr=sigwr*sigwr;
sig2i1=sigd*sigd+sig2wt+sig2wr/2;
sig2i2=sigd*sigd+sig2wt/2+sig2wr;
mi3=(delta*delta+sigd*sigd+sig2wt-sig2wr)/max(sig2wr,0.04);
call symput('MI3',mi3);
****Compute simulated values****;
junk=int(100000*ranuni(&seed));
do n=&n1 to &n2;
nu=n-1;
k=1/sqrt(2*n);
do j=1 to &nsim;
z=normal(0);
i1=2*rangam(0,nu/2);
i2=2*rangam(0,nu/2);
v=2*rangam(0,nu/2);
w=2*rangam(0,nu/2);
d=delta+k*sqrt(sig2i1+sig2i2)*z;
sswt=sig2wt*v;
sswr=sig2wr*w;
ssi1=sig2i1*i1;
ssi2=sig2i2*i2;
keep n j nu d ssi1 ssi2 sswt sswr k mi3;
output;
```

```
end;
end;
run;
/*
proc univariate;
var d ssi sswt sswr;
by n;
run;
*/
****Compute HSCI****;
data ball; set sim;
s_0=0.04;
eta0=((log(1.25))**2 + 0.05)/s_0;
s2wr=sswr/nu;
m_i1=ssi1/nu;
m_i2=ssi2/nu;
e_d=d*d;
e_1=m_i1/2;
e_2=m_i2/2;
e_t=sswt/(4*nu);
e_r=-(1.75+eta0*(s2wr gt s_0))*s2wr;
nua=(m_i1/n+m_i2/n)**2/(((m_i1/n)**2/nu)+((m_i2/n)**2/nu));
t=tinv(0.95,nua);
h_d=(abs(d)+t*k*sqrt(m_i1+m_i2))**2;
qu=cinv(0.95,nu);
ql=cinv(0.05,nu);
h_i1=nu*e_1/ql;
h_i2=nu*e_2/ql;
h_t=nu*e_t/ql;
h_r=nu*e_r/qu;
u_d=(h_d-e_d)**2;
u_i1=(h_i1-e_1)**2;
u_i2=(h_i2-e_2)**2;
u_t=(h_t-e_t)**2;
u_r=(h_r-e_r)**2;
h_mi3=(e_d+e_1+e_2+e_t+e_r-eta0*s_0*(s2wr le s_0))
+sqrt(u_d+u_i1+u_i2+u_t+u_r);
phi=(h_mi3 lt 0);
*keep n j phi h_mi3 nua t;
run;

proc univariate data=ball noprint;
var phi nua;
by n;
output out=h3 mean=hsci;
run;
****Compute GPV****;
```

```
data call; set sim;
s_0=0.04;
eta0=((log(1.25))**2 + 0.05)/s_0;
xi=0;
do j=1 to &ngp;
z=normal(0);
u1=2*rangam(0,nu/2);
u2=2*rangam(0,nu/2);
v=2*rangam(0,nu/2);
w=2*rangam(0,nu/2);
sig2wt=sswt/v;
sig2wr=sswr/w;
sig2i1=ssi1/u1;
sig2i2=ssi2/u2;
delta=d-k*sqrt(sig2i1+sig2i2)*z;
gpvmi3=(delta*delta+(sig2i1+sig2i2)/2+sig2wt/4-1.75*sig2wr)/max(sig2wr,s_0);
xi=xi+(gpvmi3 gt eta0);
end;
p=xi/&ngp;
psi05=(p lt 0.05);
run;

proc univariate data=call noprint;
var psi05;
output out=gpv mean=gpv;
by n;
run;
*merge results;
data tests; merge h3 gpv;
length test $ 5;
by n;
delta=&del;
sigma_d=&sd;
sigma_wt=&swt;
sigma_wr=&swr;
test="&ds";
mi3=&mi3;
run;

proc print data=tests;
run;

proc append base=mi3.testsid3_p data=tests;
run;

%mend;
```

```
%simibe(0.05,0.01,0.15,0.15,5000,10000,8,10,65479,p1a);
%simibe(0.05,0.1,0.15,0.15,5000,10000,11,14,47041,p1b);
%simibe(0.05,0.15,0.15,0.15,5000,10000,17,20,45823,p1c);
%simibe(0.05,0.01,0.20,0.20,5000,10000,15,17,54001,p2a);
%simibe(0.05,0.1,0.20,0.20,5000,10000,21,24,28031,p2b);
%simibe(0.05,0.15,0.20,0.20,5000,10000,29,33,18503,p2c);
%simibe(0.05,0.01,0.23,0.23,5000,10000,21,25,54577,p3a);
%simibe(0.05,0.1,0.23,0.23,5000,10000,28,31,33893,p3b);
%simibe(0.05,0.15,0.23,0.23,5000,10000,41,44,79087,p3c);
%simibe(0.05,0.01,0.3,0.3,5000,10000,28,31,47939,p4a);
%simibe(0.05,0.1,0.3,0.3,5000,10000,31,35,41243,p4b);
%simibe(0.05,0.15,0.3,0.3,5000,10000,36,41,86453,p4c);
%simibe(0.05,0.01,0.5,0.5,5000,10000,27,31,26701,p5a);
%simibe(0.05,0.1,0.5,0.5,5000,10000,28,32,23227,p5b);
%simibe(0.05,0.15,0.5,0.5,5000,10000,30,35,53551,p5c);

proc print data=mi3.testsid3_p;
title2;
run;

/*****************************************
SIMPBED3.SAS
This program does the test size simulations
for PBE/2S3P.
Chapter 4, Section 4.3, Table 4.
*****************************************/
options ps=64 ls=120 nodate nonumber nomprint;
title 'Simulation data';
libname mi3 'c:\My Documents\My SAS Files\V8';
*ibname mi3 'j:\mi3';
title 'Simulation for PBE/Design A';
run;

%macro simpbe(del,r,sbt,sbr,swt,swr,nsim,ngp,seed,ds);

title2 "delta=&del, rho=&r, sigma_bt=&sbt
sigma_br=&sbr sigma_wt=&swt sigma_wr=&swr";
run;
*create simulation data sets;
data sim;
****Input model parameter values****;
sigbt=&sbt;
sigbr=&sbr;
sigwt=&swt;
sigwr=&swr;
delta=&del;
rho=&r;
```

```
****Calculate needed parameter values****;
sig2br=sigbr*sigbr;
sig2bt=sigbt*sigbt;
sig2wr=sigwr*sigwr;
sig2wt=sigwt*sigwt;
sigsqr1=sigbr*sigbr+sigwr*sigwr;
sigsqt1=sig2bt+sig2wt/2;
sigsqt2=sigbt*sigbt+sigwt*sigwt;
sigsqr2=sig2br+sig2wr/2;
sig12=rho*sigbt*sigbr;
beta1=sig12/sigsqt1;
beta2=sig12/sigsqt2;
sigi1=sqrt(sigsqt1+sigsqr1-2*sig12);
sigi2=sqrt(sigsqt2+sigsqr2-2*sig12);
lambdasq1=sigsqr1-sig12*sig12/sigsqt1;
lambda1=sqrt(lambdasq1);
lambdasq2=sigsqr2-sig12*sig12/sigsqt2;
lambda2=sqrt(lambdasq2);
mp3=(delta*delta+sig2bt+sig2wt-sig2br-sig2wr)/max(sig2br+sig2wr,0.04);
call symput('MP3',mp3);
****Compute simulated values****;
junk=int(100000*ranuni(&seed));
do n=12,18,24;
nu=n-1;
do i=1 to &nsim;
z=normal(0);
ut1=2*rangam(0,nu/2);
ur1=2*rangam(0,(nu-1)/2);
uwt=2*rangam(0,nu/2);
z12=normal(0);
d=delta+z*(sigi1+sigi2)/(2*n);
sswt=sig2wt*uwt;
sst1=sigsqt1*ut1;
ssront1=lambdasq1*ur1;
b1=beta1+z12*lambda1/sqrt(sst1);
ssr1=b1*b1*sst1+ssront1;
sig2i1=(ssront1+sst1*(1-b1)*(1-b1))/nu;
ut2=2*rangam(0,nu/2);
ur2=2*rangam(0,(nu-1)/2);
uwr=2*rangam(0,nu/2);
z122=normal(0);
sswr=sig2wr*uwr;
sst2=sigsqt2*ut2;
ssront2=lambdasq2*ur2;
b2=beta2+z122*lambda2/sqrt(sst2);
ssr2=b2*b2*sst1+ssront2;
sig2i2=(ssront2+sst2*(1-b2)*(1-b2))/nu;
```

```
output;
end;
end;
run;
****Compute HSCI****;
data ball; set sim;
s_0=0.04;
eta0=((log(1.25))**2 + 0.02)/s_0;
e_d=d*d;
nuc=(((sig2i1+sig2i2)/n)**2)/((sig2i1**2 + sig2i2**2)/(n*n*(n-1)));
t=tinv(0.95,nuc);
h_d=(abs(d)+t*sqrt((sig2i1+sig2i2)/n)/2)**2;
qu=cinv(0.95,nu);
ql=cinv(0.05,nu);
u_d=(h_d-e_d)**2;
u_t1=(sst1*(1/ql - 1/nu)/2)**2;
u_t2=(sst2*(1/ql - 1/nu)/2)**2;
u_wt=(sswt*(1/ql - 1/nu)/4)**2;
if ((ssr1/nu+ssr2/nu+sswr/(2*nu))/2 gt s_0) then do;
u_r1=(ssr1*(1+eta0)*(1/qu - 1/nu)/2)**2;
u_r2=(ssr2*(1+eta0)*(1/qu - 1/nu)/2)**2;
u_wr=(sswr*(1+eta0)*(1/qu - 1/nu)/4)**2;
lmp3=e_d+(sst1/nu+sst2/nu+sswt/(2*nu))/2-(1+eta0)*
(ssr1/nu+ssr2/nu+sswr/(2*nu))/2;
h_mp3=lmp3+sqrt(u_d+u_t1+u_t2+u_wt+u_r1+u_r2+u_wr);
end;
else do;
u_r1=(ssr1*(1/qu - 1/nu)/2)**2;
u_r2=(ssr2*(1/qu - 1/nu)/2)**2;
u_wr=(sswr*(1/qu - 1/nu)/4)**2;
lmp3=e_d+(sst1/nu+sst2/nu)/2-(ssr1/nu+ssr2/nu+sswr/(2*nu))/2-s_0*eta0;
h_mp3=lmp3+sqrt(u_d+u_t1+u_t2+u_wt+u_r1+u_r2+u_wr);
end;
phi=(h_mp3 lt 0);
keep n i lmp3 h_mp3 phi;
run;

proc univariate data=ball noprint;
var phi;
by n;
output out=h3 mean=hsci;
run;

****Compute GPV****;
data call; set sim;
s_0=0.04;
eta0=((log(1.25))**2 + 0.02)/s_0;
```

```
xi=0;
xi2=0;
do j=1 to &ngp;
z=normal(0);
z12=normal(0);
ut1=2*rangam(0,nu/2);
ur1=2*rangam(0,(nu-1)/2);
uwt=2*rangam(0,nu/2);
sig2wt=sswr/uwt;
sigsqt1=sst1/ut1;
lambdasq1=ssront1/ur1;
beta1=b1-z12*sqrt(lambdasq1/sst1);
sigi1=sigsqt1*(beta1-1)**2 + lambdasq1;
sigsqr1=lambdasq1+beta1*beta1*sigsqt1;
z12=normal(0);
ut2=2*rangam(0,nu/2);
ur2=2*rangam(0,(nu-1)/2);
uwr=2*rangam(0,nu/2);
sig2wr=sswr/uwr;
sigsqt2=sst2/ut2;
lambdasq2=ssront2/ur2;
beta2=b2-z12*sqrt(lambdasq2/sst2);
sigi2=sigsqt2*(beta2-1)**2 + lambdasq2;
sigsqr2=lambdasq2+beta2*beta2*sigsqt1;
delta=d-z*sqrt((sigi1+sigi2)/n)/2;
gpvmp3=(delta*delta+(sigsqt1+sig2wt/2+sigsqt2-sigsqr1-sigsqr2-
sig2wr/2)/2)/max((sigsqr1+sigsqr2+sig2wr/2)/2,s_0);
xi=xi+(gpvmp3 gt eta0);
xi2=xi2+(gpvmp3 gt mp3);
end;
p=xi/&ngp;
c=xi2/&ngp;
psi=(p lt 0.05);
cover=(c gt 0.05);
run;

proc univariate data=call noprint;
var psi cover;
output out=gpv mean=gpv5 cover95;
by n;
run;
*merge results;
data tests; merge h3 gpv;
length test $ 5;
by n;
delta=&del;
r=&r;
```

```
sigma_bt=&sbt;
sigma_br=&sbr;
sigma_wt=&swt;
sigma_wr=&swr;
test="&ds";
mp3=&mp3;
run;run;

proc print data=tests;
run;

proc append base=mi3.testpd3 data=tests;
run;

%mend;

%simpbe(0.886099,0.8,0.6,0.5,0.3,0.4472136,5000,10000,55061,e3a);
%simpbe(0.886099,0.9,0.6,0.5,0.3,0.4472136,5000,10000,72277,e3b);
%simpbe(0.886099,0.8,0.5,0.6,0.4472136,0.3,5000,10000,55061,e3c);
%simpbe(0.886099,0.9,0.5,0.6,0.4472136,0.3,5000,10000,72277,e3d);
%simpbe(0,0.8,0.282476626,0.1,0.1,0.1,5000,10000,75577,f1a);
%simpbe(0,0.9,0.282476626,0.1,0.1,0.1,5000,10000,64621,f1b);
%simpbe(0,0.8,0.356708993,0.2,0.1,0.1,5000,10000,79043,f1c);
%simpbe(0,0.9,0.356708993,0.2,0.1,0.1,5000,10000,86923,f1d);
%simpbe(0,0.8,0.423775989,0.2,0.2,0.2,5000,10000,46867,f2a);
%simpbe(0,0.9,0.423775989,0.2,0.2,0.2,5000,10000,19237,f2b);
%simpbe(0,0.8,0.713417986,0.4,0.2,0.2,5000,10000,31699,f2c);
%simpbe(0,0.9,0.713417986,0.4,0.2,0.2,5000,10000,61211,f2d);
%simpbe(0,0.8,0.635663984,0.3,0.3,0.3,5000,10000,85103,f3a);
%simpbe(0,0.9,0.635663984,0.3,0.3,0.3,5000,10000,65267,f3b);
%simpbe(0,0.8,1.070126979,0.6,0.3,0.3,5000,10000,55061,f3c);
%simpbe(0,0.9,1.070126979,0.6,0.3,0.3,5000,10000,44897,f3d);

proc print;
title2;
run;

/*****************************************
SIMPBED3_P.SAS
This program does the power/sample size simulations
for PBE/2S3P.
Chapter 4, Section 4.4, Table 5.
*****************************************/
options ps=64 ls=120 nodate nonumber nomprint;
title 'Simulation data';
libname mi3 'c:\My Documents\My SAS Files\V8';
*ibname mi3 'j:\mi3';
```

```
title 'Simulation for PBE/Design A';
run;

%macro simpbe(del,r,sbt,sbr,swt,swr,nsim,ngp,seed,n1,n2,n3,n4,ds);

title2 "delta=&del, rho=&r, sigma_bt=&sbt
sigma_br=&sbr sigma_wt=&swt sigma_wr=&swr";
run;
*create simulation data sets;
data sim;
****Input model parameter values****;
sigbt=&sbt;
sigbr=&sbr;
sigwt=&swt;
sigwr=&swr;
delta=&del;
rho=&r;
****Calculate needed parameter values****;
sig2br=sigbr*sigbr;
sig2bt=sigbt*sigbt;
sig2wr=sigwr*sigwr;
sig2wt=sigwt*sigwt;
sigsqr1=sigbr*sigbr+sigwr*sigwr;
sigsqt1=sig2bt+sig2wt/2;
sigsqt2=sigbt*sigbt+sigwt*sigwt;
sigsqr2=sig2br+sig2wr/2;
sig12=rho*sigbt*sigbr;
beta1=sig12/sigsqt1;
beta2=sig12/sigsqt2;
sigi1=sqrt(sigsqt1+sigsqr1-2*sig12);
sigi2=sqrt(sigsqt2+sigsqr2-2*sig12);
lambdasq1=sigsqr1-sig12*sig12/sigsqt1;
lambda1=sqrt(lambdasq1);
lambdasq2=sigsqr2-sig12*sig12/sigsqt2;
lambda2=sqrt(lambdasq2);
mp3=(delta*delta+sig2bt+sig2wt-sig2br-sig2wr)/max(sig2br+sig2wr,0.04);
call symput('MP3',mp3);
****Compute simulated values****;
junk=int(100000*ranuni(&seed));
*o n=9 to 17;
do n=&n1,&n2,&n3,&n4;
nu=n-1;
do i=1 to &nsim;
z=normal(0);
ut1=2*rangam(0,nu/2);
ur1=2*rangam(0,(nu-1)/2);
uwt=2*rangam(0,nu/2);
```

