# LIKELIHOOD APPROACH FOR EVALUATING BIOEQUIVALENCE OF HIGHLY VARIABLE DRUGS

By

Liping Du

Thesis

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

MASTER OF SCIENCE

in

Biostatistics

August, 2013

Nashville, TN

Approved:

Professor Leena Choi

Professor William D. Dupont

## ACKNOWLEDGMENTS

First, I would like to express my most sincere gratitude to Dr. Leena Choi, my thesis adviser, for her guidance throughout this research and writing. I also wish to thank Dr. William D. Dupont, as a member of my thesis committee, for his valuable suggestions. I would like to extend my deep appreciation to the faculty in the Biostatistics Department of Vanderbilt University for their teaching and encouragement. I am grateful for my classmates and colleagues for their assistance. I specially thank Mingsheng Guo for helping me with the simulations, and Peggy Schuyler for editing. In addition, I appreciate the support by National Institutes of Health grant (R21 AG034412) awarded to Dr. Leena Choi.

Finally, I would like to express my deep gratefulness to my husband, Xiaofeng, Qi, three kids, my parents, and my parents in-law. Without their long lasting understanding and support, I could not finish this program and this thesis.

## TABLE OF CONTENTS

## LIST OF TABLES

## LIST OF FIGURES

## LIST OF ABBREVIATIONS

| | |
|---|---|
| $C_{max}$ | peak concentration |
| ABE | average bioequivalence |
| AUC | the area under the blood concentration-time curve |
| BE | bioequivalence |
| BIE | bioinequivalence |
| CI | confidence interval |
| CV | coefficient of variation |
| EMA | European Medicines Agency |
| FDA | the United States Food and Drug Administration |
| GLL | generalized law of likelihood |
| GLR | generalized likelihood ratio |
| GMR | geometric mean ratio |
| HVD | highly variable drug |
| IBE | individual bioequivalence |
| LI | likelihood interval |
| MLE | maximum likelihood estimate |
| PBE | population bioequivalence |
| PK | pharmacokinetic |
| R | reference drug |
| RSABE | reference scaled average bioequivalence |
| T | test drug |
| TOST | Two-One-Sided-Tests |

# CHAPTER I

## INTRODUCTION

Bioavailability equivalence or bioequivalence (BE) is required by the United States Food and Drug Administration (FDA) for approval and marketing of generic drugs or new formulations of an existing drug (Davit et al., 2009). A test drug is bioequivalent to a reference (such as innovator) drug if the rate and extent of absorption of the test drug do not show a significant difference from the reference drug when administered at the same molar dose of the therapeutic ingredient under similar experimental conditions in either a single dose or multiple doses (FDA, 2012). Generally, establishing BE requires *in vivo* pharmacokinetic studies that compare the blood/plasma drug concentration-time profile of the test drug with that of a reference drug in healthy adults typically in a cross-over design (FDA, 2001; Tamboli et al., 2010).

The appropriate statistical hypotheses for BE (referred as interval hypotheses) (Westlake, 1972; Hauck and Anderson, 1984; Schuirmann, 1987; Hoenig and Heisey, 2001) are:

$$\begin{aligned} H_0: &\quad \mu_T - \mu_R \leq \theta_1 \ \text{ or } \ \mu_T - \mu_R \geq \theta_2; \\ H_1: &\quad \theta_1 < \mu_T - \mu_R < \theta_2, \end{aligned} \tag{I.1}$$

where $\mu_T$ and $\mu_R$ denote the population means of a pharmacokinetic parameter (typically logarithmically transformed AUC, the area under the blood concentration-time curve, or Cmax, the peak concentration) for the test (T) and the reference (R) drugs, respectively, and $\theta_1$ and $\theta_2$ denote the specified equivalence limits; the current FDA guidance suggests $\theta_1 = \log 0.8$ and $\theta_2 = \log 1.25$ so that $-\theta_1 = \theta_2 = 0.223$.

Schuirmann's Two-One-Sided-Tests (TOST) procedure (Schuirmann, 1987), equivalently, the 90% confidence interval (CI) approach, has been used as the mainstream method for testing the above hypotheses since 1992 by the FDA and other countries (Tamboli et al., 2010). According to the FDA, BE is demonstrated if the 90% CIs for $\mu_T - \mu_R$ for all three pharmacokinetic measures ($AUC_{0-t}$, $AUC_{0-\infty}$, and Cmax), fall completely within the [-0.223, 0.223] limit. In other words, the 90% CIs of the geometric mean ratios (GMR, exponentiated mean difference in log scale) for all three pharmacokinetic measures fall completely within the limits (in %) of 80-125% (FDA, 2001). This is particularly called average bioequivalence (ABE) criterion.

Despite a review of 12 years of BE data from the FDA shows that the ABE criterion support FDA's objective of approving high quality generic drugs (Davit et al., 2009), there are some concerns over some generic drugs that have a narrow therapeutic window such as some anti-epileptic drugs and anti-coagulant drug, Warfarin (Shaw and Hartman, 2010; Talati et al., 2012; Dentali et al., 2011). The one size-fits-all criterion has

come under criticism by many authors because it does not consider the therapeutic window and intra-subject variability of a drug, and thus does not address the issues of drug prescribability [for which a concept of population bioequivalence (PBE) may be applied] and switchability [for which a concept of individual bioequivalence (IBE) may be applied] (Chow et al., 2011). For highly variable drugs (HVDs), which are defined as drugs with within-subject coefficient of variation (CV) in one or more of the pharmacokinetic parameters 30% or larger (Davit et al., 2012), the TOST has low power with a sample size of 24 or 36 (typically used in BE trials) since the width of CI is proportional to the estimated variability and reciprocally proportional to the square root of the sample size (Berger and Hsu, 1996; Davit et al., 2008). Consequently, the number of subjects required for demonstrating ABE increases dramatically in order to maintain the power. A highly variable reference drug may not be demonstrated to be bioequivalent even to itself in a typical cross-over study with a modest number of subjects (Davit et al., 2012). In general, it is believed that HVDs have wide therapeutic windows and thus, effective and safe. A review of 1,010 BE studies of 180 generic drugs submitted to the FDA during 2003-2005 suggests that 31% (57/180) of those are highly variable in terms of root mean square error (Davit et al., 2008). Therefore, it is necessary to develop an alternative to the TOST for HVDs.

Berger and Hsu (1996) and Brown et al. (1997) have constructed uniformly more powerful tests than the TOST. But those tests are difficult to interpret due to polar coordinates and the rejection region of those tests has the undesirable property of being non-convex (Liu and Chow, 1996). Recently, the FDA working group examined several alternatives and proposed a reference scaled average bioequivalence approach (RSABE), where the BE limits are scaled to the variability of the reference drug for HVDs (i.e., the limits are increasing proportionally to the variance) (Haidar et al., 2008a). The implementation of the FDA's RSABE method for HVDs is summarized in Davit et al. (2012). Either partial replicate (three-way cross-over: RTR, RRT, and TRR) or full replicate (four-way cross-over: RTRT and TRTR) design is required, with minimum subject number of 24 for RSABE. The European Medicines Agency (EMA) recently issued a guideline for BE assessment of HVDs (EMA, 2010), which is also a reference scaled approach. However, the EMA approach differs from the FDA: 1) the EMA RSABE allows only for $C_{max}$, not for AUC; 2) the EMA uses a different scaling factor for RSABE, which is smaller for the reference drug with %CV ranging from 30 to 50% and fixed at a smaller scaling factor for the drug with %CV larger than 50% (Karalis et al., 2012). Both of these scaled approaches have an additional point estimate constraint; that is, the point estimate for GMR has to be contained within [0.8, 1.25].

Even though the power is improved using the reference scaled approaches, the frequentist tests may not guarantee any level of confidence for the true difference being at the equivalence boundaries (for example, GMR = 0.8 and 1.25) (Ghosh and Gonen, 2008; de Souza et al., 2009). There are also debates on the practice

of using the 100(1-2$\alpha$)% CI rather than the test at $\alpha$-level (Berger and Hsu, 1996; Munk and Pfluger, 1999). The root of these problems is the fundamental flaw in the frequentist framework, within which p-values are confused with the strength of evidence and the observed type I error rate (Blume and Peipert, 2003; Choi et al., 2008). This motivated Choi et al. (2008) to advocate the likelihood framework for representing and interpreting BE data as evidence.

In this thesis, we extended the likelihood approach for evaluating BE proposed by Choi et al. (2008) to HVDs in full replicate $2 \times 4$ cross-over studies. We demonstrated how to present evidence for BE and interpret this evidence using the profile likelihood of the parameters of interest, including the mean difference and variance ratios (both in log scale) for pharmacokinetic measures of two drugs. This likelihood approach shows the full spectrum of evidence that supports BE or bioinequivalence (BIE). We may even evaluate the mean and variance together in a unified framework using the likelihood approach. Simulations were used to evaluate the operating characteristics of the likelihood approach when the FDA or EMA RSABE criteria were applied.

# CHAPTER II

## BACKGROUND OF LIKELIHOOD PARADIGM

The law of likelihood says (Hacking, 1965; Royall, 1997):

> If hypothesis $A$ implies that the probability that a random variable $X$ takes the value $x$ is $p_A(x)$, while hypothesis $B$ implies that the probability is $p_B(x)$, then the observation $X = x$ is evidence supporting $A$ over $B$ if and only if $p_A(x) > p_B(x)$, and the likelihood ratio, $p_A(x)/p_B(x)$, measures the strength of that evidence.

The likelihood paradigm, based on the law of likelihood, provides an appropriate framework for representing and interpreting statistical evidence. It is specifically developed to understand "what the data say" (Blume, 2002). The growing interest in the likelihood paradigm is contributed by Royall (1997, 2000) and others including Blume (2002) and Zhang (2009). The likelihood approach has also been applied to practical problems such as clinical trials [see Blume (2008), Wang and Blume (2011) and Choi et al. (2008)]. More recently, Zhang and Zhang (2013) generalized the law of likelihood (GLL) for composite hypotheses in clinical trials, such as those used in BE trials.

**Operating characteristics:**

Suppose we wish to test two simple hypotheses $H_0 : \mu = \mu_0$ versus $H_1 : \mu = \mu_1$, where $\mu$ is the parameter of interest and its likelihood function is $L_n(\mu)$ (indexed with the sample size, $n$). The probabilities $P_0\left[\frac{L_n(\mu_1)}{L_n(\mu_0)} \geq k\right]$ and $P_0\left[\frac{1}{k} < \frac{L_n(\mu_1)}{L_n(\mu_0)} < k\right]$, when the hypothesis $H_0$ is true, are called the probability of observing misleading evidence (the likelihood ratio in favor of wrong hypothesis) and the probability of observing weak evidence (the likelihood ratio in favor of neither of the hypotheses), respectively (where $k > 1$). Note that both the probabilities are functions of sample size $n$ and the threshhold $k$. Although they are closely related to the type I and type II error rates, they are distinct from the type I and type II error rates in the sense that the type I error rate is fixed in hypothesis testing while both these probabilities converge to zero as $n \to \infty$ [see Figure 2 in Blume (2002)].