```
z12=normal(0);
d=delta+z*(sigi1+sigi2)/(2*n);
sswt=sig2wt*uwt;
sst1=sigsqt1*ut1;
ssront1=lambdasq1*ur1;
b1=beta1+z12*lambda1/sqrt(sst1);
ssr1=b1*b1*sst1+ssront1;
sig2i1=(ssront1+sst1*(1-b1)*(1-b1))/nu;
ut2=2*rangam(0,nu/2);
ur2=2*rangam(0,(nu-1)/2);
uwr=2*rangam(0,nu/2);
z122=normal(0);
sswr=sig2wr*uwr;
sst2=sigsqt2*ut2;
ssront2=lambdasq2*ur2;
b2=beta2+z122*lambda2/sqrt(sst2);
ssr2=b2*b2*sst1+ssront2;
sig2i2=(ssront2+sst2*(1-b2)*(1-b2))/nu;
output;
end;
end;
run;
****Compute HSCI****;
data ball; set sim;
s_0=0.04;
eta0=((log(1.25))**2 + 0.02)/s_0;
e_d=d*d;
nuc=(((sig2i1+sig2i2)/n)**2)/((sig2i1**2 + sig2i2**2)/(n*n*(n-1)));
t=tinv(0.95,nuc);
h_d=(abs(d)+t*sqrt((sig2i1+sig2i2)/n)/2)**2;
qu=cinv(0.95,nu);
ql=cinv(0.05,nu);
u_d=(h_d-e_d)**2;
u_t1=(sst1*(1/ql - 1/nu)/2)**2;
u_t2=(sst2*(1/ql - 1/nu)/2)**2;
u_wt=(sswt*(1/ql - 1/nu)/4)**2;
if ((ssr1/nu+ssr2/nu+sswr/(2*nu))/2 gt s_0) then do;
u_r1=(ssr1*(1+2*eta0)*(1/qu - 1/nu)/2)**2;
u_r2=(ssr2*(1+2*eta0)*(1/qu - 1/nu)/2)**2;
u_wr=(sswr*(1+2*eta0)*(1/qu - 1/nu)/4)**2;
lmp3=e_d+(sst1/nu+sst2/nu+sswt/(2*nu))/2-
(1+2*eta0)*(ssr1/nu+ssr2/nu+sswr/(2*nu))/2;
h_mp3=lmp3+sqrt(u_d+u_t1+u_t2+u_wt+u_r1+u_r2+u_wr);
end;
else do;
u_r1=(ssr1*(1/qu - 1/nu)/2)**2;
u_r2=(ssr2*(1/qu - 1/nu)/2)**2;
```

```
u_wr=(sswr*(1/qu - 1/nu)/4)**2;
lmp3=e_d+(sst1/nu+sst2/nu)/2-(ssr1/nu+ssr2/nu+sswr/(2*nu))/2-s_0*eta0;
h_mp3=lmp3+sqrt(u_d+u_t1+u_t2+u_wt+u_r1+u_r2+u_wr);
end;
phi=(h_mp3 lt 0);
keep n i lmp3 h_mp3 phi;
run;

proc univariate data=ball noprint;
var phi;
by n;
output out=h3 mean=hsci;
run;

****Compute GPV****;
data call; set sim;
s_0=0.04;
eta0=((log(1.25))**2 + 0.02)/s_0;
xi=0;
xi2=0;
do j=1 to &ngp;
z=normal(0);
z12=normal(0);
ut1=2*rangam(0,nu/2);
ur1=2*rangam(0,(nu-1)/2);
uwt=2*rangam(0,nu/2);
sig2wt=sswr/uwt;
sigsqt1=sst1/ut1;
lambdasq1=ssront1/ur1;
beta1=b1-z12*sqrt(lambdasq1/sst1);
sigi1=sigsqt1*(beta1-1)**2 + lambdasq1;
sigsqr1=lambdasq1+beta1*beta1*sigsqt1;
z12=normal(0);
ut2=2*rangam(0,nu/2);
ur2=2*rangam(0,(nu-1)/2);
uwr=2*rangam(0,nu/2);
sig2wr=sswr/uwr;
sigsqt2=sst2/ut2;
lambdasq2=ssront2/ur2;
beta2=b2-z12*sqrt(lambdasq2/sst2);
sigi2=sigsqt2*(beta2-1)**2 + lambdasq2;
sigsqr2=lambdasq2+beta2*beta2*sigsqt1;
delta=d-z*sqrt((sigi1+sigi2)/n)/2;
gpvmp3=(delta*delta+(sigsqt1+sig2wt/2+sigsqt2-sigsqr1-sigsqr2-
sig2wr/2)/2)/max((sigsqr1+sigsqr2+sig2wr/2)/2,s_0);
xi=xi+(gpvmp3 gt eta0);
xi2=xi2+(gpvmp3 gt mp3);
```

```
end;
p=xi/&ngp;
c=xi2/&ngp;
psi=(p lt 0.05);
cover=(c gt 0.05);
run;

proc univariate data=call noprint;
var psi cover;
output out=gpv mean=gpv5 cover95;
by n;
run;
*merge results;
data tests; merge h3 gpv;
length test $ 5;
by n;
delta=&del;
r=&r;
sigma_bt=&sbt;
sigma_br=&sbr;
sigma_wt=&swt;
sigma_wr=&swr;
test="&ds";
mp3=&mp3;
run;run;

proc print data=tests;
run;

proc append base=mi3.powerpd3 data=tests;
run;

%mend;
/*
%simpbe(0.05,0.90,0.15,0.15,0.15,0.15,5000,10000,20297,8,9,11,12,p1a);
%simpbe(0.05,0.95,0.15,0.15,0.15,0.15,5000,10000,41651,8,9,11,12,p1b);
%simpbe(0.05,0.90,0.30,0.30,0.15,0.15,5000,10000,99487,10,11,14,15,p1c);
%simpbe(0.05,0.95,0.30,0.30,0.15,0.15,5000,10000,56821,10,11,14,15,p1d);
%simpbe(0.05,0.90,0.23,0.23,0.23,0.23,5000,10000,81847,10,11,14,15,p2a);
%simpbe(0.05,0.95,0.23,0.23,0.23,0.23,5000,10000,11597,10,11,14,15,p2b);
%simpbe(0.05,0.90,0.46,0.46,0.23,0.23,5000,10000,29059,10,11,14,15,p2c);
%simpbe(0.05,0.95,0.46,0.46,0.23,0.23,5000,10000,18593,10,11,14,15,p2d);
*/
%simpbe(0.05,0.90,0.3,0.3,0.3,0.3,5000,10000,60737,10,11,13,14,p3a);
%simpbe(0.05,0.95,0.3,0.3,0.3,0.3,5000,10000,88069,10,11,13,14,p3b);
%simpbe(0.05,0.90,0.6,0.6,0.3,0.3,5000,10000,60077,10,11,14,15,p3c);
%simpbe(0.05,0.95,0.6,0.6,0.3,0.3,5000,10000,47629,10,11,14,15,p3d);
```

```
%simpbe(0.05,0.90,0.5,0.5,0.5,0.5,5000,10000,62299,10,11,14,15,p4a);
%simpbe(0.05,0.95,0.5,0.5,0.5,0.5,5000,10000,50503,10,11,14,15,p4b);
%simpbe(0.05,0.90,1.0,1.0,0.5,0.5,5000,10000,77081,10,11,14,15,p4c);
%simpbe(0.05,0.95,1.0,1.0,0.5,0.5,5000,10000,48673,10,11,14,15,p4d);

proc print;
title2;
run;
/*****************************************
SIMPBED3_P.SAS
This program does the power/sample size simulations
for PBE/2S3P.
Chapter 4, Section 4.4, Table 5.
*****************************************/
options ps=64 ls=120 nodate nonumber nomprint;
title 'Simulation data';
libname mi3 'c:\My Documents\My SAS Files\V8';
*ibname mi3 'j:\mi3';
title 'Simulation for PBE/Design A';
run;

%macro simpbe(del,r,sbt,sbr,swt,swr,nsim,ngp,seed,n1,n2,n3,n4,ds);

title2 "delta=&del, rho=&r, sigma_bt=&sbt
sigma_br=&sbr sigma_wt=&swt sigma_wr=&swr";
run;
*create simulation data sets;
data sim;
****Input model parameter values****;
sigbt=&sbt;
sigbr=&sbr;
sigwt=&swt;
sigwr=&swr;
delta=&del;
rho=&r;
****Calculate needed parameter values****;
sig2br=sigbr*sigbr;
sig2bt=sigbt*sigbt;
sig2wr=sigwr*sigwr;
sig2wt=sigwt*sigwt;
sigsqr1=sigbr*sigbr+sigwr*sigwr;
sigsqt1=sig2bt+sig2wt/2;
sigsqt2=sigbt*sigbt+sigwt*sigwt;
sigsqr2=sig2br+sig2wr/2;
sig12=rho*sigbt*sigbr;
beta1=sig12/sigsqt1;
beta2=sig12/sigsqt2;
```

```
sigi1=sqrt(sigsqt1+sigsqr1-2*sig12);
sigi2=sqrt(sigsqt2+sigsqr2-2*sig12);
lambdasq1=sigsqr1-sig12*sig12/sigsqt1;
lambda1=sqrt(lambdasq1);
lambdasq2=sigsqr2-sig12*sig12/sigsqt2;
lambda2=sqrt(lambdasq2);
mp3=(delta*delta+sig2bt+sig2wt-sig2br-sig2wr)/max(sig2br+sig2wr,0.04);
call symput('MP3',mp3);
****Compute simulated values****;
junk=int(100000*ranuni(&seed));
*o n=9 to 17;
do n=&n1,&n2,&n3,&n4;
nu=n-1;
do i=1 to &nsim;
z=normal(0);
ut1=2*rangam(0,nu/2);
ur1=2*rangam(0,(nu-1)/2);
uwt=2*rangam(0,nu/2);
z12=normal(0);
d=delta+z*(sigi1+sigi2)/(2*n);
sswt=sig2wt*uwt;
sst1=sigsqt1*ut1;
ssront1=lambdasq1*ur1;
b1=beta1+z12*lambda1/sqrt(sst1);
ssr1=b1*b1*sst1+ssront1;
sig2i1=(ssront1+sst1*(1-b1)*(1-b1))/nu;
ut2=2*rangam(0,nu/2);
ur2=2*rangam(0,(nu-1)/2);
uwr=2*rangam(0,nu/2);
z122=normal(0);
sswr=sig2wr*uwr;
sst2=sigsqt2*ut2;
ssront2=lambdasq2*ur2;
b2=beta2+z122*lambda2/sqrt(sst2);
ssr2=b2*b2*sst1+ssront2;
sig2i2=(ssront2+sst2*(1-b2)*(1-b2))/nu;
output;
end;
end;
run;
****Compute HSCI****;
data ball; set sim;
s_0=0.04;
eta0=((log(1.25))**2 + 0.02)/s_0;
e_d=d*d;
nuc=(((sig2i1+sig2i2)/n)**2)/((sig2i1**2 + sig2i2**2)/(n*n*(n-1)));
t=tinv(0.95,nuc);
```

```
h_d=(abs(d)+t*sqrt((sig2i1+sig2i2)/n)/2)**2;
qu=cinv(0.95,nu);
ql=cinv(0.05,nu);
u_d=(h_d-e_d)**2;
u_t1=(sst1*(1/ql - 1/nu)/2)**2;
u_t2=(sst2*(1/ql - 1/nu)/2)**2;
u_wt=(sswt*(1/ql - 1/nu)/4)**2;
if ((ssr1/nu+ssr2/nu+sswr/(2*nu))/2 gt s_0) then do;
u_r1=(ssr1*(1+2*eta0)*(1/qu - 1/nu)/2)**2;
u_r2=(ssr2*(1+2*eta0)*(1/qu - 1/nu)/2)**2;
u_wr=(sswr*(1+2*eta0)*(1/qu - 1/nu)/4)**2;
lmp3=e_d+(sst1/nu+sst2/nu+sswt/(2*nu))/2-(1+2*eta0)*(ssr1/nu+ssr2/nu+sswr/(2*nu))/2;
h_mp3=lmp3+sqrt(u_d+u_t1+u_t2+u_wt+u_r1+u_r2+u_wr);
end;
else do;
u_r1=(ssr1*(1/qu - 1/nu)/2)**2;
u_r2=(ssr2*(1/qu - 1/nu)/2)**2;
u_wr=(sswr*(1/qu - 1/nu)/4)**2;
lmp3=e_d+(sst1/nu+sst2/nu)/2-
(ssr1/nu+ssr2/nu+sswr/(2*nu))/2-s_0*eta0;
h_mp3=lmp3+sqrt(u_d+u_t1+u_t2+u_wt+u_r1+u_r2+u_wr);
end;
phi=(h_mp3 lt 0);
keep n i lmp3 h_mp3 phi;
run;

proc univariate data=ball noprint;
var phi;
by n;
output out=h3 mean=hsci;
run;

****Compute GPV****;
data call; set sim;
s_0=0.04;
eta0=((log(1.25))**2 + 0.02)/s_0;
xi=0;
xi2=0;
do j=1 to &ngp;
z=normal(0);
z12=normal(0);
ut1=2*rangam(0,nu/2);
ur1=2*rangam(0,(nu-1)/2);
uwt=2*rangam(0,nu/2);
sig2wt=sswr/uwt;
sigsqt1=sst1/ut1;
lambdasq1=ssront1/ur1;
```

```
beta1=b1-z12*sqrt(lambdasq1/sst1);
sigi1=sigsqt1*(beta1-1)**2 + lambdasq1;
sigsqr1=lambdasq1+beta1*beta1*sigsqt1;
z12=normal(0);
ut2=2*rangam(0,nu/2);
ur2=2*rangam(0,(nu-1)/2);
uwr=2*rangam(0,nu/2);
sig2wr=sswr/uwr;
sigsqt2=sst2/ut2;
lambdasq2=ssront2/ur2;
beta2=b2-z12*sqrt(lambdasq2/sst2);
sigi2=sigsqt2*(beta2-1)**2 + lambdasq2;
sigsqr2=lambdasq2+beta2*beta2*sigsqt1;
delta=d-z*sqrt((sigi1+sigi2)/n)/2;
gpvmp3=(delta*delta+(sigsqt1+sig2wt/2+sigsqt2-sigsqr1-sigsqr2-
sig2wr/2)/2)/max((sigsqr1+sigsqr2+sig2wr/2)/2,s_0);
xi=xi+(gpvmp3 gt eta0);
xi2=xi2+(gpvmp3 gt mp3);
end;
p=xi/&ngp;
c=xi2/&ngp;
psi=(p lt 0.05);
cover=(c gt 0.05);
run;

proc univariate data=call noprint;
var psi cover;
output out=gpv mean=gpv5 cover95;
by n;
run;
*merge results;
data tests; merge h3 gpv;
length test $ 5;
by n;
delta=&del;
r=&r;
sigma_bt=&sbt;
sigma_br=&sbr;
sigma_wt=&swt;
sigma_wr=&swr;
test="&ds";
mp3=&mp3;
run;run;

proc print data=tests;
run;
```

```
proc append base=mi3.powerpd3 data=tests;
run;

%mend;

%simpbe(0.05,0.90,0.15,0.15,0.15,0.15,5000,10000,20297,8,9,11,12,p1a);
%simpbe(0.05,0.95,0.15,0.15,0.15,0.15,5000,10000,41651,8,9,11,12,p1b);
%simpbe(0.05,0.90,0.30,0.30,0.15,0.15,5000,10000,99487,10,11,14,15,p1c);
%simpbe(0.05,0.95,0.30,0.30,0.15,0.15,5000,10000,56821,10,11,14,15,p1d);
%simpbe(0.05,0.90,0.23,0.23,0.23,0.23,5000,10000,81847,10,11,14,15,p2a);
%simpbe(0.05,0.95,0.23,0.23,0.23,0.23,5000,10000,11597,10,11,14,15,p2b);
%simpbe(0.05,0.90,0.46,0.46,0.23,0.23,5000,10000,29059,10,11,14,15,p2c);
%simpbe(0.05,0.95,0.46,0.46,0.23,0.23,5000,10000,18593,10,11,14,15,p2d);
%simpbe(0.05,0.90,0.3,0.3,0.3,0.3,5000,10000,60737,10,11,13,14,p3a);
%simpbe(0.05,0.95,0.3,0.3,0.3,0.3,5000,10000,88069,10,11,13,14,p3b);
%simpbe(0.05,0.90,0.6,0.6,0.3,0.3,5000,10000,60077,10,11,14,15,p3c);
%simpbe(0.05,0.95,0.6,0.6,0.3,0.3,5000,10000,47629,10,11,14,15,p3d);
%simpbe(0.05,0.90,0.5,0.5,0.5,0.5,5000,10000,62299,10,11,14,15,p4a);
%simpbe(0.05,0.95,0.5,0.5,0.5,0.5,5000,10000,50503,10,11,14,15,p4b);
%simpbe(0.05,0.90,1.0,1.0,0.5,0.5,5000,10000,77081,10,11,14,15,p4c);
%simpbe(0.05,0.95,1.0,1.0,0.5,0.5,5000,10000,48673,10,11,14,15,p4d);

proc print;
title2;
run;

/*********************************************
SIMIBENB3A.SAS
This program does the test size simulations
for IBE/3S3P.
Chapter 4, Section 4.5, Table 6.
*********************************************/
options ps=64 ls=120 nodate nonumber;
libname mi3 'c:\My Documents\My SAS Files\V8';
title 'Simulation for IBE';
run;

%macro simibe(del,sd,swt,swr,nsim,ngp,seed,ds);

title2 "delta=&del, sigma_d=&sd sigma_wt=&swt sigma_wr=&swr";
*create simulation data sets;
data sim;
****Input model parameter values****;
sigd=&sd;
sigwt=&swt;
sigwr=&swr;
delta=&del;
```

```
****Calculate needed parameter values****;
sig2wt=sigwt*sigwt;
sig2wr=sigwr*sigwr;
sig2i=sigd*sigd+sig2wt+sig2wr/2;
sigi=sqrt(sig2i);
mi3=(delta*delta+sig2i-1.5*sig2wr)/max(sig2wr,0.04);
call symput('MI3',mi3);
****Compute simulated values****;
junk=int(100000*ranuni(&seed));
do n=8,12,16;
nu=3*(n-1);
k=1/sqrt(3*n);
do j=1 to &nsim;
z=normal(0);
i=2*rangam(0,nu/2);
w=2*rangam(0,nu/2);
d=delta+k*sigi*z;
sswr=sig2wr*w;
ssi=sig2i*i;
keep n j nu d ssi sswr k mi3;
output;
end;
end;
run;

****Compute HSCI****;
data ball; set sim;
s_0=0.04;
eta0=((log(1.25))**2 + 0.05)/s_0;
s2wr=sswr/nu;
e_d=d*d;
e_i=ssi/nu;
e_r=(-3.9948*(s2wr gt s_0)-1.5*(s2wr le s_0))*s2wr;
t=tinv(0.95,nu);
h_d=(abs(d)+t*k*sqrt(e_i))**2;
qu=cinv(0.95,nu);
ql=cinv(0.05,nu);
h_i=ssi/ql;
h_r=nu*e_r/qu;
u_d=(h_d-e_d)**2;
u_i=(h_i-e_i)**2;
u_r=(h_r-e_r)**2;
h_mi3=(e_d+e_i+e_r-eta0*s_0*(s2wr le s_0))+sqrt(u_d+u_i+u_r);
phi=(h_mi3 lt 0);
keep n j phi h_mi3;
run;
```

```
proc univariate data=ball noprint;
var phi;
by n;
output out=h3 mean=hsci;
run;
****Compute GPV****;
data call; set sim;
s_0=0.04;
eta0=((log(1.25))**2 + 0.05)/s_0;
xi=0;
do j=1 to &ngp;
z=normal(0);
i=2*rangam(0,nu/2);
w=2*rangam(0,nu/2);
sig2i=ssi/i;
delta=d-k*z*sqrt(sig2i);
sig2wr=sswr/w;
gpvmi3=(delta*delta+sig2i-1.5*sig2wr)/max(sig2wr,s_0);
xi=xi+(gpvmi3 gt eta0);
end;
p=xi/&ngp;
psi=(p lt 0.05);
run;

proc univariate data=call noprint;
var psi;
output out=gpv mean=gpv;
by n;
run;
*merge results;
data tests; merge h3 gpv;
length test $ 5;
by n;
delta=&del;
sigma_d=&sd;
sigma_wt=&swt;
sigma_wr=&swr;
test="&ds";
mi3=&mi3;
run;

proc print data=tests;
run;

proc append base=mi3.testsinb3a data=tests;
run;
```

```
%mend;

%simibe(0.315742,0.01,0.15,0.15,5000,10000,84431,i1a);
%simibe(0.299655,0.1,0.15,0.15,5000,10000,79087,i1b);
%simibe(0.278016,0.15,0.15,0.15,5000,10000,71089,i1c);
%simibe(0.315742,0.01,0.20,0.20,5000,10000,38693,i2a);
%simibe(0.299655,0.1,0.20,0.20,5000,10000,90617,i2b);
%simibe(0.278016,0.15,0.20,0.20,5000,10000,26627,i2c);
%simibe(0.363148,0.01,0.23,0.23,5000,10000,19157,i3a);
%simibe(0.349251,0.1,0.23,0.23,5000,10000,35393,i3b);
%simibe(0.330872,0.15,0.23,0.23,5000,10000,53507,i3c);
%simibe(0.4737450264747,0.01,0.3,0.3,5000,10000,56401,i4a);
%simibe(0.4631785294133,0.1,0.3,0.3,5000,10000,97931,i4b);
%simibe(0.4494823134557,0.15,0.3,0.3,5000,10000,90053,i4c);
%simibe(0.7896876142387,0.01,0.5,0.5,5000,10000,88211,i5a);
%simibe(0.7833942354153,0.1,0.5,0.5,5000,10000,31123,i5b);
%simibe(0.775375088639,0.15,0.5,0.5,5000,10000,36931,i5c);
%simibe(0.05,0.01,0.345966826,0.15,5000,10000,49787,i6a);
%simibe(0.05,0.1,0.331350335,0.15,5000,10000,50909,i6b);
%simibe(0.05,0.15,0.31191833,0.15,5000,10000,51973,i6c);
%simibe(0.05,0.01,0.370395794,0.20,5000,10000,53117,i7a);
%simibe(0.05,0.1,0.356781508,0.20,5000,10000,54277,i7b);
%simibe(0.05,0.15,0.338811223,0.20,5000,10000,55373,i7c);
%simibe(0.05,0.01,0.426938288,0.23,5000,10000,56479,i8a);
%simibe(0.05,0.1,0.415182251,0.23,5000,10000,57503,i8b);
%simibe(0.05,0.15,0.399845347,0.23,5000,10000,58601,i8c);
%simibe(0.05,0.01,0.558510833,0.3,5000,10000,59671,i9a);
%simibe(0.05,0.1,0.549576519,0.3,5000,10000,60887,i9b);
%simibe(0.05,0.15,0.538083962,0.3,5000,10000,62039,i9c);
%simibe(0.05,0.01,0.933330878,0.5,5000,10000,65449,i10a);
%simibe(0.05,0.1,0.928012138,0.5,5000,10000,66587,i10b);
%simibe(0.05,0.15,0.921252695,0.5,5000,10000,68711,i10c);

proc print data=mi3.testsinb3a;
title2;
run;

/*****************************************
SIMIBENB3B_P.SAS
This program does the power/sample size simulations
for IBE/3S3P.
Chapter 4, Section 4.6, Table 7.
*****************************************/
options ps=64 ls=120 nodate nonumber;
libname mi3 'c:\My Documents\My SAS Files\V8';
title 'Simulation for IBE/power';
run;
```

```
%macro simibe(del,sd,swt,swr,nsim,ngp,n1,n2,seed,ds);

title2 "delta=&del, sigma_d=&sd sigma_wt=&swt sigma_wr=&swr";
*create simulation data sets;
data sim;
****Input model parameter values****;
sigd=&sd;
sigwt=&swt;
sigwr=&swr;
delta=&del;
****Calculate needed parameter values****;
sig2wt=sigwt*sigwt;
sig2wr=sigwr*sigwr;
sig2i=sigd*sigd+sig2wt+sig2wr/2;
sigi=sqrt(sig2i);
mi3=(delta*delta+sig2i-1.5*sig2wr)/max(sig2wr,0.04);
call symput('MI3',mi3);
****Compute simulated values****;
junk=int(100000*ranuni(&seed));
do n=&n1 to &n2;
nu=3*(n-1);
k=1/sqrt(3*n);
do j=1 to &nsim;
z=normal(0);
i=2*rangam(0,nu/2);
w=2*rangam(0,nu/2);
d=delta+k*sigi*z;
sswr=sig2wr*w;
ssi=sig2i*i;
keep n j nu d ssi sswr k mi3;
output;
end;
end;
run;

****Compute HSCI****;
data ball; set sim;
s_0=0.04;
eta0=((log(1.25))**2 + 0.05)/s_0;
s2wr=sswr/nu;
e_d=d*d;
e_i=ssi/nu;
e_r=(-3.9948*(s2wr gt s_0)-1.5*(s2wr le s_0))*s2wr;
t=tinv(0.95,nu);
h_d=(abs(d)+t*k*sqrt(e_i))**2;
qu=cinv(0.95,nu);
```

```
ql=cinv(0.05,nu);
h_i=ssi/ql;
h_r=nu*e_r/qu;
u_d=(h_d-e_d)**2;
u_i=(h_i-e_i)**2;
u_r=(h_r-e_r)**2;
h_mi3=(e_d+e_i+e_r-eta0*s_0*(s2wr le s_0))+sqrt(u_d+u_i+u_r);
phi=(h_mi3 lt 0);
keep n j phi h_mi3;
run;

proc univariate data=ball noprint;
var phi;
by n;
output out=h3 mean=hsci;
run;
****Compute GPV****;
data call; set sim;
s_0=0.04;
eta0=((log(1.25))**2 + 0.05)/s_0;
xi=0;
do j=1 to &ngp;
z=normal(0);
i=2*rangam(0,nu/2);
w=2*rangam(0,nu/2);
sig2i=ssi/i;
delta=d-k*z*sqrt(sig2i);
sig2wr=sswr/w;
gpvmi3=(delta*delta+sig2i-1.5*sig2wr)/max(sig2wr,s_0);
xi=xi+(gpvmi3 gt eta0);
end;
p=xi/&ngp;
psi=(p lt 0.05);
run;

proc univariate data=call noprint;
var psi;
output out=gpv mean=gpv;
by n;
run;
*merge results;
data tests; merge h3 gpv;
length test $ 5;
by n;
delta=&del;
sigma_d=&sd;
sigma_wt=&swt;
```

```
sigma_wr=&swr;
test="&ds";
mi3=&mi3;
run;

proc print data=tests;
run;

proc append base=mi3.testsinb3b_p data=tests;
run;

%mend;

%simibe(0.05,0.01,0.15,0.15,5000,10000,4,5,75211,p1a);
%simibe(0.05,0.1,0.15,0.15,5000,10000,5,7,71719,p1b);
%simibe(0.05,0.15,0.15,0.15,5000,10000,8,10,67247,p1c);
%simibe(0.05,0.2,0.15,0.15,5000,10000,17,20,61673,p1d);
%simibe(0.05,0.01,0.20,0.20,5000,10000,6,8,70663,p2a);
%simibe(0.05,0.1,0.20,0.20,5000,10000,8,10,59369,p2b);
%simibe(0.05,0.15,0.20,0.20,5000,10000,14,16,55147,p2c);
%simibe(0.05,0.2,0.20,0.20,5000,10000,26,29,76213,p2d);
%simibe(0.05,0.01,0.23,0.23,5000,10000,9,11,86539,p3a);
%simibe(0.05,0.1,0.23,0.23,5000,10000,12,14,84299,p3b);
%simibe(0.05,0.15,0.23,0.23,5000,10000,16,19,52291,p3c);
%simibe(0.05,0.2,0.23,0.23,5000,10000,27,30,54577,p3d);
%simibe(0.05,0.01,0.3,0.3,5000,10000,11,13,52511,p4a);
%simibe(0.05,0.1,0.3,0.3,5000,10000,11,13,57943,p4b);
%simibe(0.05,0.15,0.3,0.3,5000,10000,14,16,70313,p4c);
%simibe(0.05,0.2,0.3,0.3,5000,10000,18,21,88801,p4d);
%simibe(0.05,0.01,0.5,0.5,5000,10000,10,12,82469,p5a);
%simibe(0.05,0.1,0.5,0.5,5000,10000,10,13,35771,p5b);
%simibe(0.05,0.15,0.5,0.5,5000,10000,11,14,17579,p5c);
%simibe(0.05,0.2,0.5,0.5,5000,10000,12,14,29231,p5d);

proc print data=mi3.testsinb3b_p;
title2;
run;

/*******************************************
SIMPBEND3.SAS
This program does the test size simulations
for PBE/3S3P.
Chapter 4, Section 4.7, Table 8.
*******************************************/
options ps=64 ls=120 nodate nonumber; *nomprint;
libname mi3 'c:\My Documents\My SAS Files\V8';
title 'Simulation for PBE';
```

```
run;

%macro simpbe(del,r,sbt,sbr,swt,swr,nsim,ngp,seed,ds);

title2 "delta=&del, rho=&r, sigma_bt=&sbt
sigma_br=&sbr sigma_wt=&swt sigma_wr=&swr";
run;
*create simulation data sets;
data sim;
****Input model parameter values****;
sigbt=&sbt;
sigbr=&sbr;
sigwt=&swt;
sigwr=&swr;
delta=&del;
rho=&r;
****Calculate needed parameter values****;
sig2br=sigbr*sigbr;
sig2wr=sigwr*sigwr;
sigsq1=sigbt*sigbt+sigwt*sigwt;
sigsq2=sig2br+sig2wr/2;
sig12=rho*sigbt*sigbr;
beta=sig12/sigsq1;
sigi=sqrt(sigsq1+sigsq2-2*sig12);
lambda1=sqrt(sigsq1);
lambdasq2=sigsq2-sig12*sig12/sigsq1;
lambda2=sqrt(lambdasq2);
mp3=(delta*delta+sigsq1-sig2br-sig2wr)/max(sig2br+sig2wr,0.04);
call symput('MP3',mp3);
****Compute simulated values****;
junk=int(100000*ranuni(&seed));
do n=8,12,16;
nu=3*(n-1);
k=1/sqrt(3*n);
do i=1 to &nsim;
z=normal(0);
u1=2*rangam(0,nu/2);
u2=2*rangam(0,(nu-1)/2);
ur=2*rangam(0,nu/2);
z12=normal(0);
d=delta+k*sigi*z;
sswr=sig2wr*ur;
ss1=sigsq1*u1;
ss2on1=lambdasq2*u2;
b=beta+z12*lambda2/sqrt(ss1);
ss2=b*b*ss1+ss2on1;
sig2i=(ss2on1+ss1*(1-b)*(1-b))/nu;
```

```
keep n i nu d sigsq1 ss1 ss2 ss2on1 sig2i sswr b k mp3;
output;
end;
end;
run;

****Compute HSCI****;
data ball; set sim;
s_0=0.04;
eta0=((log(1.25))**2 + 0.02)/s_0;
e_d=d*d;
t=tinv(0.95,nu);
h_d=(abs(d)+k*t*sqrt(sig2i))**2;
qu=cinv(0.95,nu);
ql=cinv(0.05,nu);
u_d=(h_d-e_d)**2;
u_1=(ss1*(1/ql - 1/nu))**2;
if ((ss2/nu+sswr/(2*nu)) gt s_0) then do;
u_2=(ss2*(1+eta0)*(1/qu - 1/nu))**2;
u_3=(0.5*sswr*(1+eta0)*(1/qu - 1/nu))**2;
lmp3=e_d+ss1/nu-(1+eta0)*(ss2/nu+sswr/(2*nu));
h_mp3=lmp3+sqrt(u_d+u_1+u_3+u_2);
end;
else do;
u_2=(ss2*(1/qu - 1/nu))**2;
u_3=(0.5*sswr*(1/qu - 1/nu))**2;
lmp3=e_d+ss1/nu-(ss2/nu+sswr/(2*nu))-s_0*eta0;
h_mp3=lmp3+sqrt(u_d+u_1+u_2+u_3);
end;
phi=(h_mp3 lt 0);
keep n i lmp3 h_mp3 phi;
run;

proc univariate data=ball noprint;
var phi;
by n;
output out=h3 mean=hsci;
run;

****Compute GPV****;
data call; set sim;
s_0=0.04;
eta0=((log(1.25))**2 + 0.02)/s_0;
xi=0;
xi2=0;
do j=1 to &ngp;
z=normal(0);
```

```
z12=normal(0);
u1=2*rangam(0,nu/2);
u2=2*rangam(0,(nu-1)/2);
ur=2*rangam(0,nu/2);
sig2wr=sswr/ur;
sigsq1=ss1/u1;
lambdasq2=ss2on1/u2;
beta=b-z12*sqrt(lambdasq2/ss1);
sigi=sqrt(sigsq1*(beta-1)**2 + lambdasq2);
delta=d-k*sigi*z;
sigsq2=lambdasq2+beta*beta*sigsq1;
gpvmp3=(delta*delta+sigsq1-sigsq2-sig2wr/2)/max(sigsq2+sig2wr/2,s_0);
xi=xi+(gpvmp3 gt eta0);
xi2=xi2+(gpvmp3 gt mp3);
end;
p=xi/&ngp;
c=xi2/&ngp;
psi=(p lt 0.05);
cover=(c gt 0.05);
run;

proc univariate data=call noprint;
var psi cover;
output out=gpv mean=gpv5 cover95;
by n;
run;
*merge results;
data tests; merge h3 gpv;
length test $ 5;
by n;
delta=&del;
r=&r;
sigma_bt=&sbt;
sigma_br=&sbr;
sigma_wt=&swt;
sigma_wr=&swr;
test="&ds";
mp3=&mp3;
run;

proc print data=tests;
run;

proc append base=mi3.testpnd3a data=tests;
run;