For normally distributed data with mean $\mu$ and known variance $\sigma^2$, if we want to test $H_0 : \mu = \mu_0$ and $H_1 : \mu = \mu_1$, the probability of observing misleading evidence when the hypothesis $H_0$ is true is:

$$P_0\left[\frac{L_n(\mu_1)}{L_n(\mu_0)} \geq k\right] = \Phi\left[-\frac{|c|\sqrt{n}}{2} - \frac{\log k}{|c|\sqrt{n}}\right], \tag{II.1}$$

where $\Delta = c\sigma = \mu_1 - \mu_0$. When $n$ is held as a constant, this probability is a function of $c$, the distance between the two hypotheses measured in standard deviation. This probability is called the bump function (Royall, 2000), which is generally much smaller than $1/k$. This bound is called the "universal bound", which is a result of applying Markov's inequality on the general form of the probability of observing misleading evidence (Royall, 1997).

The derived probabilities of observing weak evidence and strong evidence when the hypothesis $H_0$ is true are:

$$P_0\left[\frac{1}{k} < \frac{L_n(\mu_1)}{L_n(\mu_0)} < k\right] = \Phi\left[\frac{|c|\sqrt{n}}{2} + \frac{\log k}{|c|\sqrt{n}}\right] - \Phi\left[\frac{|c|\sqrt{n}}{2} - \frac{\log k}{|c|\sqrt{n}}\right] \text{ and} \quad \text{(II.2)}$$

$$P_0\left[\frac{L_n(\mu_0)}{L_n(\mu_1)} \geq k\right] = 1 - \Phi\left[-\frac{|c|\sqrt{n}}{2} + \frac{\log k}{|c|\sqrt{n}}\right]. \quad \text{(II.3)}$$

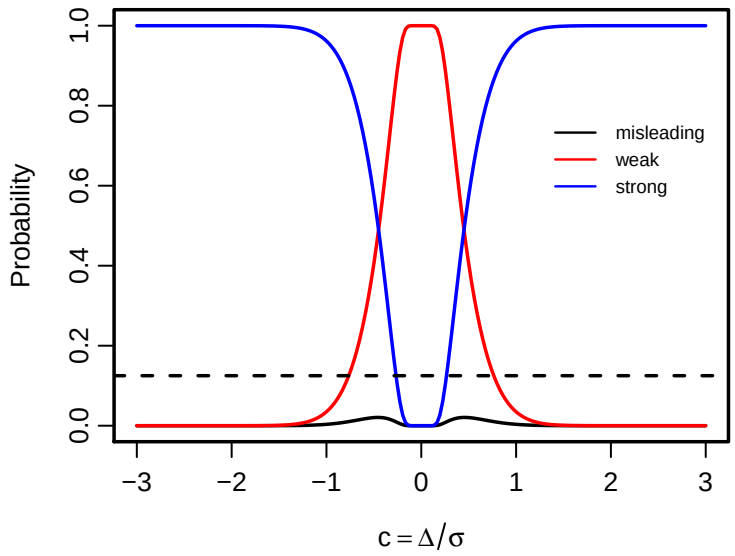


Figure II.1: Probabilities of observing misleading, weak and strong evidence at $n$=20 and $k$=8. The dashed horizontal line represents the universal bound, 1/8.

Figure II.1 shows these probabilities as a function of $c$ when $n$=20 and $k$=8. When the alternative is close to the true hypothesis ($c \cong 0$), the probability of observing weak evidence is close to 1 since the data cannot tell the small difference between the alternative and null hypotheses. The probability of misleading evidence is much smaller than the universal bound, 1/8.

**Nuisance parameters:**

Nuisance parameters are often present in the likelihood function in multi-parameter models (i.e., $\sigma$, if $\sigma$ is unknown in the normal distribution model above). There are several *ad hoc* solutions to eliminate the nuisance parameters, including conditional, marginal, profile and estimated likelihoods (Royall, 1997). The profile

likelihood function for $\mu$ is defined as $\max_{\sigma} L_n(\mu,\sigma) = L_n(\mu,\hat{\sigma}(\mu)) = L_{pn}(\mu)$. The profile likelihood function behaves like a true likelihood function under certain conditions and the limiting probability of observing misleading evidence is given by the bump function (Royall, 2000). In the context of BE analysis, Choi et al. (2008) demonstrated that the profile likelihood is a good approximation to the true likelihood.

**Likelihood intervals and their connection with confidence intervals:**

A standardized likelihood plot, which is a plot of likelihood divided by the maximum likelihood estimate (MLE) as a function of the parameter of interest, has been suggested to represent the data as evidence (Royall, 1997). From this plot, we visualize all the likelihood ratios of any alternative parameter values to the MLE. A horizontal line of $1/k$ on the plot defines a set of parameter values that are consistent with the data at $k$ level, where the standardized likelihoods are $\geq 1/k$. Any parameter values in the likelihood support set (likelihood interval, LI) are supported by the data since the best supported value, MLE, is only better supported at most by a factor of $k$.

For normally distributed data with known variance $\sigma$, the $1/k$ LI for the mean $\mu$ can be derived as:

$$\hat{\mu} \pm \sqrt{2\log k}\sigma/\sqrt{n}. \tag{II.4}$$

On the other hand, the $1-\alpha$ CI for $\mu$ can be written as:

$$\hat{\mu} \pm z_{\alpha/2}\sigma/\sqrt{n}. \tag{II.5}$$

As such, there is a one to one relationship between the LI (II.4) and CI (II.5). Thus, for normally distributed data, the $100(1-\alpha)\%$ CI can be surrogate to a LI with $k = \exp(z^2_{\alpha/2}/2)$; for example, 1/4 and 1/6.8 LIs approximately correspond to 90% and 95% CIs, respectively.

**Simple versus interval hypotheses:**

The likelihood ratio measures the strength of evidence for one simple hypothesis over another. Benchmark values of 8 and 32 for the likelihood ratio has been suggested to define weak, moderate and strong evidences (Royall, 1997). However, in practice, there is great need for testing composite hypotheses in clinical trials, such as the interval hypotheses in BE trials. Zhang and Zhang (2013) proposed a generalization of the law of likelihood using generalized likelihood ratio (GLR) for composite hypotheses in clinical trials. For two composite hypotheses $H_1 : \theta \in \Theta_1$ versus $H_2 : \theta \in \Theta_2$, where both $\Theta_1$ and $\Theta_2 \subset \Theta$, if $\sup L(\Theta_1) > \sup L(\Theta_2)$, then there is evidence supporting $H_1$ over $H_2$ based on the GLL. The GLL is proposed as an evidential tool for clinical trials, and a consequence of GLL [Theorem 3 in Zhang and Zhang (2013)] is that a hypothesis $H$ concerning $\theta$ is supported over its complement at least $k$ if and only if the $1/k$ likelihood interval is contained

in $H$. Choi et al. (2008) considered that the data present evidence at $k$ strength in favor of BE if the entire $1/k$ LI for the mean difference (in log scale) is contained within the ABE limit of [-0.223, 0.223], which is consistent with Zhang's Theorem 3.

Through simulations, Choi et al. investigated the probability of incorrectly presenting evidence supporting ABE using the frequency of $1/k$ LI being completely contained within the ABE limit, when the data are marginally BIE ($\mu_T - \mu_R = -0.223$). This is the probability of observing $k$ strength misleading evidence at specific $\mu_T - \mu_R = -0.223$, which is an analogue of the type I error rate for the hypotheses (I.1). The probability of correctly presenting evidence for ABE was also examined using the frequency of $1/k$ LI being completely contained within the ABE limit when the data are truly ABE ($\mu_T - \mu_R = 0$). This is the probability of observing evidence at $k$ strength level, and it is an analogue of the power at $\mu_T - \mu_R = 0$ for the hypotheses (I.1). Here, failing to present evidence for ABE at $k$ strength implies that either the $1/k$ LI is partially contained in the ABE limit or it is completely out of the limit. Therefore, the probability of failing to present evidence for BE (an analogue of type II error rate) includes the probabilities of observing weak evidence and misleading evidence (both at $k$ strength level).

# CHAPTER III

# METHODS

In a full replicate $2 \times 4$ cross-over study, subjects are randomized to either sequence RTRT or TRTR. Each subject then receives both drug formulations twice in the order of RTRT or TRTR to which the subject is randomized. Between the drug administrations, there is a "washout period" to avoid the possible effect of drug(s) administered in the previous period(s), called "carry-over" effect. We assume there is no "carry-over" effect here.

Let's assume $n_1$ subjects in sequence RTRT and $n_2$ subjects in sequence TRTR ($n = n_1 + n_2$). Let $Y_{ijk}$ be a random variable representing the log transformed response (i.e., $Y = \log \text{AUC}$ or $\log \text{Cmax}$) for subject $i$ in period $j$ of sequence $k$ with $i = 1, \ldots, n_k$, $j = 1,2,3,4$, and $k = 1,2$. The following model for $Y_{ijk}$ is commonly assumed (Berger and Hsu, 1996; Qcana et al., 2008):

$$Y_{ijk} = \mu + P_j + F_{jk} + S_k + \gamma_{ikR} I(\text{drug=R}) + \gamma_{ikT} I(\text{drug=T}) + e_{ijk}, \tag{III.1}$$

where $I(\cdot)$ is an indicator function, $\mu$ is an overall mean, $P_j$ is the fixed effect of $j$th period, and $F_{jk}$ is the fixed effect of drug formulation administered in period $j$ of sequence $k$. That is $F_{11} = F_{31} = F_{22} = F_{42} = F_R$ and $F_{21} = F_{41} = F_{12} = F_{32} = F_T$. Also $S_k$ denotes the fixed effect of $k$th sequence, and $\gamma_{ikR}$ and $\gamma_{ikT}$ are the random-effects of subject $i$ in sequence $k$ for the reference and test drugs, which are assumed to follow a bivariate normal distribution ($N_2$):

$$\boldsymbol{\gamma}_{ik} = (\gamma_{ikR}, \gamma_{ikT})' \sim N_2 \left[ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{BR}^2 & \rho\,\sigma_{BR}\sigma_{BT} \\ \rho\,\sigma_{BR}\sigma_{BT} & \sigma_{BT}^2 \end{pmatrix} \right].$$

The random errors, $e_{ijk}$'s, are independent, and $e_{ijk}$ is assumed to follow a normal distribution ($N$), $N(0, \sigma^2_{\tau(jk)})$. The $\boldsymbol{\gamma}_{ik}$ and $e_{ijk}$ are also independent to each other. We assume $\sigma_{\tau(11)} = \sigma_{\tau(31)} = \sigma_{\tau(22)} = \sigma_{\tau(42)} = \sigma_{WR}$ and $\sigma_{\tau(21)} = \sigma_{\tau(41)} = \sigma_{\tau(12)} = \sigma_{\tau(32)} = \sigma_{WT}$.

## III.1 Estimation of the mean difference, $\phi$, and its standard error

Table III.1 displays the expected means and the observed data in a $2 \times 4$ cross-over study.

Table III.1: Expected means and the observed responses, $y_{ijk}$ (in parentheses), in a $2 \times 4$ full replicate cross-over study.

| | Period | | | |
|---|---|---|---|---|
| Seq. | 1 | 2 | 3 | 4 |
| 1 | $\mu+P_1+F_R+S_1$ $(y_{i11})$ | $\mu+P_2+F_T+S_1$ $(y_{i21})$ | $\mu+P_3+F_R+S_1$ $(y_{i31})$ | $\mu+P_4+F_T+S_1$ $(y_{i41})$ |
| 2 | $\mu+P_1+F_T+S_2$ $(y_{i12})$ | $\mu+P_2+F_R+S_2$ $(y_{i22})$ | $\mu+P_3+F_T+S_2$ $(y_{i32})$ | $\mu+P_4+F_R+S_2$ $(y_{i42})$ |

Let

$$\bar{d}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1}\left(\frac{Y_{i21}+Y_{i41}}{2}-\frac{Y_{i11}+Y_{i31}}{2}\right) \text{ and} \tag{III.2}$$

$$\bar{d}_2 = \frac{1}{n_2}\sum_{i=1}^{n_2}\left(\frac{Y_{i12}+Y_{i32}}{2}-\frac{Y_{i22}+Y_{i42}}{2}\right). \tag{III.3}$$

Then

$$\mathrm{E}[\bar{d}_1] = \phi+\frac{P_2+P_4}{2}-\frac{P_1+P_3}{2} \text{ and} \tag{III.4}$$

$$\mathrm{E}[\bar{d}_2] = \phi+\frac{P_1+P_3}{2}-\frac{P_2+P_4}{2}, \tag{III.5}$$

where $\phi = F_T - F_R$ is the mean difference of the two drugs and it can be estimated by:

$$\hat{\phi} = \frac{\bar{d}_1+\bar{d}_2}{2}, \tag{III.6}$$

which is an unbiased estimator for $\phi$ with complete data. The variance of $\hat{\phi}$ is

$$\begin{aligned}\mathrm{Var}(\hat{\phi}) &= \frac{1}{4}\mathrm{Var}(\bar{d}_1)+\frac{1}{4}\mathrm{Var}(\bar{d}_2)\\ &= \frac{1}{4}\frac{1}{n_1}\mathrm{Var}\left(\frac{Y_{i21}+Y_{i41}}{2}-\frac{Y_{i11}+Y_{i31}}{2}\right)+\frac{1}{4}\frac{1}{n_2}\mathrm{Var}\left(\frac{Y_{i12}+Y_{i32}}{2}-\frac{Y_{i22}+Y_{i42}}{2}\right).\end{aligned} \tag{III.7}$$

The $\mathrm{Var}\left(\frac{Y_{i21}+Y_{i41}}{2}-\frac{Y_{i11}+Y_{i31}}{2}\right) = \mathrm{Var}\left(\frac{Y_{i12}+Y_{i32}}{2}-\frac{Y_{i22}+Y_{i42}}{2}\right)$ can be estimated by the pooled sample variance of average difference within a subject:

$$\begin{aligned}s^2 = &\frac{1}{n_1+n_2-2}\sum_{i=1}^{n_1}\left[\left(\frac{y_{i21}+y_{i41}}{2}-\frac{y_{i11}+y_{i31}}{2}\right)-\left(\frac{\bar{y}_{\cdot 21}+\bar{y}_{\cdot 41}}{2}-\frac{\bar{y}_{\cdot 11}+\bar{y}_{\cdot 31}}{2}\right)\right]^2+\\ &\frac{1}{n_1+n_2-2}\sum_{i=1}^{n_2}\left[\left(\frac{y_{i12}+y_{i32}}{2}-\frac{y_{i22}+y_{i42}}{2}\right)-\left(\frac{\bar{y}_{\cdot 12}+\bar{y}_{\cdot 32}}{2}-\frac{\bar{y}_{\cdot 22}+\bar{y}_{\cdot 42}}{2}\right)\right]^2,\end{aligned} \tag{III.8}$$

where $\bar{y}_{\cdot jk}$ denotes the sample mean of $j$ period of sequence $k$ (each cell of Table III.1).

Therefore, the estimated standard error for $\hat{\phi}$ is:

$$\widehat{\mathrm{SE}\hat{\phi}} = \frac{1}{2} s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}. \tag{III.9}$$

In addition, $\sigma^2_{WR} = \frac{\mathrm{Var}(Y_{i11} - Y_{i31})}{2} = \frac{\mathrm{Var}(Y_{i22} - Y_{i42})}{2}$ can be estimated by $s^2_{WR}$:

$$s^2_{WR} = \frac{1}{2(n_1+n_2-2)} \left\{ \sum_{i=1}^{n_1} \left[ (y_{i11} - y_{i31}) - (\bar{y}_{\cdot 11} - \bar{y}_{\cdot 31}) \right]^2 + \sum_{i=1}^{n_2} \left[ (y_{i22} - y_{i42}) - (\bar{y}_{\cdot 22} - \bar{y}_{\cdot 42}) \right]^2 \right\} \tag{III.10}$$

Similarly, $\sigma^2_{WT} = \frac{\mathrm{Var}(Y_{i21} - Y_{i41})}{2} = \frac{\mathrm{Var}(Y_{i12} - Y_{i32})}{2}$ can be estimated by $s^2_{WT}$:

$$s^2_{WT} = \frac{1}{2(n_1+n_2-2)} \left\{ \sum_{i=1}^{n_1} \left[ (y_{i21} - y_{i41}) - (\bar{y}_{\cdot 21} - \bar{y}_{\cdot 41}) \right]^2 + \sum_{i=1}^{n_2} \left[ (y_{i12} - y_{i32}) - (\bar{y}_{\cdot 12} - \bar{y}_{\cdot 32}) \right]^2 \right\} \tag{III.11}$$

The equations (III.6), and (III.9) to (III.11) are the simple moment estimators that can be obtained using a random-effects model with a random intercept as well.

### III.2 FDA and EMA's RSABEs for HVDs

In practice, the FDA recommends implementing RSABE for HVDs (Davit et al., 2012; FDA, 2011). If the observed within-subject variability of the reference drug, $s_{WR}$, for log AUC or log Cmax is equal or greater than 0.294, the reference scaling method may be used. Otherwise, the original ABE analysis (i.e., TOST) must be used. The value 0.294 is determined using the conversion formula of $\sigma^2 = \log(CV^2 + 1)$ when CV=30%. There is no penalty if the applicant uses a partial or full replicate design with the intent of using RSABE.

The criterion for RSABE is $\frac{(\mu_T - \mu_R)^2}{\sigma^2_{WR}} \leq \theta_S$, when $s_{WR}$ for a pharmacokinetic parameter is equal or greater than 0.294. Equivalently, the upper 95% confidence bound for $(\mu_T - \mu_R)^2 - \theta_S \sigma^2_{WR}$ must be $\leq 0$. Appendix A lists the Howe method (Howe, 1974) to determine this criterion bound. In addition, the point estimate for GMR must fall within [0.8, 1.25]. Here, $\theta_S = (\log 1.25)^2 / \sigma^2_{W0}$ and $\sigma_{W0} = 0.25$, yielding $\theta_S = 0.892^2$.

The reference scaled EMA BE limits for HVDs are required only for log Cmax, but for log AUC, the unscaled ABE limit of [-0.223, 0.223] must be used. When $s_{WR}$ is equal or larger than 0.294 (i.e., %CV=30%) but less than 0.472 (i.e., %CV=50%), the 90% CI for log Cmax is required to fall completely within the limit of [$-0.76 s_{WR}$, $0.76 s_{WR}$] to claim BE; when $s_{WR}$ is equal or larger than 0.472, a fixed limit of [-0.359, 0.359] ($0.359 = 0.76 \times 0.472$) should be used. EMA approach requires the point estimate constraint as well.

### III.3 Likelihood function and likelihood approach

Model (III.1) can be written as:

$$\begin{aligned} Y_{ijk} &= \mu + P_2^* I(j=2) + P_3^* I(j=3) + P_4^* I(j=4) + S^* I(k=2) + \phi I(\text{drug=T}) \\ &\quad + \gamma_{ikR} I(\text{drug=R}) + \gamma_{ikT} I(\text{drug=T}) + e_{ijk}. \end{aligned} \tag{III.12}$$

In vector/matrix notation, (III.12) becomes:

$$\mathbf{Y}_{ik} = \mathbf{X}_{ik}\boldsymbol{\beta} + \boldsymbol{\gamma}_{ik}^* + \mathbf{e}_{ik}, \tag{III.13}$$

where $\mathbf{Y}_{ik} = (Y_{i1k}, Y_{i2k}, Y_{i3k}, Y_{i4k})'$; $\boldsymbol{\beta} = (\mu, P_2^*, P_3^*, P_4^*, S^*, \phi)'$; $\boldsymbol{\gamma}_{i1}^* = (\gamma_{i1R}, \gamma_{i1T}, \gamma_{i1R}, \gamma_{i1T})'$ and $\boldsymbol{\gamma}_{i2}^* = (\gamma_{i2T}, \gamma_{i2R}, \gamma_{i2T}, \gamma_{i2R})'$.

$\mathbf{X}_{i1} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 \end{pmatrix}$ and $\mathbf{X}_{i2} = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 \end{pmatrix}$ are the design matrices.

We assume that both $\boldsymbol{\gamma}_{ik}^*$ and $\mathbf{e}_{ik}$ are normally distributed with mean $\mathbf{0}$. Their variance and covariance matrices are: $\text{Cov}(\boldsymbol{\gamma}_{i1}^*) = \begin{pmatrix} \sigma_{BR}^2 & \rho\sigma_{BR}\sigma_{BT} & \sigma_{BR}^2 & \rho\sigma_{BR}\sigma_{BT} \\ \rho\sigma_{BR}\sigma_{BT} & \sigma_{BT}^2 & \rho\sigma_{BR}\sigma_{BT} & \sigma_{BT}^2 \\ \sigma_{BR}^2 & \rho\sigma_{BR}\sigma_{BT} & \sigma_{BR}^2 & \rho\sigma_{BR}\sigma_{BT} \\ \rho\sigma_{BR}\sigma_{BT} & \sigma_{BT}^2 & \rho\sigma_{BR}\sigma_{BT} & \sigma_{BT}^2 \end{pmatrix}$ and $\text{Cov}(\mathbf{e}_{i1}) = \begin{pmatrix} \sigma_{WR}^2 & 0 & 0 & 0 \\ 0 & \sigma_{WT}^2 & 0 & 0 \\ 0 & 0 & \sigma_{WR}^2 & 0 \\ 0 & 0 & 0 & \sigma_{WT}^2 \end{pmatrix}$,

respectively, and for $\text{Cov}(\boldsymbol{\gamma}_{i2}^*)$ and $\text{Cov}(\mathbf{e}_{i2})$, the $R$ and $T$ subscripts should be swapped.