%mend;
```

```
%simpbe(0.295366392,0.8,0.2,0.173205,0.1,0.141412,5000,10000,45557,e1a);
%simpbe(0.295366392,0.9,0.2,0.173205,0.1,0.141412,5000,10000,32143,e1b);
%simpbe(0.295366392,0.8,0.173205,0.2,0.141421,0.1,5000,10000,14197,e1c);
%simpbe(0.295366392,0.9,0.173205,0.2,0.141421,0.1,5000,10000,33343,e1d);
%simpbe(0.590733,0.8,0.387298,0.4,0.223607,0.2,5000,10000,98321,e2a);
%simpbe(0.590733,0.9,0.387298,0.4,0.223607,0.2,5000,10000,54517,e2b);
%simpbe(0.590733,0.8,0.4,0.387298,0.2,0.223607,5000,10000,69497,e2c);
%simpbe(0.590733,0.9,0.4,0.387298,0.2,0.223607,5000,10000,80917,e2d);
%simpbe(0.886099,0.8,0.6,0.5,0.3,0.4472136,5000,10000,55061,e3a);
%simpbe(0.886099,0.9,0.6,0.5,0.3,0.4472136,5000,10000,72277,e3a);
%simpbe(0.886099,0.8,0.5,0.6,0.4472136,0.3,5000,10000,55061,e3c);
%simpbe(0.886099,0.9,0.5,0.6,0.4472136,0.3,5000,10000,72277,e3d);
%simpbe(0,0.8,0.282476626,0.1,0.1,0.1,5000,10000,75577,f1a);
%simpbe(0,0.9,0.282476626,0.1,0.1,0.1,5000,10000,64621,f1b);
%simpbe(0,0.8,0.356708993,0.2,0.1,0.1,5000,10000,79043,f1c);
%simpbe(0,0.9,0.356708993,0.2,0.1,0.1,5000,10000,86923,f1d);
%simpbe(0,0.8,0.423775989,0.2,0.2,0.2,5000,10000,46867,f2a);
%simpbe(0,0.9,0.423775989,0.2,0.2,0.2,5000,10000,19237,f2b);
%simpbe(0,0.8,0.713417986,0.4,0.2,0.2,5000,10000,31699,f2c);
%simpbe(0,0.9,0.713417986,0.4,0.2,0.2,5000,10000,61211,f2d);
%simpbe(0,0.8,0.635663984,0.3,0.3,0.3,5000,10000,85103,f3a);
%simpbe(0,0.9,0.635663984,0.3,0.3,0.3,5000,10000,65267,f3b);
%simpbe(0,0.8,1.070126979,0.6,0.3,0.3,5000,10000,55061,f3c);
%simpbe(0,0.9,1.070126979,0.6,0.3,0.3,5000,10000,44897,f3d);

proc print;
where 1.7 lt mp3 lt 1.8;
title2;
run;

/*****************************************
SIMPBEND3_P.SAS
This program does the power/sample size simulations
for PBE.
Chapter 4, Section 4.8, Table 9.
*****************************************/
options ps=64 ls=120 nodate nonumber; *nomprint;
libname mi3 'c:\My Documents\My SAS Files\V8';
title 'Simulation for PBE/Table 9';
run;

%macro simpbe(del,r,sbt,sbr,swt,swr,nsim,ngp,seed,n1,n2,n3,n4,ds);

title2 "delta=&del, rho=&r, sigma_bt=&sbt
sigma_br=&sbr sigma_wt=&swt sigma_wr=&swr";
run;
```

```
*create simulation data sets;
data sim;
****Input model parameter values****;
sigbt=&sbt;
sigbr=&sbr;
sigwt=&swt;
sigwr=&swr;
delta=&del;
rho=&r;
****Calculate needed parameter values****;
sig2br=sigbr*sigbr;
sig2wr=sigwr*sigwr;
sigsq1=sigbt*sigbt+sigwt*sigwt;
sigsq2=sig2br+sig2wr/2;
sig12=rho*sigbt*sigbr;
beta=sig12/sigsq1;
sigi=sqrt(sigsq1+sigsq2-2*sig12);
lambda1=sqrt(sigsq1);
lambdasq2=sigsq2-sig12*sig12/sigsq1;
lambda2=sqrt(lambdasq2);
mp3=(delta*delta+sigsq1-sig2br-sig2wr)/max(sig2br+sig2wr,0.04);
call symput('MP3',mp3);
****Compute simulated values****;
junk=int(100000*ranuni(&seed));
do n=&n1,&n2,&n3,&n4;
nu=3*(n-1);
k=1/sqrt(3*n);
do i=1 to &nsim;
z=normal(0);
u1=2*rangam(0,nu/2);
u2=2*rangam(0,(nu-1)/2);
ur=2*rangam(0,nu/2);
z12=normal(0);
d=delta+k*sigi*z;
sswr=sig2wr*ur;
ss1=sigsq1*u1;
ss2on1=lambdasq2*u2;
b=beta+z12*lambda2/sqrt(ss1);
ss2=b*b*ss1+ss2on1;
sig2i=(ss2on1+ss1*(1-b)*(1-b))/nu;
keep n i nu d sigsq1 ss1 ss2 ss2on1 sig2i sswr b k mp3;
output;
end;
end;
run;

****Compute HSCI****;
```

```
data ball; set sim;
s_0=0.04;
eta0=((log(1.25))**2 + 0.02)/s_0;
e_d=d*d;
t=tinv(0.95,nu);
h_d=(abs(d)+k*t*sqrt(sig2i))**2;
qu=cinv(0.95,nu);
ql=cinv(0.05,nu);
u_d=(h_d-e_d)**2;
u_1=(ss1*(1/ql - 1/nu))**2;
if ((ss2/nu+sswr/(2*nu)) gt s_0) then do;
u_2=(ss2*(1+eta0)*(1/qu - 1/nu))**2;
u_3=(0.5*sswr*(1+eta0)*(1/qu - 1/nu))**2;
lmp3=e_d+ss1/nu-(1+eta0)*(ss2/nu+sswr/(2*nu));
h_mp3=lmp3+sqrt(u_d+u_1+u_3+u_2);
end;
else do;
u_2=(ss2*(1/qu - 1/nu))**2;
u_3=(0.5*sswr*(1/qu - 1/nu))**2;
lmp3=e_d+ss1/nu-(ss2/nu+sswr/(2*nu))-s_0*eta0;
h_mp3=lmp3+sqrt(u_d+u_1+u_2+u_3);
end;
phi=(h_mp3 lt 0);
keep n i lmp3 h_mp3 phi;
run;

proc univariate data=ball noprint;
var phi;
by n;
output out=h3 mean=hsci;
run;

****Compute GPV****;
data call; set sim;
s_0=0.04;
eta0=((log(1.25))**2 + 0.02)/s_0;
xi=0;
xi2=0;
do j=1 to &ngp;
z=normal(0);
z12=normal(0);
u1=2*rangam(0,nu/2);
u2=2*rangam(0,(nu-1)/2);
ur=2*rangam(0,nu/2);
sig2wr=sswr/ur;
sigsq1=ss1/u1;
lambdasq2=ss2on1/u2;
```

```
beta=b-z12*sqrt(lambdasq2/ss1);
sigi=sqrt(sigsq1*(beta-1)**2 + lambdasq2);
delta=d-k*sigi*z;
sigsq2=lambdasq2+beta*beta*sigsq1;
gpvmp3=(delta*delta+sigsq1-sigsq2-sig2wr/2)/max(sigsq2+sig2wr/2,s_0);
xi=xi+(gpvmp3 gt eta0);
xi2=xi2+(gpvmp3 gt mp3);
end;
p=xi/&ngp;
c=xi2/&ngp;
psi=(p lt 0.05);
cover=(c gt 0.05);
run;

proc univariate data=call noprint;
var psi cover;
output out=gpv mean=gpv5 cover95;
by n;
run;
*merge results;
data tests; merge h3 gpv;
length test $ 5;
by n;
delta=&del;
r=&r;
sigma_bt=&sbt;
sigma_br=&sbr;
sigma_wt=&swt;
sigma_wr=&swr;
test="&ds";
mp3=&mp3;
run;run;

proc print data=tests;
run;

proc append base=mi3.pwerpnd3a data=tests;
run;

%mend;

%simpbe(0.05,0.90,0.15,0.15,0.15,0.15,5000,10000,20297,4,5,6,7,p1a);
%simpbe(0.05,0.95,0.15,0.15,0.15,0.15,5000,10000,41659,4,5,6,7,p1b);
%simpbe(0.05,0.90,0.30,0.30,0.15,0.15,5000,10000,99487,4,5,7,8,p1c);
%simpbe(0.05,0.95,0.30,0.30,0.15,0.15,5000,10000,56821,4,5,7,8,p1d);
%simpbe(0.05,0.90,0.23,0.23,0.23,0.23,5000,10000,81847,5,6,8,9,p2a);
%simpbe(0.05,0.95,0.23,0.23,0.23,0.23,5000,10000,11597,5,6,8,9,p2b);
```

```
%simpbe(0.05,0.90,0.46,0.46,0.23,0.23,5000,10000,29059,4,5,7,8,p2c);
%simpbe(0.05,0.95,0.46,0.46,0.23,0.23,5000,10000,18593,4,5,7,8,p2d);
%simpbe(0.05,0.90,0.3,0.3,0.3,0.3,5000,10000,60737,6,7,8,9,p3a);
%simpbe(0.05,0.95,0.3,0.3,0.3,0.3,5000,10000,88069,6,7,8,9,p3b);
%simpbe(0.05,0.90,0.6,0.6,0.3,0.3,5000,10000,60077,4,5,7,8,p3c);
%simpbe(0.05,0.95,0.6,0.6,0.3,0.3,5000,10000,47629,4,5,7,8,p3d);
%simpbe(0.05,0.90,0.5,0.5,0.5,0.5,5000,10000,62299,5,6,7,8,p4a);
%simpbe(0.05,0.95,0.5,0.5,0.5,0.5,5000,10000,50503,5,6,7,8,p4b);
%simpbe(0.05,0.90,1.0,1.0,0.5,0.5,5000,10000,77081,4,5,7,8,p4c);
%simpbe(0.05,0.95,1.0,1.0,0.5,0.5,5000,10000,48673,4,5,7,8,p4d);

proc print;
title2;
run;

/****************************************
EX2S2P_PBE.SAS
This program does FDA (HSCI) and GPV procedures for PBE
for 2x2 data (Chow and Liu, 2000, Table 3.6.1, pg. 73).
This is the example in Chapter 5, Section 3.
****************************************/
options nodate nonumber ls=120 ps=64;
title;
*Enter, transform data and define model parameters;
data chow_liu;
input seq subject period trt auc;
lauc=log(auc);
gamma1=(trt=1)*((seq=1)-(seq=2));
gamma2=(trt=2)*((seq=1)-(seq=2));
cards;
1 1 1 2 74.675
1 1 2 1 73.675
1 4 1 2 96.4
1 4 2 1 93.25
1 5 1 2 101.95
1 5 2 1 102.125
1 6 1 2 79.05
1 6 2 1 69.45
1 11 1 2 79.05
1 11 2 1 69.025
1 12 1 2 85.95
1 12 2 1 68.7
1 15 1 2 69.725
1 15 2 1 59.425
1 16 1 2 86.275
1 16 2 1 76.125
1 19 1 2 112.675
```

```
1 19 2 1 114.875
1 20 1 2 99.525
1 20 2 1 116.25
1 23 1 2 89.425
1 23 2 1 64.175
1 24 1 2 55.175
1 24 2 1 74.575
2 2 1 1 74.825
2 2 2 2 37.35
2 3 1 1 86.875
2 3 2 2 51.925
2 7 1 1 81.675
2 7 2 2 72.175
2 8 1 1 92.7
2 8 2 2 77.5
2 9 1 1 50.45
2 9 2 2 71.875
2 10 1 1 66.125
2 10 2 2 94.025
2 13 1 1 122.45
2 13 2 2 124.975
2 14 1 1 99.075
2 14 2 2 85.225
2 17 1 1 86.35
2 17 2 2 95.925
2 18 1 1 49.925
2 18 2 2 67.1
2 21 1 1 42.7
2 21 2 2 59.425
2 22 1 1 91.725
2 22 2 2 114.05
;;;;
run;

proc transpose data=chow_liu out=tcl prefix=lc;
var lauc;
by seq subject;
id trt;
run;
*compute statistics;
data tcl; set tcl;
xp=lc1*lc2;
d=lc1-lc2;
run;

proc univariate noprint;
var lc1 lc2 d;
```

```
output out=st mean=i1-i3 css=s1-s3;
by seq;
run;
*compute cross product;
data xp; merge tc1(in=a) st;
by seq;
if a;
if first.seq then sxp=0;
xp=(lc1-i1)*(lc2-i2);
sxp=sxp+xp;
retain sxp;
if last.seq then output;
run;
*finish computing summary statistics;
proc univariate noprint;
var i3 s1-s3 sxp;
output out=st2 sum=i sst ssr ssi ss12;
run;

data st2; set st2;
nu=22;
mut=sst/nu;
mur=ssr/nu;
mi=ssi/nu;
ssront=ssr-ss12*ss12/sst;
b=ss12/sst;
d=i/2;
run;
*compute FDA (HSCI) test;
data ball; set st2;
s_0=0.04;
eta0=((log(1.25))**2 + 0.02)/s_0;
e_d=d*d;
e_1=mut;
e_3=-(1+eta0)*mur;
t=tinv(0.95,nu);
h_d=(abs(d)+t*sqrt(mi)/24)**2;
qu=cinv(0.95,nu);
ql=cinv(0.05,nu);
h_1=nu*e_1/ql;
h_3=nu*e_3/qu;
u_d=(h_d-e_d)**2;
u_1=(h_1-e_1)**2;
u_3=(h_3-e_3)**2;
mp3=(e_d+e_1-mur)/mur;
lmp3=e_d+e_1+e_3;
h_mp3=lmp3+sqrt(u_d+u_1+u_3);
```

```
run;

proc print data=ball;
run;
*compute GPV;
data call; set st2;
k=sqrt(1/24);
s_0=0.04;
eta0=((log(1.25))**2 + 0.02)/s_0;
do j=1 to 50000;
z=normal(0);
z12=normal(0);
ut=2*rangam(0,nu/2);
ur=2*rangam(0,(nu-1)/2);
sigsq1=sst/ut;
lambdasq2=ssront/ur;
beta=b-z12*sqrt(lambdasq2/sst);
sigi=sqrt(sigsq1*(beta-1)**2 + lambdasq2);
delta=d-k*sigi*z;
sigsq2=lambdasq2+beta*beta*sigsq1;
gpvmp3=(delta*delta+sigsq1-sigsq2)/max(sigsq2,s_0);
psi=(gpvmp3 gt eta0);
output;
end;
run;

proc univariate;
var psi gpvmp3;
run;

/*****************************************
SIMPBEP2.SAS
This program does the test size simulations
for PBE/2x2.
Chapter 5, Section 4, Table 1.
*****************************************/
options ps=64 ls=120 nodate nonumber; *nomprint;
libname mi3 'c:\My Documents\My SAS Files\V8';
title 'Simulation for PBE';
run;

%macro simpbe(del,r,sbt,sbr,swt,swr,nsim,ngp,seed,ds);

title2 "delta=&del, rho=&r, sigma_bt=&sbt
sigma_br=&sbr sigma_wt=&swt sigma_wr=&swr";
run;
*create simulation data sets;
```

```
data sim;
****Input model parameter values****;
sigbt=&sbt;
sigbr=&sbr;
sigwt=&swt;
sigwr=&swr;
delta=&del;
rho=&r;
****Calculate needed parameter values****;
sig2br=sigbr*sigbr;
sig2wr=sigwr*sigwr;
sigsq1=sigbt*sigbt+sigwt*sigwt;
sigsq2=sig2br+sig2wr;
sig12=rho*sigbt*sigbr;
beta=sig12/sigsq1;
sigi=sqrt(sigsq1+sigsq2-2*sig12);
lambda1=sqrt(sigsq1);
lambdasq2=sigsq2-beta*beta*sigsq1;
lambda2=sqrt(lambdasq2);
mp3=(delta*delta+sigsq1-sigsq2)/max(sig2br+sig2wr,0.04);
call symput('MP3',mp3);
****Compute simulated values****;
junk=int(100000*ranuni(&seed));
do n=8,12,16;
nu=2*(n-1);
k=1/sqrt(2*n);
do i=1 to &nsim;
z=normal(0);
u1=2*rangam(0,nu/2);
u2=2*rangam(0,(nu-1)/2);
z12=normal(0);
d=delta+k*sigi*z;
ss1=sigsq1*u1;
ss2on1=lambdasq2*u2;
b=beta+z12*lambda2/sqrt(ss1);
ss2=b*b*ss1+ss2on1;
sig2i=(ss2on1+ss1*(1-b)*(1-b))/nu;
keep n i nu d sigsq1 ss1 ss2 ss2on1 sig2i b k mp3;
output;
end;
end;
run;