The likelihood function for $\boldsymbol{\beta}$ based on model (III.12) is:

$$L(\boldsymbol{\beta};\mathbf{y}) = \prod_{k=1}^{2}\prod_{i=1}^{n_k} \frac{1}{(\sqrt{2\pi})^4 |\mathbf{V}_{ik}|^{1/2}} \exp[-(\mathbf{y}_{ik} - \mathbf{X}_{ik}\boldsymbol{\beta})' \mathbf{V}_{ik}^{-1} (\mathbf{y}_{ik} - \mathbf{X}_{ik}\boldsymbol{\beta})/2], \tag{III.14}$$

where $\mathbf{V_{i1}} = \begin{pmatrix} \sigma_{BR}^2 + \sigma_{WR}^2 & \rho\sigma_{BR}\sigma_{BT} & \sigma_{BR}^2 & \rho\sigma_{BR}\sigma_{BT} \\ \rho\sigma_{BR}\sigma_{BT} & \sigma_{BT}^2 + \sigma_{WT}^2 & \rho\sigma_{BR}\sigma_{BT} & \sigma_{BT}^2 \\ \sigma_{BR}^2 & \rho\sigma_{BR}\sigma_{BT} & \sigma_{BR}^2 + \sigma_{WR}^2 & \rho\sigma_{BR}\sigma_{BT} \\ \rho\sigma_{BR}\sigma_{BT} & \sigma_{BT}^2 & \rho\sigma_{BR}\sigma_{BT} & \sigma_{BT}^2 + \sigma_{WT}^2 \end{pmatrix}$

$$\text{and } \mathbf{V_{i2}} = \begin{pmatrix} \sigma_{BT}^2 + \sigma_{WT}^2 & \rho\sigma_{BR}\sigma_{BT} & \sigma_{BT}^2 & \rho\sigma_{BR}\sigma_{BT} \\ \rho\sigma_{BR}\sigma_{BT} & \sigma_{BR}^2 + \sigma_{WR}^2 & \rho\sigma_{BR}\sigma_{BT} & \sigma_{BR}^2 \\ \sigma_{BT}^2 & \rho\sigma_{BR}\sigma_{BT} & \sigma_{BT}^2 + \sigma_{WT}^2 & \rho\sigma_{BR}\sigma_{BT} \\ \rho\sigma_{BR}\sigma_{BT} & \sigma_{BR}^2 & \rho\sigma_{BR}\sigma_{BT} & \sigma_{BR}^2 + \sigma_{WR}^2 \end{pmatrix}.$$

The profile likelihood function for $\phi$ will be:

$$L_p(\phi;\mathbf{y}) = L(\phi,\hat{\boldsymbol{\beta}}_\phi;\mathbf{y}) = \max_{\boldsymbol{\beta}_\phi} L(\phi,\boldsymbol{\beta}_\phi;\mathbf{y}) \quad \text{for fixed } \phi, \tag{III.15}$$

where $\hat{\boldsymbol{\beta}}_\phi = \text{argmax}[L_\phi(\boldsymbol{\beta}_\phi;\mathbf{y})]$ at fixed $\phi$, and $\boldsymbol{\beta}_\phi$ denotes all $\beta$'s except for $\phi$. Utilizing the `optim()` function in R, we obtained the profile likelihoods for the mean difference $\phi$ and the variance ratios for an example of $2 \times 4$ full replicate BE data, which are presented in CHAPTER IV.

According to Choi et al. (2008), if the $1/k$ LI for $\phi$ lies completely within the ABE limit of [-0.223, 0.223], the data support ABE; the larger the k, the greater the strength of evidence. We used the largest $k$, $k_{\max}$, at which the $1/k$ LI lies completely within the BE limit (for the mean or variance), as the evidence for BE in the example. The larger the $k_{\max}$, the stronger the evidence. We also used Zhang's generalized likelihood ratio, GLR = $\sup_{H_1} L_n(\phi) / \sup_{H_0} L_n(\phi)$ to provide evidence for BE or BIE, where $H_1$ and $H_0$ are the interval hypotheses (I.1); the larger this GLR, the stronger the evidence for BE. In fact, when GLR $> 1$, it is exactly the $k_{\max}$. When GLR $< 1$, there is no such $k_{\max}$ exists.

### III.4 Simulations

We used simulations to evaluate the operating characteristics of the likelihood approach as shown in Choi et al. (2008). We calculated the power as the percentage of simulations that present evidence for ABE. The data present evidence for ABE if the $1/k$ profile LI for the mean difference fall completely within the BE limit (unscaled or scaled), with or without the point estimate constraint.

In the simulations, we generated data from the random-effects model (III.12) assuming that:

1. Equal number of subjects in each sequence (RTRT or TRTR) and complete data for each subject;
2. There is no period and sequence fixed effects;
3. Random-effects $\gamma_{i1R} = \gamma_{i1T}$ and $\gamma_{i2R} = \gamma_{i2T}$, thus $\rho = 1$; and the between-subject variances $\sigma_{BR}^2 = \sigma_{BT}^2 = 0.04$;
4. Other parameter values: $\mu = 5$; the total sample size $n$ = 24, 36, and 72; within-subject standard deviation $\sigma_{WR} = 0.198, 0.294, 0.385, 0.472, 0.631$ corresponding to %CV= 20, 30, 40, 50, 70%, respectively;

and $\sigma_{WT}/\sigma_{WR}$ = 0.7, 1.0 and 1.3. The true test/reference GMRs are 1.0, 1.1, 1.15, 1.2, 1.25, 1.3, 1.4 and 1.6.

For each simulated data, we applied 4 approaches to evaluate BE, including ABE (TOST), FDA RSABE, EMA RSABE, and the likelihood approach. For the likelihood approach, we calculated the 1/8, 1/6.8 and 1/4 profile LIs for the mean difference $\phi$. According to the $s_{WR}$, we evaluated whether these intervals lie within the corresponding FDA or EMA RSABE limits (with or without point estimate constraint). The powers for different approaches under different scenarios were plotted and compared.

Due to the relatively slow computaion speed of `optim()` function, we used an alternative method (`lme()` function, Appendix B) to obtain the profile likelihood for the mean difference and the corresponding profile LIs in the simulations. We confirmed that the profile likelihoods from the two computing methods using the example data were identical (Figure B.1).

## CHAPTER IV

## EXAMPLE FOR 2 × 4 CROSS-OVER BIOEQUIVALENCE DATA ANALYSIS

Example data are taken from a full replicate 2 × 4 cross-over design where sequence 1 is RTRT and sequence 2 is TRTR [modified from an example data in Chapter 4 in Patterson and Jones (2005) with only complete data for 27 subjects for each sequence]. The BEs for both AUC and Cmax were evaluated using the frequentist and likelihood approaches. For the frequentist approach, the ABE was evaluated using TOST, FDA and EMA RSABE methods. For the likelihood approach, in addition to ABE, PBE and IBE were also evaluated using the total variances for the reference ($\sigma^2_{TR} = \sigma^2_{BR} + \sigma^2_{WR}$) and the test ($\sigma^2_{TT} = \sigma^2_{BT} + \sigma^2_{WT}$) drugs, and the within-subject variances for the reference ($\sigma^2_{WR}$) and the test ($\sigma^2_{WT}$) drugs.

### IV.1 TOST, FDA and EMA RSABEs

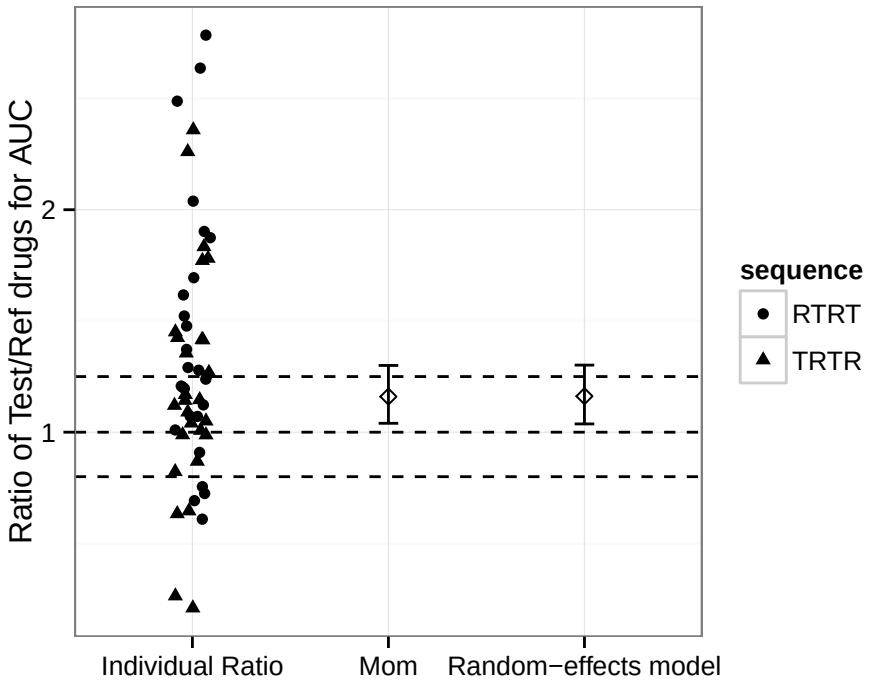


Figure IV.1: The ratio of the AUC means of the test to the reference drugs for each subject, and the GMR estimates with the 90% CIs from the moment based method (Mom) and the random-effects model. The dashed horizontal lines represent the line of 1.0, and the ABE limits of 0.8 and 1.25.

Figures IV.1 and IV.2 present the ratios of the means of the test to the reference drugs for AUC and Cmax for each individual, the estimated GMRs and the 90% CIs using the moment based method and the

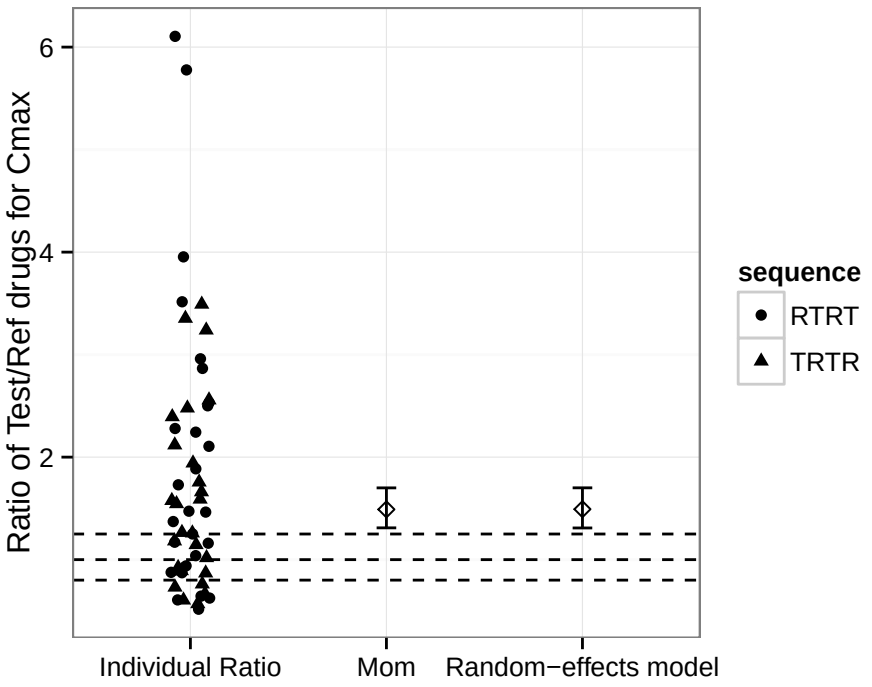


Figure IV.2: The ratio of the Cmax means of the test to the reference drugs for each subject, and the GMR estimates with the 90% CIs from the moment based method (Mom) and the random-effects model. The dashed horizontal lines represent the line of 1.0, and the ABE limits of 0.8 and 1.25.

Table IV.1: FDA RSABE and EMA RSABE results for the example data.

| | PK Parameter | GMR (90% CI) | $s_{WR}$ | BE limits | criteria bound | BE result |
|---|---|---|---|---|---|---|
| FDA | AUC | 1.16 (1.04, 1.3) | 0.337 | [0.74, 1.35] | -0.016 | passed |
| | Cmax | 1.49 (1.31, 1.7) | 0.546 | [0.61, 1.63] | 0.061 | failed |
| EMA | AUC | 1.16 (1.04, 1.3) | 0.337 | [0.8, 1.25] | | failed |
| | Cmax | 1.49 (1.31, 1.7) | 0.546 | [0.7, 1.43] | | failed |

random-effects model. The estimates from the moment based method are almost same as those from the random-effects model. If the conventional BE test method is to be used, we would fail to conclude ABE for AUC and Cmax since neither of the 90% CIs fall completely within the [0.8, 1.25] limit. The individual Cmax ratios show that there are two extreme values (Figure IV.2) from sequence RTRT. They may be influential, but we do not study their influence here.

If the RSABE is used with consideration of the within-subject variability, the conclusion may be different, as summarized in Table IV.1. The reference drug is indeed highly variable for both AUC and Cmax with $s_{WR}$ being equal to 0.337 and 0.546, respectively.

According to the FDA criteria, we would conclude BE for AUC since the 95% upper bound (criteria

bound) for AUC is less than 0 and the GMR point estimate, 1.16, is within the range of [0.8, 1.25]. But for Cmax, the criteria bound is larger than 0 and the GMR point estimate, 1.49, is out of the range of [0.8, 1.25], and hence we failed to demonstrate BE for Cmax. The overall conclusion is that the test drug failed to demonstrate BE to its reference drug.

For EMA, the RSABE is only allowed for Cmax and the BE limit for AUC is always [0.8, 1.25] regardless of $s_{WR}$. Thus, we failed to demonstrate BE for AUC. For Cmax, the scaled limit is expanded to [$\exp(-0.359)$, $\exp(0.359)$], or [0.7, 1.43], since $s_{WR}$ is larger than 0.472 (%CV=50%). Neither the 90% CI nor the GMR point estimate of Cmax is completely contained within their corresponding limits, failing to conclude BE for Cmax as well.

Just for comparison with the FDA RSABE for AUC, if we applied the EMA scaled limits to AUC, which are $\exp(-0.76 \times 0.337) = 0.77$ and $\exp(0.76 \times 0.337) = 1.29$, we would still not conclude BE since the 90% CI is not completely contained within these limits. Thus, BE is concluded for AUC according to FDA RSABE criteria, but we failed to conclude BE according to EMA RSABE criteria. This is because the FDA RSABE is more permissive due to the larger scaling factor. Note that in frequentist framework, if the data fail to show BE, it does not imply BIE.

### IV.2 Likelihood approach

#### IV.2.1 Profile likelihoods for log AUC

In order to evaluate PBE and IBE, we reparameterized the total variances and the within-subject variances as the ratio of their standard deviations of the test to the reference drugs, $\sigma_{TT}/\sigma_{TR}$ and $\sigma_{WT}/\sigma_{WR}$, in the likelihood function (III.14). The profile likelihoods for $\phi$, $\sigma_{TT}/\sigma_{TR}$ and $\sigma_{WT}/\sigma_{WR}$ for log AUC were obtained, which are displayed in Figures IV.3 to IV.5.

Figure IV.3 shows that the MLE for the mean difference (in log scale) is 0.149 [GMR = $\exp(0.149) = 1.16$], which is the same as the moment based estimate. The 1/8 LI is only partially contained within the ABE limit [-0.223, 0.223]. The data fail to present evidence for ABE at $k = 8$ strength level according to Choi et al. (2008). We identified the $k_{\max}$ is 1.85, which can be used as the strength of evidence for ABE for log AUC. If we applied the GLL for composite hypotheses (Zhang and Zhang, 2013), the GLR = $k_{\max}$= 1.85, which indicates weak evidence for ABE.

Similarly, we calculated the $k_{\max}$ = GLR = 3.43 from the profile likelihood for $\sigma_{TT}/\sigma_{TR}$ in Figure IV.4, demonstrating weak evidence for BE. However, for the within-subject standard deviation ratio, $\sigma_{WT}/\sigma_{WR}$, as shown in Figure IV.5, there is little evidence for BE since the MLE is slightly out of the limit, none of the $1/k$ LIs are within the limit, and Zhang's GLR is close to 1, indicating almost equal evidence for BE and BIE. Here we assume the BE limit for the standard deviation ratio is [0.7, 1.3], but note that no explicit BE criteria

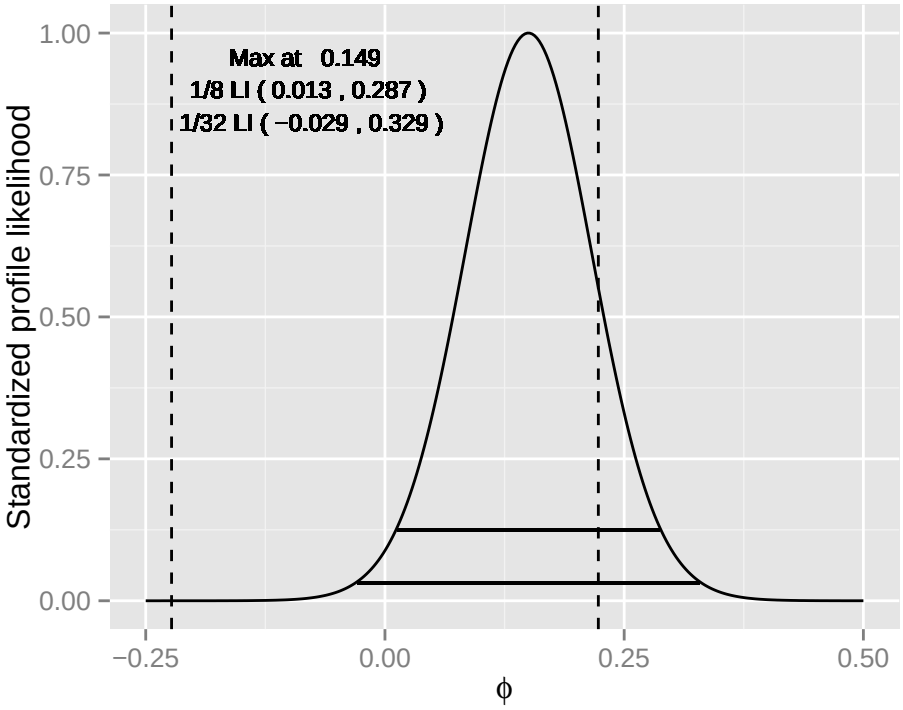


Figure IV.3: The profile likelihood for the mean difference of the test and reference drugs for log AUC. The dashed vertical lines represent the ABE limits of -0.223 and 0.223.

has been suggested for this quantity by any regulatory authorities in any countries.

Within the likelihood paradigm, it is easy to evaluate evidence for BE for the mean and variance together to address the PBE or IBE issues. For AUC, there is weak evidence for PBE when we combine the evidence for $\phi$ and $\sigma_{TT}/\sigma_{TR}$ together. If both evidence were strong, the evidence for PBE would be strong. There is little evidence for IBE when we consider the evidence for $\phi$ and $\sigma_{WT}/\sigma_{WR}$ together. The evidence for PBE and IBE for AUC is summarized in Table IV.2.

### IV.2.2 Profile likelihoods for log Cmax

Figures IV.6 to IV.8 display the profile likelihoods of $\phi$, $\sigma_{TT}/\sigma_{TR}$ and $\sigma_{WT}/\sigma_{WR}$ for log Cmax. None of $1/k$ LIs for $\phi$ is completely contained within the ABE limit of [-0.223, 0.223] (Figure IV.6), indicating little evidence for BE for Cmax. Note that the MLE is also out of the limit, and hence a $k_{\max}$ cannot be determined. Zhang's GLR in this case is 0.08, suggesting moderate evidence for supporting BIE (1/0.08=12.5). Again to evaluate PBE or IBE for Cmax, we may combine the evidence from Figures IV.6 and IV.7, or evidence from Figures IV.6 and IV.8, which is summarized in Table IV.2.

Overall, we may conclude that the data present only weak evidence for BE. However, with the likelihood

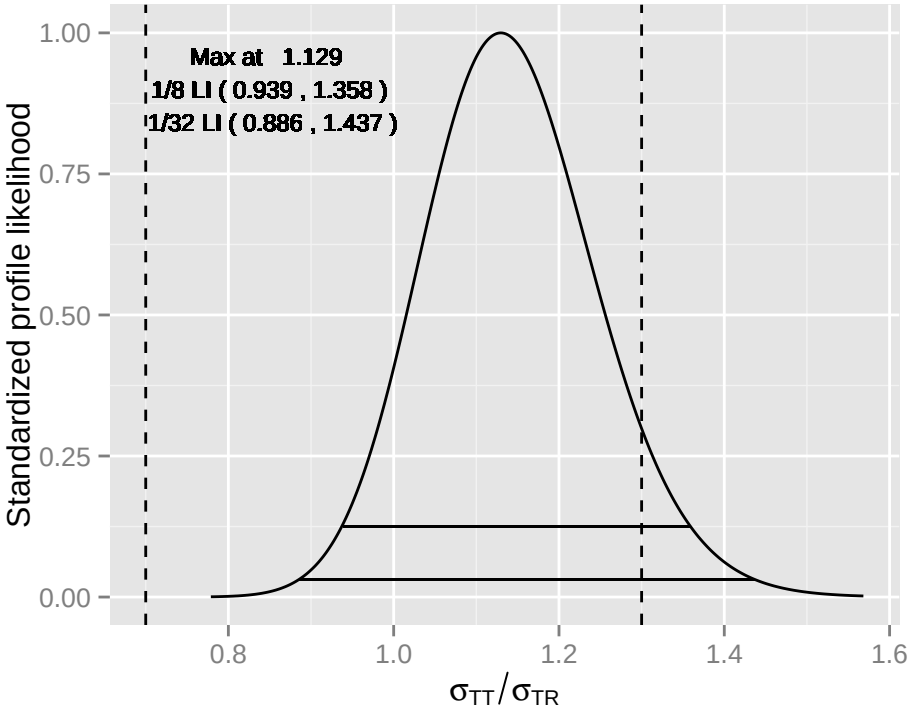


Figure IV.4: The profile likelihood for the ratio of total standard deviations of test/reference drugs for log AUC. The dashed vertical lines represent the limits of 0.7 and 1.3.

approach, we also present the specific strength level and direction of the evidence for each parameter. We even evaluate BE for the variance and mean together within a unified framework. For highly variable drugs, the LI will be wider due to larger variance, and consequently attenuate evidence for ABE. However, with sample size increase, the evidence will be stronger if the two drugs are truly ABE. Likelihood approach may not reduce sample size. A benefit of using the likelihood approach is that if more data are collected, we can combine them with the existing data to re-evaluate the strength of evidence. There is no need to adjust p-values in contrast to the methods in the frequentist framework. It is impossible to present evidence for BE or BIE using FDA or EMA's RSABE approaches. It is also difficult to evaluate BE for the variance in the frequentist framework.