****Compute HSCI****;
data ball; set sim;
s_0=0.04;
eta0=((log(1.25))**2 + 0.02)/s_0;
```

```
e_d=d*d;
t=tinv(0.95,nu);
h_d=(abs(d)+k*t*sqrt(sig2i))**2;
qu=cinv(0.95,nu);
ql=cinv(0.05,nu);
u_d=(h_d-e_d)**2;
u_1=(ss1*(1/ql - 1/nu))**2;
if (ss2/nu gt s_0) then do;
u_2=(ss2*(1+eta0)*(1/qu - 1/nu))**2;
lmp3=e_d+ss1/nu-(1+eta0)*(ss2/nu);
h_mp3=lmp3+sqrt(u_d+u_1+u_2);
end;
else do;
u_2=(ss2*(1/qu - 1/nu))**2;
lmp3=e_d+ss1/nu-(ss2/nu)-s_0*eta0;
h_mp3=lmp3+sqrt(u_d+u_1+u_2);
end;
phi=(h_mp3 lt 0);
keep n i lmp3 h_mp3 phi;
run;

proc univariate data=ball noprint;
var phi;
by n;
output out=h3 mean=hsci;
run;

****Compute GPV****;
data call; set sim;
s_0=0.04;
eta0=((log(1.25))**2 + 0.02)/s_0;
xi=0;
xi2=0;
do j=1 to &ngp;
z=normal(0);
z12=normal(0);
u1=2*rangam(0,nu/2);
u2=2*rangam(0,(nu-1)/2);
sigsq1=ss1/u1;
lambdasq2=ss2on1/u2;
beta=b-z12*sqrt(lambdasq2/ss1);
sigi=sqrt(sigsq1*(beta-1)**2 + lambdasq2);
delta=d-k*sigi*z;
sigsq2=lambdasq2+beta*beta*sigsq1;
gpvmp3=(delta*delta+sigsq1-sigsq2)/max(sigsq2,s_0);
xi=xi+(gpvmp3 gt eta0);
xi2=xi2+(gpvmp3 gt mp3);
```

```
end;
p=xi/&ngp;
c=xi2/&ngp;
psi=(p lt 0.05);
cover=(c gt 0.05);
run;

proc univariate data=call noprint;
var psi cover;
output out=gpv mean=gpv5 cover95;
by n;
run;
*merge results;
data tests; merge h3 gpv;
length test $ 5;
by n;
delta=&del;
r=&r;
sigma_bt=&sbt;
sigma_br=&sbr;
sigma_wt=&swt;
sigma_wr=&swr;
test="&ds";
mp3=&mp3;
run;run;

proc print data=tests;
run;

proc append base=mi3.testp2 data=tests;
run;

%mend;

%simpbe(0.2641837,0.8,0.1,0.1,0.1,0.1,5000,10000,71363,d1a);
%simpbe(0.2641837,0.9,0.1,0.1,0.1,0.1,5000,10000,88241,d1b);
%simpbe(0.295366,0.8,0.2,0.2,0.1,0.1,5000,10000,54559,d1c);
%simpbe(0.295366,0.9,0.2,0.2,0.1,0.1,5000,10000,32917,d1d);
%simpbe(0.373612217,0.8,0.2,0.2,0.2,0.2,5000,10000,46867,d2a);
%simpbe(0.373612217,0.9,0.2,0.2,0.2,0.2,5000,10000,19237,d2b);
%simpbe(0.590733,0.8,0.4,0.4,0.2,0.2,5000,10000,31699,d2c);
%simpbe(0.590733,0.9,0.4,0.4,0.2,0.2,5000,10000,61211,d2d);
%simpbe(0.560418,0.8,0.3,0.3,0.3,0.3,5000,10000,85103,d3a);
%simpbe(0.560418,0.9,0.3,0.3,0.3,0.3,5000,10000,65267,d3b);
%simpbe(0.886099,0.8,0.6,0.6,0.3,0.3,5000,10000,55061,d3c);
%simpbe(0.886099,0.9,0.6,0.6,0.3,0.3,5000,10000,72277,d3d);
%simpbe(0.295366392,0.8,0.2,0.173205,0.1,0.141412,5000,10000,45557,e1a);
```

```
%simpbe(0.295366392,0.9,0.2,0.173205,0.1,0.141412,5000,10000,32143,e1b);
%simpbe(0.295366392,0.8,0.173205,0.2,0.141421,0.1,5000,10000,14197,e1c);
%simpbe(0.295366392,0.9,0.173205,0.2,0.141421,0.1,5000,10000,33343,e1d);
%simpbe(0.590733,0.8,0.387298,0.4,0.223607,0.2,5000,10000,98321,e2a);
%simpbe(0.590733,0.9,0.387298,0.4,0.223607,0.2,5000,10000,54517,e2b);
%simpbe(0.590733,0.8,0.4,0.387298,0.2,0.223607,5000,10000,69497,e2c);
%simpbe(0.590733,0.9,0.4,0.387298,0.2,0.223607,5000,10000,80917,e2d);
%simpbe(0.886099,0.8,0.6,0.5,0.3,0.4472136,5000,10000,55061,e3a);
%simpbe(0.886099,0.9,0.6,0.5,0.3,0.4472136,5000,10000,72277,e3a);
%simpbe(0.886099,0.8,0.5,0.6,0.4472136,0.3,5000,10000,10037,e3c);
%simpbe(0.886099,0.9,0.5,0.6,0.4472136,0.3,5000,10000,80173,e3d);
%simpbe(0,0.8,0.282476626,0.1,0.1,0.1,5000,10000,75577,f1a);
%simpbe(0,0.9,0.282476626,0.1,0.1,0.1,5000,10000,64621,f1b);
%simpbe(0,0.8,0.356708993,0.2,0.1,0.1,5000,10000,79043,f1c);
%simpbe(0,0.9,0.356708993,0.2,0.1,0.1,5000,10000,86923,f1d);
%simpbe(0,0.8,0.423775989,0.2,0.2,0.2,5000,10000,46867,f2a);
%simpbe(0,0.9,0.423775989,0.2,0.2,0.2,5000,10000,19237,f2b);
%simpbe(0,0.8,0.713417986,0.4,0.2,0.2,5000,10000,31699,f2c);
%simpbe(0,0.9,0.713417986,0.4,0.2,0.2,5000,10000,61211,f2d);
%simpbe(0,0.8,0.635663984,0.3,0.3,0.3,5000,10000,85103,f3a);
%simpbe(0,0.9,0.635663984,0.3,0.3,0.3,5000,10000,65267,f3b);
%simpbe(0,0.8,1.070126979,0.6,0.3,0.3,5000,10000,55061,f3c);
%simpbe(0,0.9,1.070126979,0.6,0.3,0.3,5000,10000,44897,f3d);

proc print;
title2;
run;

/*******************************************
SIMPBEP2_P.SAS
This program does the power/sample size simulations
for PBE/2x2.
Chapter 5, Section 4, Table 2.
*******************************************/
options ps=64 ls=120 nodate nonumber nomprint;
libname mi3 'c:\My Documents\My SAS Files\V8';
title 'Simulation for PBE';
run;

%macro simpbe(del,r,sbt,sbr,swt,swr,nsim,ngp,seed,n1,n2,n3,n4,ds);

title2 "delta=&del, rho=&r, sigma_bt=&sbt sigma_br=&sbr
sigma_wt=&swt sigma_wr=&swr";
run;
*create simulation data sets;
data sim;
****Input model parameter values****;
```

```
sigbt=&sbt;
sigbr=&sbr;
sigwt=&swt;
sigwr=&swr;
delta=&del;
rho=&r;
****Calculate needed parameter values****;
sig2br=sigbr*sigbr;
sig2wr=sigwr*sigwr;
sigsq1=sigbt*sigbt+sigwt*sigwt;
sigsq2=sig2br+sig2wr;
sig12=rho*sigbt*sigbr;
beta=sig12/sigsq1;
sigi=sqrt(sigsq1+sigsq2-2*sig12);
lambda1=sqrt(sigsq1);
lambdasq2=sigsq2-beta*beta*sigsq1;
lambda2=sqrt(lambdasq2);
mp3=(delta*delta+sigsq1-sigsq2)/max(sig2br+sig2wr,0.04);
call symput('MP3',mp3);
****Compute simulated values****;
junk=int(100000*ranuni(&seed));
do n=&n1,&n2,&n3,&n4;
*do n=5 to 20;
nu=2*(n-1);
k=1/sqrt(2*n);
do i=1 to &nsim;
z=normal(0);
u1=2*rangam(0,nu/2);
u2=2*rangam(0,(nu-1)/2);
z12=normal(0);
d=delta+k*sigi*z;
ss1=sigsq1*u1;
ss2on1=lambdasq2*u2;
b=beta+z12*lambda2/sqrt(ss1);
ss2=b*b*ss1+ss2on1;
sig2i=(ss2on1+ss1*(1-b)*(1-b))/nu;
keep n i nu d sigsq1 ss1 ss2 ss2on1 sig2i b k mp3;
output;
end;
end;
run;

****Compute HSCI****;
data ball; set sim;
s_0=0.04;
eta0=((log(1.25))**2 + 0.02)/s_0;
e_d=d*d;
```

```
t=tinv(0.95,nu);
h_d=(abs(d)+k*t*sqrt(sig2i))**2;
qu=cinv(0.95,nu);
ql=cinv(0.05,nu);
u_d=(h_d-e_d)**2;
u_1=(ss1*(1/ql - 1/nu))**2;
if ((ss2/nu) gt s_0) then do;
u_2=(ss2*(1+eta0)*(1/qu - 1/nu))**2;
lmp3=e_d+ss1/nu-(1+eta0)*(ss2/nu);
h_mp3=lmp3+sqrt(u_d+u_1+u_2);
end;
else do;
u_2=(ss2*(1/qu - 1/nu))**2;
lmp3=e_d+ss1/nu-(ss2/nu)-s_0*eta0;
h_mp3=lmp3+sqrt(u_d+u_1+u_2);
end;
phi=(h_mp3 lt 0);
keep n i lmp3 h_mp3 phi;
run;

proc univariate data=ball noprint;
var phi;
by n;
output out=h3 mean=hsci;
run;

****Compute GPV****;
data call; set sim;
s_0=0.04;
eta0=((log(1.25))**2 + 0.02)/s_0;
xi=0;
xi2=0;
do j=1 to &ngp;
z=normal(0);
z12=normal(0);
u1=2*rangam(0,nu/2);
u2=2*rangam(0,(nu-1)/2);
sigsq1=ss1/u1;
lambdasq2=ss2on1/u2;
beta=b-z12*sqrt(lambdasq2/ss1);
sigi=sqrt(sigsq1*(beta-1)**2 + lambdasq2);
delta=d-k*sigi*z;
sigsq2=lambdasq2+beta*beta*sigsq1;
gpvmp3=(delta*delta+sigsq1-sigsq2)/max(sigsq2,s_0);
xi=xi+(gpvmp3 gt eta0);
xi2=xi2+(gpvmp3 gt mp3);
end;
```

```
p=xi/&ngp;
c=xi2/&ngp;
psi=(p lt 0.05);
cover=(c gt 0.05);
run;

proc univariate data=call noprint;
var psi cover;
output out=gpv mean=gpv5 cover95;
by n;
run;
*merge results;
data tests; merge h3 gpv;
length test $ 5;
by n;
delta=&del;
r=&r;
sigma_bt=&sbt;
sigma_br=&sbr;
sigma_wt=&swt;
sigma_wr=&swr;
test="&ds";
mp3=&mp3;
run;run;

proc print data=tests;
run;

proc append base=mi3.testp2_p data=tests;
run;

%mend;

%simpbe(0.05,0.90,0.15,0.15,0.15,0.15,5000,10000,20297,7,8,10,11,p1a);
%simpbe(0.05,0.95,0.15,0.15,0.15,0.15,5000,10000,41651,8,9,11,12,p1b);
%simpbe(0.05,0.90,0.30,0.30,0.15,0.15,5000,10000,99487,7,8,12,13,p1c);
%simpbe(0.05,0.95,0.30,0.30,0.15,0.15,5000,10000,56821,6,7,11,12,p1d);
%simpbe(0.05,0.90,0.23,0.23,0.23,0.23,5000,10000,81847,11,12,14,15,p2a);
%simpbe(0.05,0.95,0.23,0.23,0.23,0.23,5000,10000,11597,11,12,14,15,p2b);
%simpbe(0.05,0.90,0.46,0.46,0.23,0.23,5000,10000,29059,7,8,12,13,p2c);
%simpbe(0.05,0.95,0.46,0.46,0.23,0.23,5000,10000,18593,7,8,12,13,p2d);
%simpbe(0.05,0.90,0.3,0.3,0.3,0.3,5000,10000,60737,11,12,14,15,p3a);
%simpbe(0.05,0.95,0.3,0.3,0.3,0.3,5000,10000,88069,11,12,14,15,p3b);
%simpbe(0.05,0.90,0.6,0.6,0.3,0.3,5000,10000,60077,7,8,12,13,p3c);
%simpbe(0.05,0.95,0.6,0.6,0.3,0.3,5000,10000,47629,7,8,12,13,p3d);
%simpbe(0.05,0.90,0.5,0.5,0.5,0.5,5000,10000,62299,11,12,14,15,p4a);
%simpbe(0.05,0.95,0.5,0.5,0.5,0.5,5000,10000,50503,11,12,14,15,p4b);
```

```
%simpbe(0.05,0.90,1.0,1.0,0.5,0.5,5000,10000,77081,7,8,12,13,p4c);
%simpbe(0.05,0.95,1.0,1.0,0.5,0.5,5000,10000,48673,7,8,12,13,p4d);

proc print;
*where 1.7 lt mp3 lt 1.8;
title2;
run;

/*****************************************
SIMIBEC.SAS
This program does the interval coverage simulations
for IBE/2S4P.
Chapter 7, Section 4.1, Table 2.
*****************************************/
options ps=64 ls=120 nodate nonumber mprint;
libname mi3 'c:\My Documents\My SAS Files\V8';
title 'Simulation for IBE intervals';
run;

%macro simibe(del,sd,swt,swr,nsim,ngp,n1,n2,seed,ds);

title2 "delta=&del, sigma_d=&sd sigma_wt=&swt sigma_wr=&swr";
*create simulation data sets;
data sim;
****Input model parameter values****;
sigd=&sd;
sigwt=&swt;
sigwr=&swr;
delta=&del;
****Calculate needed parameter values****;
sig2wt=sigwt*sigwt;
sig2wr=sigwr*sigwr;
sig2i=sigd*sigd+sig2wt/2+sig2wr/2;
sigi=sqrt(sig2i);
mi3=(delta*delta+sig2i+0.5*sig2wt-1.5*sig2wr)/max(sig2wr,0.04);
call symput('MI3',mi3);
****Compute simulated values****;
junk=int(100000*ranuni(&seed));
do n=8,12,16,20;
*do n=5 to 35 by 3;
nu=2*(n-1);
k=1/sqrt(2*n);
do i=1 to &nsim;
z=normal(0);
u=2*rangam(0,nu/2);
v=2*rangam(0,nu/2);
w=2*rangam(0,nu/2);
```

```
d=delta+k*sigi*z;
sswt=sig2wt*u;
sswr=sig2wr*v;
ssi=sig2i*w;
keep n i nu d ssi sswt sswr k mi3;
output;
end;
end;
run;
****Compute Ting-Burdick****;
data ting; set sim;
s_0=0.04;
eta0=((log(1.25))**2 + 0.05)/s_0;
sn=2*n;
m_i=ssi/nu;
m_t=sswt/nu;
m_r=sswr/nu;
ssdrug=sn*d*d;
****This stuff is direct from Burdick simulation program****;
lambhat=ssdrug/(2*M_i);
nstar=int((1+2*lambhat)**2/(1+4*lambhat))+1;
f15=cinv(0.05,nu)/nu;
fu2=cinv(0.95,nu)/nu;
f16=finv(0.05,nstar,nu);
f17=finv(0.05,nu,nu);
hd=nstar/cinv(0.05,nstar)-1;
hi=1/f15-1;
ht=hi;
gr=1-1/fu2;
hdr=((1-f16)**2-(hd*f16)**2-gr**2)/f16;
hir=((1-f17)**2-(hi*f17)**2-gr**2)/f17;
htr=hir;
k1=ssdrug/(sn)+(1-1/(sn))*m_i+.5*m_t;
k2=m_r;
rhohtibe=k1/k2;
k6=(hd*ssdrug/(sn))**2+(hi*(1-1/(sn))*m_i)**2+(ht*.5*m_t)**2;
k7=hdr/(sn)*ssdrug*m_r+ hir*(1-1/(sn))*m_i*m_r+htr*.5*m_t*m_r;
k8=(gr*m_r)**2;
if (m_r gt s_0) then do;
vuiber=(2+k7/(k1*k2))**2-4*(1-k8/k2**2)*(1-k6/k1**2);
umlsi=rhohtibe*((2+k7/(k1*k2)+sqrt(vuiber))/(2* (1-k8/k2**2)))-1.5;
end;
else do;
vustar=k6+(1.5)**2*k8+1.5*k7;
umlsi=(k1-1.5*k2+sqrt(vustar))/s_0;
end;
chi=(umlsi gt mi3);
```

```
run;

proc univariate data=ting noprint;
var chi;
output out=mls mean=mls;
by n;
run;
****Compute GCI****;
data call; set sim;
s_0=0.04;
eta0=((log(1.25))**2 + 0.05)/s_0;
xi=0;
xi2=0;
do j=1 to &ngp;
z=normal(0);
u=2*rangam(0,nu/2);
v=2*rangam(0,nu/2);
w=2*rangam(0,nu/2);
sig2i=ssi/w;
delta=d-k*z*sqrt(sig2i);
sig2wt=sswt/u;
sig2wr=sswr/v;
gpvmi3=(delta*delta+sig2i+0.5*sig2wt-1.5*sig2wr)/max(sig2wr,s_0);
del2=1.5*sig2wr+mi3*max(sig2wr,s_0)-sig2i-sig2wt/2;
psi2=(probnorm((d+sqrt(abs(del2)))/(k*sqrt(sig2i)))
-probnorm((d-sqrt(abs(del2)))/(k*sqrt(sig2i))))*(del2>0);
xi=xi+(gpvmi3 gt mi3);
xi2=xi2+psi2;
end;
c=xi/&ngp;
c2=xi2/&ngp;
cover=(c gt 0.05);
cover2=(c2 lt 0.95);
run;

proc univariate data=call noprint;
var cover cover2;
output out=gpv mean=cover1 cover2;
by n;
run;
*merge results;
data tests; merge /*h3*/ mls gpv;
length test $ 5;
by n;
delta=&del;
sigma_d=&sd;
sigma_wt=&swt;
```

```
sigma_wr=&swr;
test="&ds";
mi3=&mi3;
run;

proc print data=tests;
run;

proc append base=mi3.coveria data=tests;
run;

%mend;

%simibe(0.05,0.01,0.15,0.15,5000,10000,5,6,65479,p1a);
%simibe(0.05,0.1,0.15,0.15,5000,10000,6,9,47041,p1b);
%simibe(0.05,0.2,0.15,0.15,5000,10000,10,13,45823,p1c);
%simibe(0.05,0.3,0.15,0.15,5000,10000,10,13,45823,p1d);
%simibe(0.05,0.01,0.20,0.20,5000,10000,8,10,54001,p2a);
%simibe(0.05,0.1,0.20,0.20,5000,10000,11,14,28031,p2b);
%simibe(0.05,0.2,0.20,0.20,5000,10000,17,21,18503,p2c);
%simibe(0.05,0.3,0.20,0.20,5000,10000,11,14,54577,p2d);
%simibe(0.05,0.01,0.23,0.23,5000,10000,14,17,33893,p3a);
%simibe(0.05,0.1,0.23,0.23,5000,10000,14,17,33893,p3b);
%simibe(0.05,0.2,0.23,0.23,5000,10000,20,24,79087,p3c);
%simibe(0.05,0.3,0.23,0.23,5000,10000,20,24,79087,p3d);
%simibe(0.05,0.01,0.3,0.3,5000,10000,13,15,47939,p4a);
%simibe(0.05,0.1,0.3,0.3,5000,10000,14,16,41243,p4b);
%simibe(0.05,0.2,0.3,0.3,5000,10000,18,20,86453,p4c);
%simibe(0.05,0.3,0.3,0.3,5000,10000,18,20,86453,p4d);
%simibe(0.05,0.01,0.5,0.5,5000,10000,12,15,26701,p5a);
%simibe(0.05,0.1,0.5,0.5,5000,10000,13,15,23227,p5b);
%simibe(0.05,0.2,0.5,0.5,5000,10000,14,16,53551,p5c);
%simibe(0.05,0.3,0.5,0.5,5000,10000,14,16,53551,p5d);

proc print data=mi3.coveria;
title2;
run;

/*****************************************
SIMPBEC.SAS
This program does the interval coverage simulations
for PBE/2S4P.
Chapter 7, Section 4.2, Tables 3 & 4.
*****************************************/
options ps=64 ls=120 nodate nonumber;*print;
libname mi3 'c:\My Documents\My SAS Files\V8';
title 'Simulation for PBE intervals';
```

```
run;

%macro simpbe(del,r,sbt,sbr,swt,swr,nsim,ngp,seed,n1,n2,n3,n4,ds);

data sim;
****Input model parameter values****;
sigbt=&sbt;
sigbr=&sbr;
sigwt=&swt;
sigwr=&swr;
delta=&del;
rho=&r;
****Calculate needed parameter values****;
sig2bt=sigbt*sigbt;
sig2br=sigbr*sigbr;
sig2wt=sigwt*sigwt;
sig2wr=sigwr*sigwr;
sigsq1=sig2bt+sig2wt/2;
sigsq2=sig2br+sig2wr/2;
sig12=rho*sigbt*sigbr;
beta=sig12/sigsq1;
sigi=sqrt(sigsq1+sigsq2-2*sig12);
lambda1=sqrt(sigsq1);
lambdasq2=sigsq2-sig12*sig12/sigsq1;
lambda2=sqrt(lambdasq2);
mp3=(delta*delta+sig2bt+sig2wt-sig2br-sig2wr)/max(sig2br+sig2wr,0.04);
****Compute simulated values****;
junk=int(100000*ranuni(&seed));
do n=8,12,16,20;
nu=2*(n-1);
k=1/sqrt(2*n);
do i=1 to &nsim;
z=normal(0);
u1=2*rangam(0,nu/2);
u2=2*rangam(0,(nu-1)/2);
ut=2*rangam(0,nu/2);
ur=2*rangam(0,nu/2);
z12=normal(0);
d=delta+k*sigi*z;
sswt=sig2wt*ut;
sswr=sig2wr*ur;
ss1=sigsq1*u1;
ss2on1=lambdasq2*u2;
b=beta+z12*lambda2/sqrt(ss1);
ss2=b*b*ss1+ss2on1;
sig2i=(ss2+ss1*(1-2*b))/nu;
keep n i nu d ss1 ss2 ss2on1 sig2i sswt sswr b k mp3;
```

```
output;
end;
end;
run;
****Compute Ting-Burdick****;
data ting; set sim;
s_0=0.04;
eta0=((log(1.25))**2 + 0.02)/s_0;
stt=2*ss1/nu;
srr=2*ss2/nu;
st=sswt/nu;
sr=sswr/nu;
sn=2*n;
ssd=sn*d*d;
lambda=ssd/(2*sig2i);
nstar=ceil((1+2*lambda)*(1+2*lambda)/(1+4*lambda));
sdt=2*ssd/sn+stt+st;
ndt=ceil(sdt*sdt/(4*ssd*ssd/(sn*sn*nstar)+(stt*stt/nu)+(st*st/nu)));
if (srr+sr gt 2*s_0) then do;
src=srr+sr;
nrc=ceil(src*src/((srr*srr/nu)+(sr*sr/nu)));
f1=finv(0.05,ndt,nu);
f2=finv(0.95,nu,ndt);
f3=finv(0.05,nrc,nu);
f4=finv(0.95,nrc,ndt);
c0=1-2/nrc;
c2=(c0-f3)**2;
c1=((c0*(f2-1))**2)-c2*f2*f2;
c3=(f4-c0)**2;
c4=((c0*(1-f1))**2)-c3*f1*f1;
rhohat=(sdt-2*sig2i/sn)/src;
vu1=c1*sdt*sdt+4*c2*sig2i*sig2i/(sn*sn);
vu2=c3*sdt*sdt+4*c4*sig2i*sig2i/(sn*sn);
dt=sn*sdt/(2*sig2i);
if (dt le f1) then u_pber=c0*rhohat+sqrt(vu1)/src-1;
else if (dt gt f1) then u_pber=c0*rhohat+sqrt(vu2)/src-1;
chi=(u_pber gt mp3);
end;
else do;
sir=srr+sr+2*sig2i/sn;
nir=ceil(sir*sir/((4*sig2i*sig2i/(sn*sn*nu))+(srr*srr/nu)+(sr*sr/nu)));
h1=(ndt/cinv(0.05,ndt))-1;
g2=1-nir/cinv(0.95,nir);
c5=((1-finv(0.05,ndt,nir))**2-finv(0.05,ndt,nir)
*finv(0.05,ndt,nir)*h1*h1-g2*g2)/finv(0.05,ndt,nir);
vm3=sdt*sdt*h1*h1+sir*sir*g2*g2+c5*sdt*sir;
u_pbec=(sdt-sir+sqrt(vm3))/(2*s_0);
```

```
chi=(u_pbec gt mp3);
end;
run;

proc univariate data=ting noprint;
var chi;
output out=mls mean=mls;
by n;
run;
****Compute GPV****;
data call; set sim;
s_0=0.04;
eta0=((log(1.25))**2 + 0.02)/s_0;
xi=0;
do j=1 to &ngp;
z=normal(0);
z12=normal(0);
u1=2*rangam(0,nu/2);
u2=2*rangam(0,(nu-1)/2);
ut=2*rangam(0,nu/2);
ur=2*rangam(0,nu/2);
sig2wt=sswt/ut;
sig2wr=sswr/ur;
sigsq1=ss1/u1;
lambdasq2=ss2on1/u2;
beta=b-z12*sqrt(lambdasq2/ss1);
sigi=sqrt(sigsq1*(beta-1)**2 + lambdasq2);
delta=d-k*sigi*z;
sigsq2=lambdasq2+beta*beta*sigsq1;
gpvmp3=(delta*delta+sigsq1+sig2wt/2-sigsq2-sig2wr/2)/max(sigsq2+sig2wr/2,s_0);
xi=xi+(gpvmp3 lt mp3);
end;
p=xi/&ngp;
cover=(p le 0.95);
run;

proc univariate data=call noprint;
var cover;
output out=gpv mean=cover_gpv;
by n;
run;
*merge results;
data tests; merge mls gpv;
length test $ 5;
by n;
delta=&del;
r=&r;
```

```
sigma_bt=&sbt;
sigma_br=&sbr;
sigma_wt=&swt;
sigma_wr=&swr;
test="&ds";
run;

proc print data=tests;
run;

proc append base=mi3.coverp data=tests;
run;

%mend;

%simpbe(0.05,0.70,0.15,0.15,0.15,0.15,1000,1000,20291,6,7,8,9,p1a);
%simpbe(0.05,0.80,0.15,0.15,0.15,0.15,5000,10000,20297,6,7,8,9,p1b);
%simpbe(0.05,0.90,0.15,0.15,0.15,0.15,5000,10000,41659,6,7,8,9,p1c);
%simpbe(0.05,0.70,0.30,0.30,0.15,0.15,5000,10000,99487,5,6,9,10,p1d);
%simpbe(0.05,0.80,0.30,0.30,0.15,0.15,5000,10000,99481,5,6,9,10,p1e);
%simpbe(0.05,0.90,0.30,0.30,0.15,0.15,5000,10000,56821,5,8,9,10,p1f);
%simpbe(0.05,0.70,0.23,0.23,0.23,0.23,5000,10000,81847,7,8,9,10,p2a);
%simpbe(0.05,0.80,0.23,0.23,0.23,0.23,5000,10000,81841,7,8,9,10,p2b);
%simpbe(0.05,0.90,0.23,0.23,0.23,0.23,5000,10000,11597,7,8,9,10,p2c);
%simpbe(0.05,0.70,0.46,0.46,0.23,0.23,5000,10000,29053,6,8,9,10,p2d);
%simpbe(0.05,0.80,0.46,0.46,0.23,0.23,5000,10000,29059,6,8,9,10,p2e);
%simpbe(0.05,0.90,0.46,0.46,0.23,0.23,5000,10000,18593,5,6,9,10,p2f);
%simpbe(0.05,0.70,0.3,0.3,0.3,0.3,5000,10000,60743,7,8,9,10,p3a);
%simpbe(0.05,0.80,0.3,0.3,0.3,0.3,5000,10000,60737,7,8,9,10,p3b);
%simpbe(0.05,0.90,0.3,0.3,0.3,0.3,5000,10000,88069,7,8,9,10,p3c);
%simpbe(0.05,0.70,0.6,0.6,0.3,0.3,5000,10000,60083,6,8,9,10,p3d);
%simpbe(0.05,0.80,0.6,0.6,0.3,0.3,5000,10000,60077,6,8,9,10,p3e);
%simpbe(0.05,0.90,0.6,0.6,0.3,0.3,5000,10000,47629,5,8,9,10,p3f);
%simpbe(0.05,0.70,0.5,0.5,0.5,0.5,5000,10000,62293,7,8,9,10,p4a);
%simpbe(0.05,0.80,0.5,0.5,0.5,0.5,5000,10000,62299,7,8,9,10,p4b);
%simpbe(0.05,0.90,0.5,0.5,0.5,0.5,5000,10000,50503,7,8,9,10,p4c);
%simpbe(0.05,0.70,1.0,1.0,0.5,0.5,5000,10000,77087,6,7,9,10,p4d);
%simpbe(0.05,0.80,1.0,1.0,0.5,0.5,5000,10000,77081,6,7,9,10,p4e);
%simpbe(0.05,0.90,1.0,1.0,0.5,0.5,5000,10000,48673,5,6,8,9,p4f);