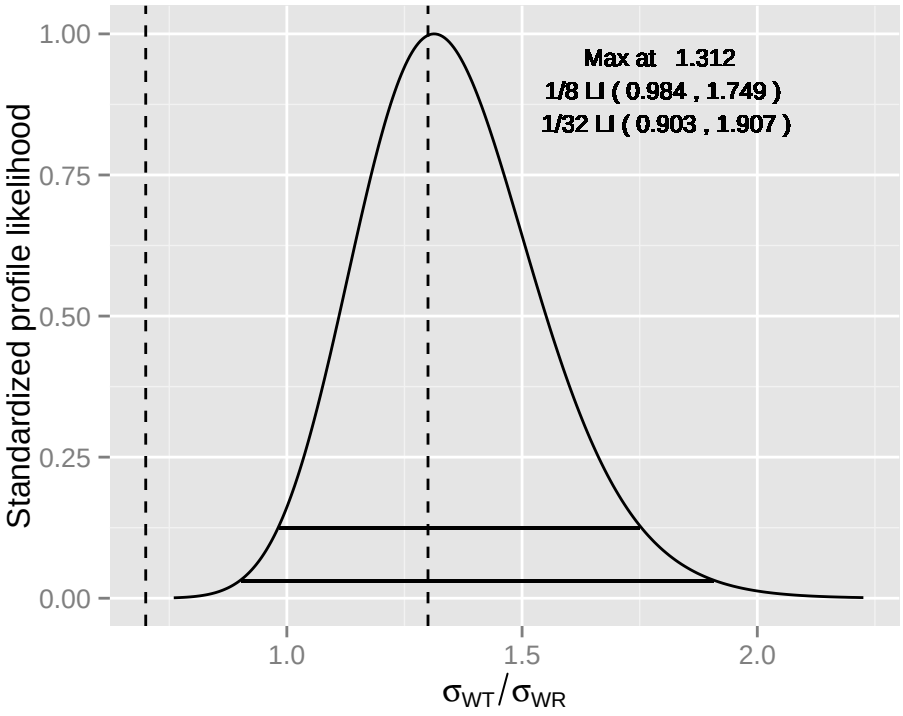


Figure IV.5: The profile likelihood for the ratio of within-subject standard deviations of test/reference drugs for log AUC. The dashed vertical lines represent the limits of 0.7 and 1.3.

Table IV.2: Evidence for BE using the profile likelihood approach for the example data.

| PK parameter | Profile likelihood | $k_{\max}$ | GLR | interpretation |
|---|---|---|---|---|
| AUC | $\phi$ | 1.85 | 1.85 | weak evidence for BE |
| | $\sigma_{TT}/\sigma_{TR}$ | 3.43 | 3.43 | weak evidence for BE |
| | $\sigma_{WT}/\sigma_{WR}$ | | 1 | equal evidence for BE and BIE |
| Cmax | $\phi$ | | 0.08 | moderate evidence for BIE |
| | $\sigma_{TT}/\sigma_{TR}$ | 8.68 | 8.68 | moderate evidence for BE |
| | $\sigma_{WT}/\sigma_{WR}$ | 9.51 | 9.51 | moderate evidence for BE |

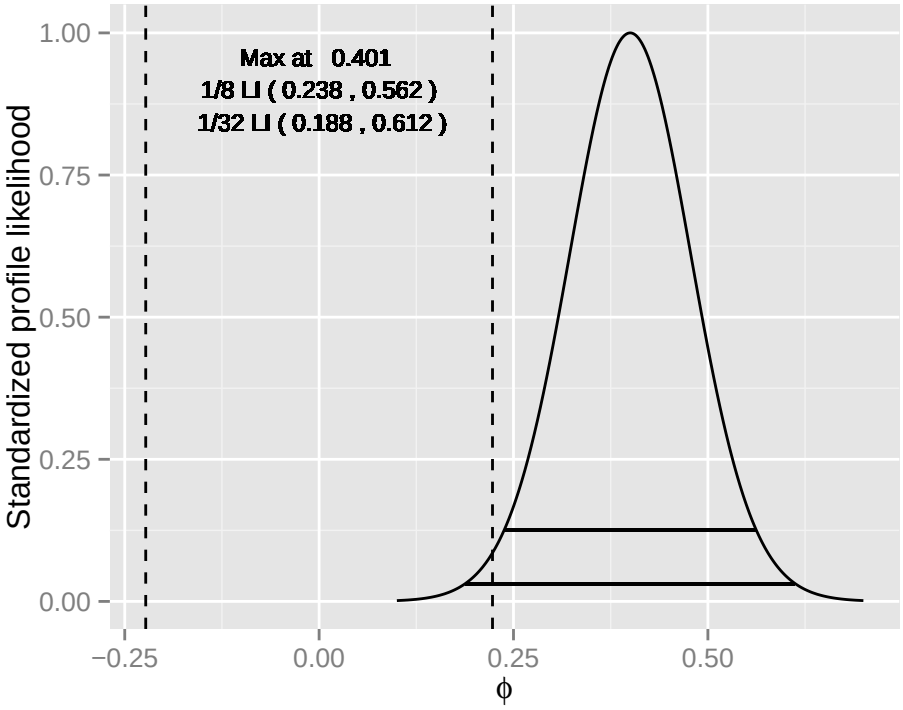


Figure IV.6: The profile likelihood for the mean difference of the test and reference drugs for $\log C_{max}$. The dashed vertical lines represent the ABE limits of -0.223 and 0.223.

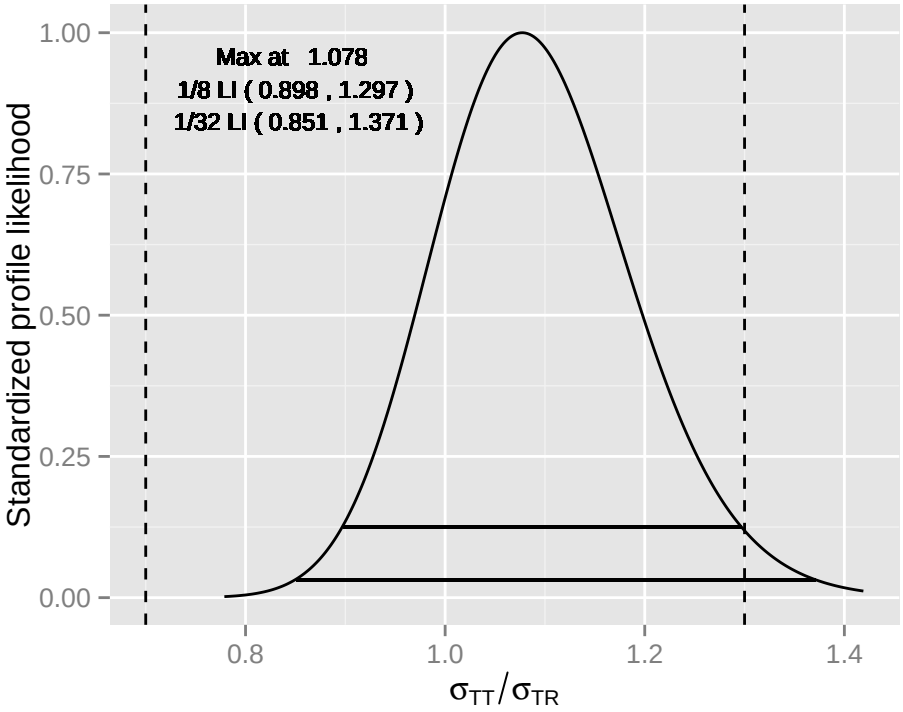


Figure IV.7: The profile likelihood for the ratio of total standard deviations of test/reference drugs for log Cmax. The dashed vertical lines represent the limits of 0.7 and 1.3.

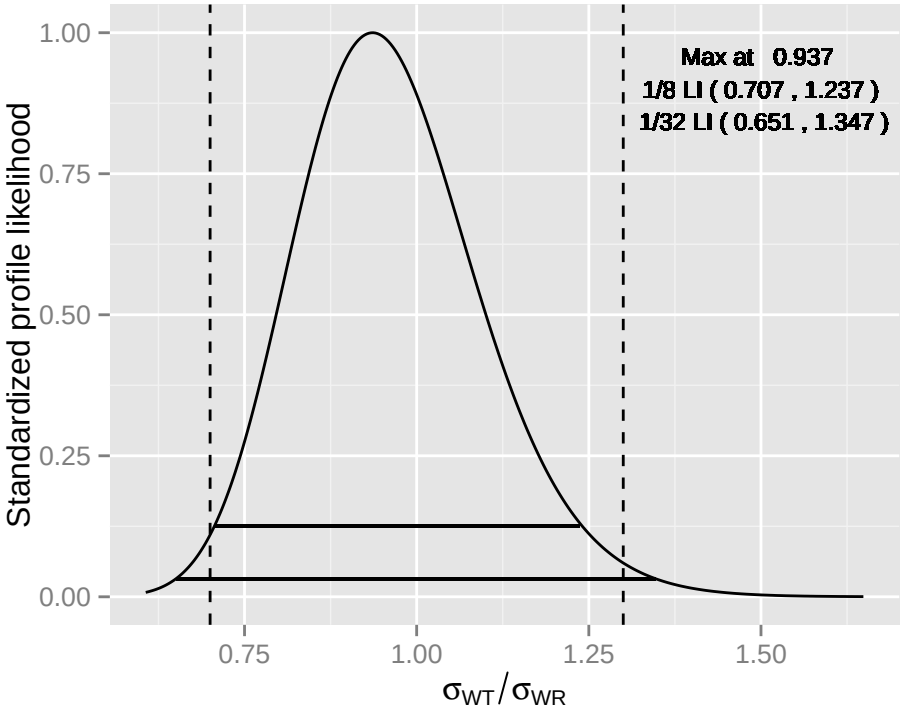


Figure IV.8: The profile likelihood for the ratio of within-subject standard deviations of test/reference drugs for log Cmax. The dashed vertical lines represent the limits of 0.7 and 1.3.

# CHAPTER V

## SIMULATION RESULTS

We performed simulations to evaluate the operating characteristics of the likelihood approach and to compare them with the FDA and EMA RSABE approaches. We investigated the effects of within-subject variabilities of the reference and test drugs (the ratio between those variabilities), and the sample size. For each simulation, the 1/8, 1/6.8 and 1/4 LIs from the profile likelihood of $\phi$ were obtained. Based on the $s_{WR}$, we compared these intervals with the corresponding FDA or EMA RSABE limits. If they fall within the limits (and the MLE is contained as well in the case of using the point estimate constraint), we conclude BE. The percentage of simulations that demonstrated BE was calculated from 1000 simulations at each set of parameters. The power curves shown in Figures V.1 to V.3 are the results of ABE and RSABEs with the point estimate constraint, and the power curves shown in Figures C.1 to C.3 are the results of ABE and RSABEs without that constraint. The red and blue dot-dash vertical lines in these figures represent the FDA and EMA RSABE limits for the corresponding %CV. The black-dashed horizontal line is drawn at 0.05 in each plot to show the type I error rate. The FDA RSABE limit for %CV=70%, 1.76, is out of the range of $x$-axis in these figures, and hence it is not shown.