proc print;
run;
```

Appendix C

METHYLPHENIDATE DATA FROM MEYER *ET AL.*

Week	Subject	Treatment	Sequence	Replication	Cmax	AUC
1	1	1	1	1	10.61	40.13
2	1	1	1	2	11.85	45.62
3	1	2	1	1	9.38	43.34
4	1	2	1	2	8.52	53.05
1	2	1	1	1	6.54	34.54
2	2	1	1	2	7.53	34.24
3	2	2	1	1	8.03	39.3
4	2	2	1	2	8.42	28.66
1	3	1	1	1	6.82	33.3
2	3	1	1	2	6.87	39.85
3	3	2	1	1	8.74	35.78
4	3	2	1	2	8.04	42.13
1	4	1	1	1	10.02	72.52
2	4	1	1	2	12.44	70.15
3	4	2	1	1	8.29	64.33
4	4	2	1	2	18.23	93.55
1	5	1	1	1	7.94	27.53
2	5	1	1	2	9.48	30.92
3	5	2	1	1	5.92	24.12
4	5	2	1	2	6.08	22.87
1	6	1	2	1	6.74	28.77
2	6	2	2	1	7.69	34.83
3	6	2	2	2	4.24	21.38
4	6	1	2	2	4.36	19.55
1	7	1	2	1	8.14	33.25
2	7	2	2	1	10.87	39.26
3	7	2	2	2	11.89	35.56
4	7	1	2	2	6.15	22.85
1	8	1	2	1	5.3	33.09
2	8	2	2	1	6.03	33.09
3	8	2	2	2	2.26	16.04
4	8	1	2	2	4.03	19.55
1	9	1	2	1	3.85	18.67
2	9	2	2	1	6.1	24.44
3	9	2	2	2	4.1	29.61
4	9	1	2	2	3.12	20.51
1	10	1	2	1	6.14	28.89
2	10	2	2	1	7.38	32.74
3	10	2	2	2	6.76	30.06
4	10	1	2	2	7.7	32.42
1	11	2	3	1	5.87	36.82
2	11	2	3	2	5.15	24.5
3	11	1	3	1	6.6	28.83
4	11	1	3	2	6.7	26.41
1	12	2	3	1	10.03	53.19
2	12	2	3	2	9.02	47.59

3	12	1	3	1	8.82	37.71
4	12	1	3	2	7.23	41.8
1	13	2	3	1	7.89	27.95
2	13	2	3	2	4.25	24
3	13	1	3	1	6.7	29.56
4	13	1	3	2	5.56	33.17
1	14	2	3	1	7.06	20.48
2	14	2	3	2	11	32.9
3	14	1	3	1	3.65	24.24
4	14	1	3	2	5.01	21.5
1	15	2	3	1	6.58	37.49
2	15	2	3	2	8.33	45.96
3	15	1	3	1	7.51	34.47
4	15	1	3	2	5.27	35.35
1	16	2	4	1	6.65	27.87
2	16	1	4	1	6.01	23.97
3	16	1	4	2	8.82	47.7
4	16	2	4	2	7.51	35.61
1	17	2	4	1	6.85	29.37
2	17	1	4	1	5.82	23.08
3	17	1	4	2	4.9	23.41
4	17	2	4	2	6.25	37.04
1	18	2	4	1	5.21	16.55
2	18	1	4	1	3.99	17.57
3	18	1	4	2	2.99	15.48
4	18	2	4	2	3.1	14.94
1	19	2	4	1	6.91	39.77
2	19	1	4	1	7.86	33.83
3	19	1	4	2	4.71	22.75
4	19	2	4	2	5.12	24.17
1	20	2	4	1	12.8	52.82
2	20	1	4	1	11.1	47.39
3	20	1	4	2	9.22	56.32
4	20	2	4	2	10.7	58.61

Appendix D

BAYESIAN ANALYSIS OF POPULATION BIOEQUIVALENCE USING THE INDEPENDENCE CHAIN ALGORTHM IN PROC MIXED

Bayesian Analysis of Population Bioequivalence Using the Independence

Chain Algorithm in PROC MIXED

Richard McNally, Colorado State University, Fort Collins, CO

ABSTRACT

The statistical test for population bioequivalence given in the current FDA guidance document on this matter ignores the dependence between summary statistics. It is very conservative in cases where there is a high degree of correlation of subject responses between test and reference formulations, which is usually the case in bioequivalence studies. Adapting the ideas of Kass and Wolfinger (2000), we use the **PRIOR** statement in **PROC MIXED** to approximate the distribution of the population bioequivalence parameter Θ_{PBE}, from which we can construct tests based either on the probability of the upper tail region (*i.e.*, $Pr(\Theta_{PBE}>\Theta_0)$), or on the one-sided upper 95% confidence bound of Θ_{PBE},. We show that this test can conclude population bioequivalence when the FDA method does not. We also examine the statistical properties of this approach. The Kass and Wolfinger approach is also compared to another alternative given by McNally *et al* which is also more powerful than the FDA method, but is slightly anti-conservative.

INTRODUCTION

On 31 January 2001, the US Food and Drug Administration (FDA) issued a new guidance document *Statistical Considerations for Bioequivalence Studies* (2001). This document codified the requirement that bioequivalence studies included the assessment of population and individual bioequivalence, in addition to average bioequivalence. Two formulations of the same drug product are *average bioequivalent* if their mean difference is within allowed limits with reasonable assurance. More specifically, two products are average bioequivant if the 90% confidence interval of μ_T-μ_R is completely inside the interval (-0.22314,0.22314), assuming that the original data has been log-transformed. This test procedure is equivalent to the Two One-Sided Tests proposed by Schuirmann (1987), the formal statement of this is contained in the 1992 FDA guidance document on statistical analyses of bioequivalence studies. But even as it was being formalized, the adequacy of average bioequivalence (ABE) alone to assure that patients whose prescriptions are switched from a brand name drug to a generic product would receive the same therapeutic effect was being questioned. Also questioned was whether ABE was adequate to assure that a patient who received a generic product as a new prescription would receive the same therapeutic benefit as with the brand-name product. In response to these concerns, the respective concepts of individual bioequivalence (IBE) and population bioequivalence (PBE) were developed. The recent FDA guidance document on bioequivalence gives criteria and suggested testing methods for IBE and PBE. The test for IBE is taken from Hyslop, Hsuan, and Holder (2000), who constructed a method for calculating an upper 95% confidence bound based on the FDA criteria. In the guidance document, the agency adapted this procedure to construct a test for PBE. Unfortunately, some of the summary statistics used to construct the FDA test are not statistically independent, whereas in Hyslop *et al*, the summary statistics are independent. It appears that, when observations within a subject are highly correlated, which is usually the case in bioequivalence studies, the FDA test is very conservative. In fact, it is possible that this test will not conclude for the same data that provides evidence of IBE. This appears to the conceptual hierarchy that IBE is a higher standard than PBE, which in turn is a higher standard than ABE.

One approach that deals with the dependency problem was developed by McNally, Iyer, and Mathew (2001). This approach applied the generalized p-value (GPV) methodology introduced by Tsui and Weerahandi (1989). Using the GPV, McNally *et al* were able to develop a more powerful test for the PBE criterion, although this test appeared at times to be slightly anti-conservative. The present investigation examines a second approach. We extend the work of Kass and Wolfinger (2000), who applied a variant of Monte Carlo Markov chain (MCMC) methodology, the independence chain (IC) algorithm, to construct confidence intervals (which Bayesians would call credible regions) for individual variance components and general functions of these components. Best of all, the IC algorithm can be implemented using the **PRIOR** statement in **PROC MIXED**. Applying the IC algorithm to the PBE criterion produces a test that appears to have properties similar to those of the GPV, and is much more powerful than the FDA test.

STATISTICAL DESIGN AND MODEL

Consider the following 2-sequence/4-period crossover design

Period	1	2	3	4
Sequence 1	**T**	**R**	**T**	**R**
Sequence 2	**R**	**T**	**R**	**T**

where **T** represents the test (generic) formulation and **R** represents the reference (innovator) formulation. To assess population bioequivalence for data from this study design, we use the following model

$$Y_{ijkl}=\mu_k + \gamma_{ikl} + \eta_{ijk} + \varepsilon_{ijkl} \qquad (1)$$

where i=1,2 is the sequence number, j=1..n_i is the j^{th} subject in sequence i, k=T,R is the treatment, and l=1,2 is the replicate number. That is to say, Y_{ijkl} is the response of subject j in sequence i receiving replicate l of treatment k. The fixed effect of treatment k is μ_k, and the fixed effect of cell ikl is γ_{ikl}. The random subject effect is η_{ijk}, and the random within-subject effect is ε_{ijkl}. The random subject effect has a bivariate normal distribution for each subject with a zero mean vector and the following covariances

$Var(\eta_{ijT})=\sigma^2_{BT}$ $Var(\eta_{ijR})=\sigma^2_{BR}$ $Cov(\eta_{ijT}, \eta_{ijR})=\rho\,\sigma_{BT}\sigma_{BR}$

The within-subject effect has a normal distribution with mean 0 and variance σ^2_{Wk} (k is the treatment). This is the statistical model used by both Hyslop *et al* and the FDA guidance document for analyzing PBE and IBE data. For details on this model, see Chinchilli and Esinhart (1996).

The FDA test for PBE is based on the following summary statistics

$U_{ijT}=(Y_{ijT1}+Y_{ijT2})/2$ $U_{ijR}=(Y_{ijR1}+Y_{ijR2})/2$
$V_{ijT}=(Y_{ijT1}-Y_{ijT2})/\sqrt{2}$ $V_{ijR}=(Y_{ijR1}-Y_{ijR2})/\sqrt{2}$

The problem with the FDA test arise with the joint distribution of U_{ijT} and U_{ijR}; they have covariance $\rho\sigma_{BT}\sigma_{BR}$, so that they are independent only when ρ=0. Now Hyslop *et al* (2000) computed a confidence interval for the IBE criteria, which is a linear combination of independent components. Their exposition is based on the work of Howe (1974), who computed a confidence interval for the difference of 2 independent mean squares. But consider the PBE criterion

$$\Theta_{PBE} = \frac{\delta^2 + \sigma^2_{TT} - \sigma^2_{TR}}{\max(\sigma^2_{TR}, \sigma^2_0)} \qquad (2)$$

where $\delta=\mu_T-\mu_R$, $\sigma^2_{TT}=\sigma^2_{BT}+\sigma^2_{WT}$ and $\sigma^2_{TR}=\sigma^2_{BR}+\sigma^2_{WR}$. By ignoring the dependence noted above, the FDA test appears to put much too high of a bound on linear combinations involving σ^2_{BT} and σ^2_{BR}, so that the FDA test is very conservative. The same is true of the test based on the confidence intervals developed by Quiroz *et al* (2001). McNally *et al* (2001) solved this problem by decomposing the covariance matrix of the joint normal distribution of the (U_{ijT},U_{ijR}) pairs, which produced a test more powerful than the FDA test. This test was based on the generalized p-value (GPV) methodology of Tsui and Weerahandi (1989). Tsui and Weerahandi noted similarities between the numerical calculations given by the GPV method for a given problem (*e.g.*, the Behren-Fisher problem) and by various Bayesian formulations of the same problem. These remarks raised our interest in approaching the PBE problem from a Bayesian perspective.

Wolfinger and Kass (2000) approached the problem of calculating confidence intervals for general functions of variance components using a Monte Carlo Markov Chain (MCMC) method called the independence chain algorithm (IC). The idea is that in certain mixed models (*e.g.*, the one-way random-effects model with balanced group sizes) the mean squares have well-defined (*i.e.*, Chi-squared) distributions. Therefore, we can implement MCMC by sampling directly from the posterior distribution of the random effects. Once a set of a random effects is generated, they can be plugged into a multivariate normal likelihood to generate random realizations of the fixed effects. In other words, the posterior of the fixed effects is the likelihood times a flat prior (f(x)=1). The base distribution (*i.e.*, the distribution that generates MCMC candidate points) for the random-effects is exact, so that MCMC samples are rejected only if they are not in the parameter space, which may occur, for example, if an individual variance component is estimated to be negative. Even for modestly unbalanced data, the base distribution is usually a good approximation to the true prior. This algorithm is implemented in SAS using the PRIOR statement in PROC MIXED.

The FDA guidance document recommends that sponsors use method of moments (MoM) estimators for the mean difference and variance components instead of restricted maximum likelihood (REML) estimators when assessing IBE and PBE. The reason for this is given by Chen *et al* (2000), that REML relies heavily on the assumption of normality, whereas MoM does not. But Chinchilli and Esinhart showed that, for the model given above for the 4-period crossover design, the MoM and the unrestricted REML estimates are the same (the unrestricted REML estimates for the random-subject effects are those given by using `type=un` in the **RANDOM** statement). Therefore, the distinction between REML and MoM may seem unimportant here. This could be considered auspicious, since the posterior that we are using to generate MCMC samples is conditioned on the REML estimates. The MCMC samples are also generated within the context of normal linear models theory, which may appear to be at odds with regulatory reluctance to assume normality. But the

only inferential method proposed for PBE that does not use normality for inference is the nonparametric bootstrap.

INDEPENDENCE CHAIN ALGORITHM

The independence chain (IC) algorithm is a special case of rejection sampling. See Tierney (1994) for a good overview of MCMC methodology, including rejection sampling and, in particular, the independence chain. Model (1) is a special case of the general linear mixed model

$$y=\mathbf{X}\beta + \mathbf{Z}\gamma + \varepsilon \qquad (3)$$

where y is a vector of observed responses, **X** is the design matrix corresponding to the fixed-effects parameter vector β, **Z** is the design matrix corresponding to the random-effects parameter vector γ, and ε is the residual error vector. We assume that γ and ε are independent and have multivariate normal distributions with mean 0 vectors and covariance matrices G and R, respectively. In the variance component setup, G and R are diagonal matrices (*i.e.*, the effects associated with the variance components are independent), but for the population bioequivalence problem we will need G to be a general covariance matrix. Let θ be the vector of variance components, *i.e.*, the components of G and R. Since we are working in the Bayesian framework, the analysis revolves around the posterior distribution of the model parameters. The joint posterior density of (β,γ,θ) is

$$f(\beta,\gamma,\theta|y)=f(\beta,\gamma|\theta,y)f(\theta|y) \qquad (4)$$

The basic procedure is to simulate the density $f(\theta|y)$ of the variance components using an independence chain, then, conditioning on θ, to simulate from $f(\beta,\gamma|\theta,y)$ as a multivariate normal distribution. We assume, with Wolfinger and Kass, that β has a uniform prior $f(\beta)=1$. Now

$$f(\theta|y)\propto f(y|\theta)f(\theta) \qquad (5)$$

is the posterior on θ. The prior distribution $f(\theta)$ that we use is the Jeffreys' prior (Jeffreys' prior is actually a class of uninformative and often improper prior distributions).

Wolfinger and Kass summarized the IC algorithm for variance components in the following way:

1. Transform the variance components to approximate independence using the ratios established in the MIVQUE(0) coefficient matrix.
2. Determine inverted-gamma densities for each transformed parameter by performing a linear regression on a series of grid evaluations of the log-marginal posterior of the variance components, $f(\theta|y)$. This posterior is computed by taking the product of the restricted likelihood and some prior.
3. Form the base sampling density for the transformed variance components, $g(\theta|y)$, by taking the product of the inverted gamma densities from step 2.
4. Use an independence chain to generate a pseudo-random sample from $f(\theta|y)$.
5. Given sampled values for the variance components, use the multivariate normal distribution to generate observations for the mean [*i.e.*, fixed] parameters.

In the independence chain algorithm (step 4), MCMC proposals are drawn from a fixed based sampling density $g(\theta|y)$. Since the proposal density is exactly the posterior for model (1), the acceptance probability of a MCMC proposal should be close to 1. What complicates the situation for population bioequivalence is that the functions of the variance components in which we are interested are not independent. So we modify the above method by allowing an inverted-Wishart density to be used in step 2. This should give tests constructed based on the IC algorithm an advantage over tests that do not take statistical dependence into account.

Since we are interested in tests of hypothesis, we use the posterior distribution to construct tests for population bioequivalence. For a general function $T(\beta,\theta)$ of the mixed model parameters, if H_0:$T(\beta,\theta)>T_0$ is the null hypothesis, then we can compute an estimate of the 95-percentile of this function and compare it to the pre-specified upper bound T_0. The values less than this percentile would form a one-sided 95% credible region, the Bayesian equivalent of a confidence interval, for $T(\beta,\theta)$. An alternative is to compute the area of the extreme or tail region, *i.e.*, $D(X)=Pr(T(\beta,\theta)>T_0)$, where D(X) is a function of the data, called the discrepancy of the observed data (Meng, 1994). This procedure is similar to the calculation of a frequentist p–value, and is even more similar to the generalized p-value of Tsui and Weerahandi (1989), the methodology on which McNally *et al* based their test of population bioequivalence. Calculations of this type are not generally accepted by Bayesians, but Rubin (1984) defended this type of test procedure as being Bayesianly justifiable under certain circumstances, and cited the use of tail probabilities by Box and Tiao to evaluate model differences. Rubin subsequently gave this measure of discrepancy a name, the *posterior-predictive p-value*. Meng develops some of the theory of posterior-predictive p-values, and he also notes their similarity to generalized p-

values.

IMPLEMENTING THE IC ALGORITHM IN PROC MIXED

The following is the SAS® code used to implement the IC algorithm.

```
data parmtrans;
input covp1-covp5;
cards;
2 0 0 1 0
0 2 0 0 0
0 0 2 0 1
0 0 0 1 0
0 0 0 0 1
;;;;
run;

proc mixed data=mph;
class treatment;
model lnauc=treatment gamma1-gamma6 / ddfm=satterth;
random treatment/subject=subject type=un g;
repeated/group=treatment;
estimate 'F' treatment -1 1 / cl alpha=0.1;
prior / alg=ic nsample=1e4 lognote=1e3 tdata=parmtrans out=sample;
run;

data pbe; set sample;
s_0=0.04;
theta0=((log(1.25))**2 + 0.02)/s_0;
thetap=((est1*est1)-covp1+covp3-covp4+covp5)/max(covp1+covp4,s_0);
psi=(thetap gt theta0);
run;

proc univariate data=pbe;
var thetap psi;
run;
```

The first data step is needed so that **PROC MIXED** uses the proper transformation of the covariance parameters. The default action is for the procedure to generate this transformation automatically. For balanced random-effects models, the coefficients of this transformation are those of the estimated mean squares. In general, these transformations are the MIVQUE(0) equations, normalized by dividing through by the coefficient of the residual effect. But in the population bioequivalence model, the residual effects depend on treatment, so the procedure cannot properly generate this transformation. Fortunately, the `tdata=parmtrans` option in the **PRIOR** statement allows the user to specify the correct transformation.

The **PROC MIXED** call fits the model in equation (1) above to the log-transformed pharmacokinetic data. Here the pharmacokinetic parameter is area under the plasma-concentration/time curve (AUC). The `gamma1-gamma6` are indicator variables that must be created in a data step. Since the model (1) has a mean effect for each treatment, there are estimability conditions on the gammas. Chinchilli and Esinhart have the details on these conditions. Using `type=un` in the **RANDOM** statement gives the method of moments estimates for the variance components in this design. The `group=treatment` option in the **REPEATED** statement allows us to have separate residual terms for each treatment group. The use of the `group` option is what necessitates the use of the `tdata` option in the **PRIOR** statement. The treatment contrast in the **ESTIMATE** statement is needed so that estimates of the δ parameter in (2) will be included in the MCMC samples. Here the **ESTIMATE** statement options request a 90% confidence interval for the estimate of δ. This is considered equivalent to the two one-sided tests for average bioequivalence of Schuirmann (1987).

The **PRIOR** statement implements the MCMC sampling routine in **PROC MIXED**. Here we are using the default prior distribution in MCMC sample generation, which is the Jeffreys', or uninformative, prior. The user can specify their own prior by creating a data set with the distribution name and the parameters of the prior, which is called using `data=<data set>`. An example of a user-specified prior is given in Wolfinger and Kass. The user can also specific a `flat` prior, $f(\theta)=1$, in which case the posterior used to generate the MCMC samples for the variance component is the likelihood function. We have also specified the following options on the **PRIOR** statement. The option `alg=ic` specifies that the independence chain algorithm is to be used to generate MCMC samples as described in the previous section. Technically, this is not needed as the independent chain is the default. The `nsample=10e4` statement specifies that 10000 MCMC samples be generated and put into the `bayes` data set specified in the `out=` statement. The default number of samples is

1000, but Wolfinger and Kass recommend 10000 samples in order to get more accurate results. We follow the progress of the sample generation using the `lognote=10e3` option, which causes a message to be printed on the LOG for every 1000 samples generated. We have already discussed the use of `tdata=parmtrans`. Other useful options that give information on the generation of MCMC samples can be found in the SAS online help on **PROC MIXED**.

The data step that follows computes the population bioequivalence function, and both the boundary and the indicator of the tail region. In this data step est1=δ, covp1=σ^2_{BR}, covp2=$\rho\sigma_{BT}\sigma_{BR}$, covp3=$\sigma^2_{BT}$, covp4=$\sigma^2_{WR}$, and covp5=$\sigma^2_{WT}$, which can be used to calculate estimates of the population bioequivalence criterion (equation (2)). We then use **PROC UNIVARIATE** to compute the percentiles of the population bioequivalence function and the posterior-predictive p-value.

EXAMPLE: METHYLPHENIDATE

Meyer *et al* (2000) studied the individual bioequivalence of brand name and generic versions of methylphenidate (better known by the brand name Ritalin™), which was evaluated using the method of Hyslop *et al* (2000), whose methodology is recommended by the FDA. McNally *et al* (2001) also examined the individual and population bioequivalence of these of these drug formulations. In examining the pharmacokinetic metric Area Under the concentration-time Curve (AUC), McNally *et al* used the generalized p-value (GPV) to conclude that the brand name and generic formulations were population bioequivalent, contrary to what would be concluded according to the methodology recommended in the FDA guidance document (2001), and also contrary to the conceptual hierarchy that IBE is a higher standard than PBE, since Meyer *et al* concluded the formulations were individual bioequivalent with respect to AUC. Bootstrap analysis seemed to confirm the conclusions of McNally *et al*. The difference between the GPV and FDA tests is that the GPV test accounts for the dependence between summary statistics, which appear to yield a more powerful test for population bioequivalence. Since the independence chain methodology also accounts for dependence (technically, in the Bayesian context the dependence is between the variance components), its performance should be similar to that of the GPV test.

The parameters for the base density are ν=16, σ_{11}=4.1570, σ_{12}=3.1467, σ_{22}=2.9777 for the inverted Wishart, α=8 and β=0.4150 for the inverted gamma corresponding to σ^2_{WT}, and α=8 and β=0.2723 for the inverted gamma corresponding to σ^2_{WR}. The values of parameters of the inverted gamma distribution are, respectively, one-half the degrees of freedom for calculating the variance components and one-half the sum of squares of the corresponding variance component. Table 1 summarizes the methylphenidate AUC population bioequivalence results. The first group of numbers are upper 95% confidence bounds based, respectively, on the generalized confidence interval (Weerahandi (1993), GCI), the independence chain (ICCR), and the FDA method. The upper bound for both the GCI and ICCR are similar and well under the regulatory upper bound of 1.7448, so that we can conclude population bioequivalence from both methods. The FDA interval is on a different scale, which requires that the upper bound be less than 0 in order to conclude population bioequivalence, which it is not in the table. Therefore, from the FDA test we would draw a conclusion contrary to that of the other 2 methods. The next 2 numbers are the areas of tail or extreme regions (*i.e.*, $Pr(\Theta_{PBE}>1.7448)$, which are similar to the conventional frequentist p-value. The first is generalized p-value (GPV); the second is the posterior predictive p-value (PPPV). In practice, these values can be used in the same manner as conventional p-values to assess statistical hypotheses. Again, the values are similar and both lead credence to the belief that the 2 formulations of methylphenidate are population bioequivalent. The FDA method does not yield anything comparable to a p-value.

Table 1. Population Bioequivalence of Methylphenidate

95% Upper Bound			Area of Critical Region	
ICCR	GCI	FDA	GPV	PPPV
1.122	1.101	0.02836	**0.0038**	**0.0040**

*Values in **bold** indicate the conclusion of population bioequivalence.*

SIMULATIONS

We did some small simulation studies to evaluate the statistical properties of the IC testing methodology for population bioequivalence. The size of the test was evaluated by examining various points on the boundary of the null hypothesis for 6 subjects per sequence. For each set of parameter values 500 sets of individual observations were generated. For each set of observations, 2000 MCMC points were generated which were used to assess population bioequivalence. The estimated size of the test is percentage of replicated data sets out of 500 that were judged to be population bioequivalent by evaluating the posterior predictive p-value at the α=0.05 level. The test size estimates are summarized in Table 2.

The results in this table are based on a relatively small simulation samples. Therefore, the estimated sizes should be treated as rough approximations. However, it is clear from the above table that the size of the IC test for population bioequivalence is slightly anti-conservative for at least a good part of the parameter space. This is most notable when δ is outside of the acceptable region for average bioequivalence (*i.e.*, $|\delta|>\log(1.25)$), as was observed by McNally *et al* for the GPV test. The estimated sizes are consistent with a result of Meng, that the size of a test based on the posterior predictive p-value has an upper bound of 2α, where α is the nominal significance level of the test procedure.

Table 2. Estimated Size of IC Test for Population Bioequivalence n=6

	δ=0.2954			δ=0
	σ_{BT}=0.2 σ_{BR}=0.2 σ_{WT}=0.1	σ_{BT}=0.1732 σ_{BR}=0.2 σ_{WT}=0.1414	σ_{BT}=0.2 σ_{BR}=0.1732 σ_{WT}=0.1	σ_{BT}=0.3567 σ_{BR}=0.2 σ_{WT}=0.1
ρ	σ_{WR}=0.1	σ_{WR}=0.1	σ_{WR}=0.1414	σ_{WR}=0.1
0.8	0.088	0.042	0.084	0.046
0.9	0.092	0.066	0.074	0.062
	δ=0.5907			δ=0
	σ_{BT}=0.4 σ_{BR}=0.4 σ_{WT}=0.2	σ_{BT}=0.3873 σ_{BR}=0.4 σ_{WT}=0.2236	σ_{BT}=0.4 σ_{BR}=0.3873 σ_{WT}=0.2	σ_{BT}=0.7134 σ_{BR}=0.4 σ_{WT}=0.2
ρ	σ_{WR}=0.2	σ_{WR}=0.2	σ_{WR}=0.2236	σ_{WR}=0.2
0.8	0.066	0.060	0.056	0.046
0.9	0.066	0.070	0.054	0.044

	δ=0.8861	δ=0
	σ_{BT}=0.6 σ_{BR}=0.6	σ_{BT}=1.0701 σ_{BR}=0.6
ρ	σ_{WT}=σ_{WR}=0.3	σ_{WT}=σ_{WR}=0.3
0.8	0.060	0.038
0.9	0.058	0.046

The power of the IC test for population bioequivalence was evaluated for δ=0.05, σ_{BT}=σ_{BR}=0.3,0.46,0.6,1.0, σ_{WT}=σ_{WR}=0.15,0.23,0.3,0.5, and ρ=0.9,0.95. For each set of parameter values 1000 sets of individual observations were generated and assessed for population bioequivalence in the same manner as for the evaluation of test size. The power of the is the number of replicated data sets out of 1000 that were judged to be population bioequivalent, based on evaluating the posterior predictive p-value at the α=0.05 level. The power summary in Table 3 shows the minimum number of subjects per sequence needed to achieve 80% power, and the corresponding power.

Table 3. Power of IC Test for Population Bioequivalence

δ=0.05

	$\sigma_{BT}=\sigma_{BR}=0.3$		$\sigma_{BT}=\sigma_{BR}=0.46$	
	$\sigma_{WT}=\sigma_{WR}=0.15$		$\sigma_{WT}=\sigma_{WR}=0.23$	
ρ	N	Power	N	Power
0.9	6	0.805	6	0.803
0.95	6	0.843	6	0.866
	$\sigma_{BT}=\sigma_{BR}=0.6$		$\sigma_{BT}=\sigma_{BR}=1.0$	
	$\sigma_{WT}=\sigma_{WR}=0.3$		$\sigma_{WT}=\sigma_{WR}=0.5$	
ρ	N	Power	N	Power
0.9	6	0.826	6	0.809
0.95	6	0.856	6	0.876

If we compare these sample size requirements and power numbers to those in McNally *et al*, we find that the IC test produces similar power results to that of the generalized p-value test (GPV), and has much more power than the FDA test. The GPV test appears to be slightly more powerful than the IC test, but the IC algorithm appears to produce a much more acceptable test for population bioequivalence than the methodology currently recommended by the FDA.

DISCUSSION

In the debate over the new bioequivalence standards now mandated by the FDA, individual bioequivalence has been given most of the attention; population bioequivalence has largely been ignored. The test for individual bioequivalence proposed by Hyslop *et al* and recommended by the FDA is relatively straightforward and works very well for that parameter. It would be perfectly sensible to apply basically the same test to population bioequivalence, as the FDA has done, if not for the statistical dependence between the summary statistics. The new bioequivalence criteria are meant to address the needs of patients who could be exposed to a generic version of a drug product under the assumption that they will derive the same therapeutic effect as they would from the innovator product. The most critical scenario to be considered is whether a patient who is already receiving the innovator product will derive the same therapeutic benefit if they are switched to the generic product, which leads to individual bioequivalence. This 'switchability' is especially crucial when a patient has been titrated to a specific dose. Less critical but still important is whether a patient who is prescribed the generic product *de novo* will derive the same therapeutic benefit as he would had the innovator product been prescribed. This is the idea behind population bioequivalence. Although the bioequivalence criteria do not have any structural hierarchy, there is still a conceptual hierarchy that individual bioequivalence is a higher standard than population bioequivalence, which in turn is a higher standard than average bioequivalence. This is the seeming inconsistency in the methylphenidate example. Meyer *et al* were able to use the Hyslop method to conclude that the test and reference formulations of methylphenidate were individual bioequivalence with respect to AUC. McNally *et al* showed that the same conclusion could be reached using the GPV method and even the confidence interval methodology of Quiroz *et al* (2001). It is troubling that we cannot used the FDA test to conclude population bioequivalence for the methylphenidate, and it seems unlikely that this would be an isolated case, given what seems to be overwhelming evidence (i.e., very low p-values and upper bounds much lower than the upper limit of Θ_{PBE}) in support of population bioequivalence that seems to come from the two test procedures that take statistical dependence into account, the GPV and the IC tests. Therefore, these tests would appear to be a significant improvement on tests, such as the FDA test, that assume that summary statistics are independent.

We may go further to assert that assuming that U_{ijT} and U_{ijR} (the sums of observed values for the respective treatments within subjects) are independent goes against the rationale for using crossover designs to assess bioequivalence, that observations within a subject are highly correlated. Even in data sets where we could not reject the null hypothesis (*i.e.*, where we could not conclude bioequivalence), we have observed values of the correlation coefficient in the range of r=0.8. This is why we have assumed that the observations within subjects are strongly correlated even when we have evaluated the size of tests for population bioequivalence. In fact, using the method of moments estimators we have commonly observed estimated values for ρ of greater than 1 (this may be considered a drawback of using the method of moments, an issue that goes beyond the scope of this paper). The point is that a testing strategy should reflect the statistical structure of the problem. We feel that that the GPV and IC tests for population bioequivalence are superior to the FDA test because they acknowledge the statistical dependency in the problem.

Naturally the question arises which is the better test for population bioequivalence, the GPV or the IC algorithm? Both Tsui and Weerahandi (1989) and Meng (1994) cite examples where the GPV gives the same answer as the Bayesian approach with a non-informative prior, but this does not seem to be the case here. Our tendency is to favor the GPV test, mainly for the following reasons.

1. The GPV test appears to have better power (see Table 4 for sample size requirements and power for the GPV test).

2. Despite seeming to be slightly less powerful, the IC algorithm does not appear uniformly to maintain the nominal level of significance.
3. Our SAS code for the GPV test generally takes a few seconds to produce a result based on 50000 simulation points. The IC test generally takes 2 to 5 minutes to produce a result based on 10000 MCMC points. More points generally mean more accurate results.

Table 4. Power of GPV test for Population Bioequivalence
(from McNally *et al*) δ=0.05

	$\sigma_{BT}=\sigma_{BR}=0.3$		$\sigma_{BT}=\sigma_{BR}=0.46$	
	$\sigma_{WT}=\sigma_{WR}=0.15$		$\sigma_{WT}=\sigma_{WR}=0.23$	
ρ	N	Power	N	Power
0.9	6	0.854	6	0.860
0.95	5	0.847	5	0.841
	$\sigma_{BT}=\sigma_{BR}=0.6$		$\sigma_{BT}=\sigma_{BR}=1.0$	
	$\sigma_{WT}=\sigma_{WR}=0.3$		$\sigma_{WT}=\sigma_{WR}=0.5$	
ρ	N	Power	N	Power
0.9	6	0.852	6	0.861
0.95	5	0.839	5	0.842

Our conclusions with respect to the power and size of the 2 tests parallel those of Chiang (2001), who compared the method of surrogate variables to the IC algorithm for variance components (the method of surrogate variables is essentially the same as the methodology for Weerahandi's generalized confidence intervals). These conclusions may appear to be rash, especially as regards the size of the IC test, since they are based relatively small simulations for the IC test. Indeed, a more extensive investigation of the statistical properties of the IC test for population is probably warranted, similar to what McNally *et al* have done for the GPV test. An extensive examination of the use of informative priors in population and individual bioequivalence may also be useful, although informative priors may be regarded as problematic in a regulatory setting.

The main advantage of the IC test for population over the GPV test is that, thanks to the **PRIOR** statement in **PROC MIXED**, it is relatively straightforward to implement. There is a caveat, that the user needs to know what covariance parameters correspond to what treatments. This depends on how the treatments are ordered in SAS. On the other hand, the GPV test for population bioequivalence involves using a very complicated test function. This is in contrast to the GPV test for individual bioequivalence, which is also somewhat straightforward to program using SAS data steps. SAS code for the GPV tests for individual and population bioequivalence is available from the author.

CONCLUSION

The independence chain (IC) algorithm provides a relatively straightforward methodology for assessing population bioequivalence. Although the IC test may not be the best method available, it can easily be implemented using the **PRIOR** statement in **PROC MIXED**, and it is superior to the test currently recommended by the FDA.

REFERENCES

Chen ML, Patnaik R, Hauck WW, Schuirmann DJ, Hyslop T, Williams R (2000). An individual bioequivalence criteria: regulatory considerations. Statistics in Medicine, **19**:2821-2842.

Chiang AKL (2001). A simple general method for constructing confidence intervals for functions of variance components. Technometrics, **43**:356-367.

Chinchilli VM, Esinhart JD (1996). Design and analysis of intra-subject variability in cross-over experiments. Statistics in Medicine, **15**:1619-1634.

Howe WG (1974). Approximate confidence limits on the mean of X+Y where X and Y are two tabled independent random variables. Journal of the American Statistical Association, **69**:789-794.

Hyslop T, Hsuan F, Holder DJ (2000). A small sample confidence interval approach to assess individual bioequivalence. Statistics in Medicine, **19**:2885-2897.

McNally RJ, Iyer H, Mathew T (2001). Tests for individual and population bioequivalence based on generalized p-values. Colorado State Univ., Dept. of Statistics Technical Report 2001/11.

Meng XL (1994). Posterior predictive p-values. Annals of Statistics, **22**:1142-1160.

Meyer MC, Straughn AB, Jarvi EJ, Patrick KS, Pelsor FR, Williams RL, Patnaik R, Chen ML, Shah VP (2000). Bioequivalence of methylphenidate immediate-release tablets using a replicated study design to characterize intrasubject variability. Pharmaceutical Research, **17**:381-384.

Quiroz J, Ting N, Wei GCG, Burdick RK. Alternative confidence intervals for the assessment of bioequivalence in four-period crossover designs. Statistics in Medicine, *to appear.*

Rubin DB (1984). Bayesianly justifiable and relevant frequency calculations for the applied statistician. Annals of Statistics, **12**:1151-1172.

Schuirmann DJ (1987). A comparison of the two one-sided tests procedure and the power approach for

assessing the equivalence of average bioavailability. J Pharmacokin Biopharm; 715:657-680.
Tierney L (1994). Markov chains for exploring posterior distributions (with discussion). Annals of Statistics, **23**:1701-1762.
Tsui KW, Weerahandi S (1989). Generalized p-values in significance testing of hypotheses in the presence of nuisance parameters. Journal of the American Statistical Association, **84**:602-607.
US Food and Drug Administration (1992). Statistical procedures for bioequivalence studies using a standard two-treatment crossover design. Rockville, Md.
US Food and Drug Administration (2001). Statistical considerations for bioequivalence studies. Rockville, Md.
Weerahandi S (1993). Generalized confidence intervals. Journal of the American Statistical Association, **88**:899-905.
Wolfinger RD, Kass RE (2000). Non-conjugate Bayesian analysis of variance component models. Biometrics, **56**:768-774.

ACKNOWLEDGMENTS

I would like to thank Marvin Meyer and Arthur Straughn of the University of Tennessee, Memphis for providing the methylphenidate data. I would also like to thank Hari Iyer of Colorado State University for his helpful comments.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Richard McNally
Department of Statistics
Colorado State University
Fort Collins, CO 80523-1877
Work Phone: 970-491-5702
Email: mcnally@stat.colostate.edu