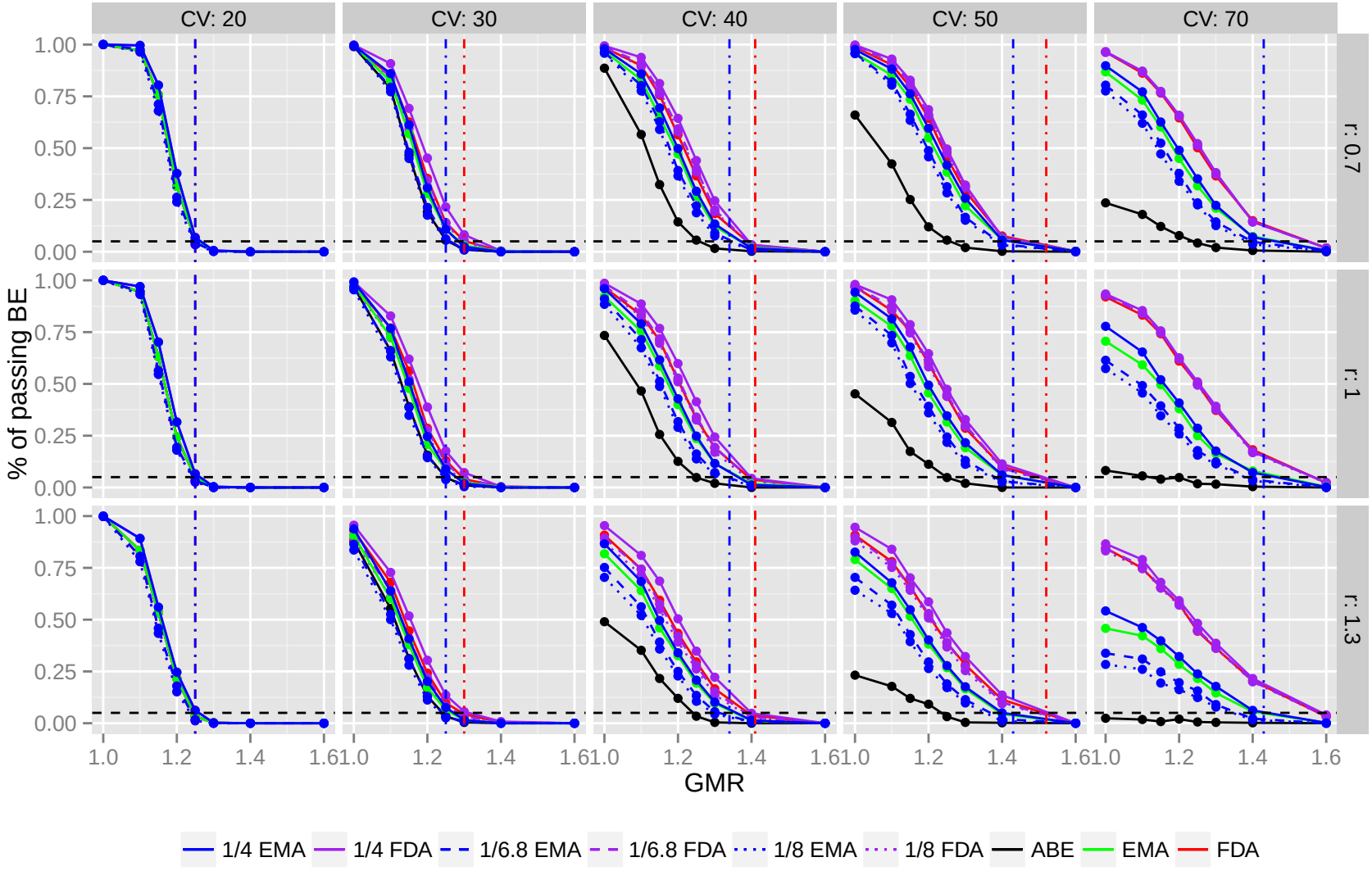


Figure V.1: Power curves ($n$=24) using different methods (with the point estimate constraint when applicable) under different scenarios for $\sigma_{WR}$ and $\sigma_{WT}/\sigma_{WR}$.

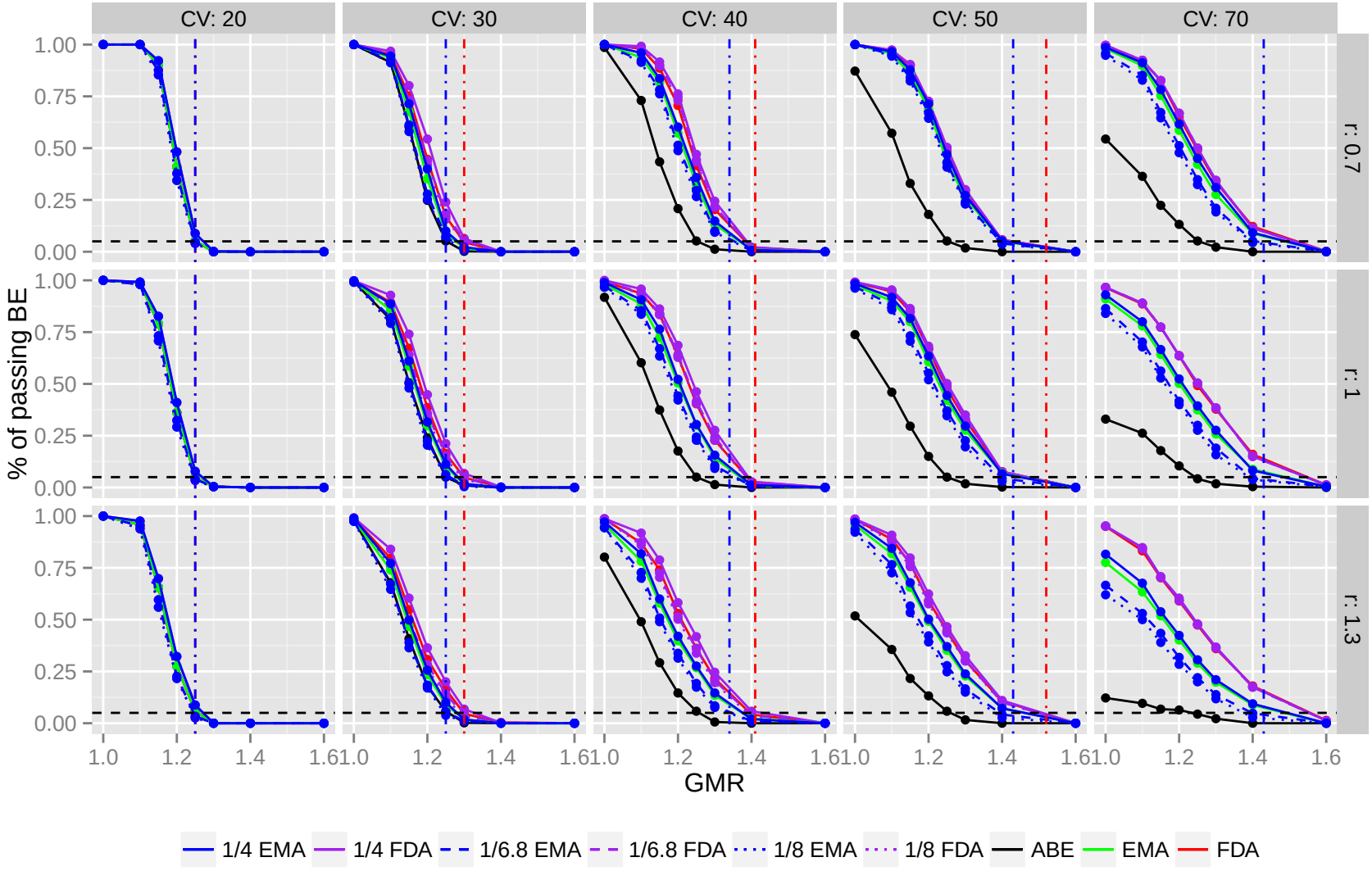


Figure V.2: Power curves ($n$=36) using different methods (with the point estimate constraint when applicable) under different scenarios for $\sigma_{WR}$ and $\sigma_{WT}/\sigma_{WR}$.

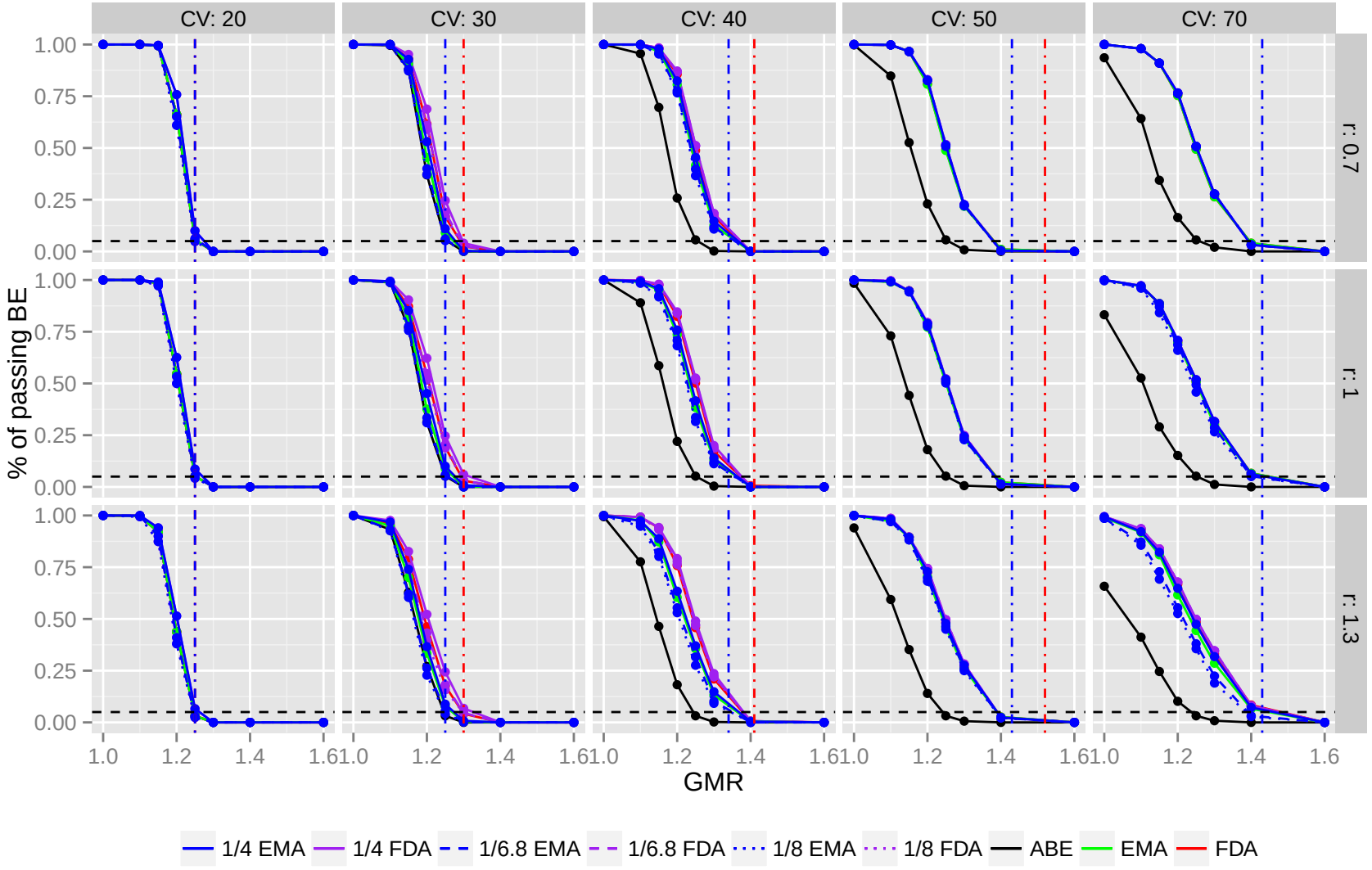


Figure V.3: Power curves ($n$=72) using different methods (with the point estimate constraint when applicable) under different scenarios for $\sigma_{WR}$ and $\sigma_{WT}/\sigma_{WR}$.

**Simulation results for RSABEs with the point estimate constraint:**

When the within-subject variability of the reference drug is small (i.e., 20% CV), the powers of ABE, FDA RSABE, EMA RSABE, and the LI methods are close to each other, especially when the sample size is large. These results are reasonable since CV $<$ 30% (RSABEs are basicly the ABE), and under the normal model the LIs for $k$ = 4, 6.8 and 8 approximately correspond to the frequentist 90, 95 and 96% CIs respectively.

When the sample size is small ($n$ = 24 and 36), the larger the %CV of the reference drug and $\sigma_{WT}/\sigma_{WR}$, the larger the differences between those methods that use FDA RSABE limits and the methods that use EMA RSABE limits. When the sample size is large ($n = 72$) and %CV is large (50-70%), these differences disappear. This reflects the basic difference between FDA RSABE and EMA RSABE methods as shown by Karalis et al. (2012).

The comparison between FDA RSABE and the LI methods using the FDA RSABE limits show that, when the sample size is small ($n = 24$) and %CV is between 30-50%, the powers of 1/4 LI are slightly higher than the FDA RSABE. This difference becomes smaller with the increase in sample size and %CV, and even disappears at %CV=70 for $n = 24$. The power of FDA RSABE is about same as the 1/6.8 and 1/8 LIs. We used the 95% criteria bound along with the point estimate constraint to determine BE for the FDA RSABE here. This may make the power of the FDA RSABE closer to the 1/6.8 LI (corresponding to 95% CI) than to the 1/4 LI (about 90% CI) at intermediate range of %CV and small sample size. For higher %CV (i.e., 70%), the point estimate constraint in the FDA RSABE and the related LI methods is very important and determines the power, which can be seen by comparing Figures V.1 to V.3 with Figures C.1 to C.3. Therefore, the FDA RSABE and related LI methods have almost same power regardless of the $k$ , the sample size and $\sigma_{WT}/\sigma_{WR}$.

Indeed, the comparison between the EMA RSABE and LI methods using the EMA RSABE limits show that the power of the 1/4 LI is closer to the EMA RSABE (which uses the 90% CI), and it is almost the same as the EMA RSABE when the sample size is $\geq 36$, regardless of the $\sigma_{WR}$ and $\sigma_{WT}/\sigma_{WR}$.

As expected, the larger $k$ for the LIs, the smaller the power. In addition, we observe that the $\sigma_{WT}/\sigma_{WR}$ affects the power for all approaches, especially when the %CV of the reference drug is large. Notice that the current FDA and EMA RSABE approaches do not consider the variability of the test drug, which would be problematic.

More interestingly, there is a trade-off between the power and the type I error rate. When the power is large, and if we still consider GMR=1.25 as the upper BE limit, then the type I error rate (power at GMR =1.25) is also large. However, we may argue that the type I error rate should be the power at the expanded limit boundaries, such as GMR = 1.30, 1.41, 1.52 and 1.76 for %CV= 30, 40, 50 and 70% using the FDA RSABE. In latter case, it appears that the type I error rate is reserved for FDA RSABE, thanks to the point estimate constraint at larger %CV (see Figure C.1 to C.3 for results without the point estimate constraint).

Then, we are left with two questions: “Are the two drugs really bioequivalent when their GMR is ‘1.76’?” and “Is the type I error rate really a type I error rate with this point estimate constraint?”.

In all, if the same BE limit criteria are applied, then the power of the LI approach is comparable with the FDA RSABE or EMA RSABE.

# CHAPTER VI

# DISCUSSIONS

The FDA working group developed and recommended the RSABE for HVDs. Haidar et al. (2008b) demonstrated improved power of RSABE as the within-subject variability increases, compared with the conventional ABE test method (i.e., TOST) for $3 \times 3$ partial replicate design. Karalis et al. (2012) compared the performance of FDA RSABE with the EMA scaled approach for the partial replicate study design. They found that: 1) FDA RSABE and EMA RSABE are basically same for %CV less than 30%; 2) FDA RSABE is more permissive when %CV $> 30\%$; 3) The major difference was found for %CV $> 50\%$, where the point estimate constraint is necessary for FDA RSABE, but less important for EMA approach. The point estimate constraint is effective only for large sample size and %CV $> 50\%$ for EMA approach. Their findings are consistent with our simulation results for the full replicate $2 \times 4$ cross-over design. We found that the FDA RSABE is more permissive than EMA RSABE, especially for small sample size ($n = 24$) and large %CV (50-70%), with the largest difference at 70%. Patterson and Jones (2011) argued against the FDA RSABE for the following reasons. When the point estimate constraint drives the inference for large sample size and large within-subject variability in FDA RSABE, this method may not protect the public from the potential large changes in exposure when patients switch between generic drugs. They are also concerned that the scaled approach relies on the observed $s_{WR}$ so heavily that it may make a "bad" study easier to show BE.

The FDA RSABE and EMA approaches do not consider the variability of the test drug. Haidar et al. (2008b) and our simulations show that $\sigma_{WT}/\sigma_{WR}$ affects the power when the within-subject variability of the reference drug increases; the higher the ratio, the smaller the power.

Our simulations, along with studies of Haidar et al. (2008b) and Karalis et al. (2012), all show that, if we still consider the upper limit for BE is 1.25, then when the power is improved over the BE range ($1 \leq$ GMR $<$ 1.25), the type I error rate (power at GMR =1.25) also increases. Clearly, these frequentist approaches for BE may not guarantee any level of confidence for the GMR at the ABE boundaries of 0.8 and 1.25, even when RSABE is concluded. Thus, alternatively, we advocate the evidential likelihood framework for evaluating BE, which can show a full spectrum of evidence for BE or BIE. This approach does show what the data say regarding the strength of evidence for BE or BIE, even though it does not make any conclusions of BE or BIE. It does not conflict between the evidence and the type I & II error rates. Moreover, the probability of observing misleading evidence is small and bounded (Royall, 1997).

We used a more general numeric method [`optim()` function] to obtain the profile likelihoods for the mean difference and variance ratios in the likelihood function (III.14). For the full replicate example data, we

showed how to obtain these profile likelihoods by profiling out other nuisance parameters (computer code is provided in Appendix D). We demonstrated that, for the composite hypothesis such as the interval hypotheses in BE testing, the evidence can be represented by $k_{\max}$ (the largest $k$ so that $1/k$ LI are completely contained within the BE limit) or Zhang's GLR. The evidence can be interpreted as supporting BE (large $k_{\max}$ = GLR $> 1$), or supporting BIE (no $k_{\max}$ and very small GLR $< 1$), or as weak evidence for BE or BIE (no $k_{\max}$ and small GLR$<$1, or small $k_{\max}$ = GLR $>$1). Note that, in the cases where the data tend to support BIE (GLR$<$1), $k_{\max}$ does not exist. Therefore, we may prefer using GLR as the measure of strength of evidence for the interval hypotheses. It is impossible to show evidence for BIE using frequentist approaches such as the FDA or EMA RSABE since failing to reject the null does not imply that the null (BIE) is true.

By considering the evidence for the mean and the variability together, our likelihood approach provide evidence for PBE or IBE also, and hence easily address the prescribability or switchability issue. The current FDA position on prescribability is that the ABE criterion ensures the safety and efficacy of the generic drug and thus its prescribability (Chow et al., 2011). At present, the FDA does not consider the IBE as an applicable approach, perhaps due to the difficulty in its implementation (Chow et al., 2011). These issues are often the target of criticism around the current ABE method. Not only we may consider the mean and the variance for one pharmacokinetic parameter together, we can also consider multiple pharmacokinetic measures (AUC and Cmax) simultaneously using their joint likelihood function (Choi et al., 2008).

We used simulations to demonstrate that, if we would apply the same FDA's or EMA's limit criteria, the operating characteristics for LI approach at fixed $k$ would be comparable with the FDA RSABE or EMA approach. Note that we do not recommend using the likelihood approach for evaluating BE in this way. The simulations are for comparison with the frequentist method in terms of the operating characteristics. Instead, we advocate employing the likelihood approach to present evidence for BE in the manner that we have illustrated using the example data (CHAPTER IV).

We extended Choi et al. (2008)'s work to HVDs. The likelihood approach proposed in this thesis can be used for any BE evaluations, regardless of variability. It is true that large variability may widen the LI of the mean difference, which makes it harder to present evidence for BE. However, we may also consider the evidence for variability along with the mean. With the increase in sample size, the evidence will be stronger if the two drugs are truly equivalent. The sample size and the strength of evidence measured by GLR for the interval hypotheses determine the probabilities of observing misleading evidence, weak evidence and strong evidence (CHAPTER II) for BE. We suggest that the evaluation of evidence for BE should be decoupled from the decision making for approval of generic drugs. With the objective of minimizing the consumer risk or producer risk, regulatory authorities can work with scientists to decide how strong (how large/small GLR should be) the evidence should be to conclude BE or BIE.

# CHAPTER VII

## CONCLUSIONS

In this thesis, we used the likelihood approach to present evidence for bioequivalence or bioinequivalence for a full replicate $2 \times 4$ bioequivalence data for highly variable drugs within the evidential framework. We recommend using the generalized likelihood ratio (or $k_{\max}$ when it exists) as evidence for the interval hypotheses in bioequivalence testing. Similar to Figure II.1 for the simple hypotheses, in the future we need to find how the probabilities of observing misleading evidence and weak evidence would behave for the interval hypotheses (I.1) as a function of geometric mean ratio, sample size and generalized likelihood ratio cut off.

## Appendix A

### FDA RSABE 95% criteria bound determination

The 95 % upper confidence bound for $(\mu_T - \mu_R)^2 - \theta_S \sigma_{WR}^2$ can be derived as follows (Howe, 1974; Hyslop et al., 2000; Tothfalus et al., 2001).

The two independent terms can be estimated by their respective expected values:

$$\begin{aligned} E_m &= (m_T - m_R)^2 = \hat{\phi}^2 \text{ and} \\ E_w &= \theta_S s_{WR}^2, \end{aligned}$$

where $m_T$ and $m_R$ are the observed overall means of the test and reference formulations, respectively, and $m_T - m_R = \hat{\phi}$.

The confidence limits for the two terms are:

$$\begin{aligned} C_m &= [|m_T - m_R| + t_{1-\alpha,N-2}\mathrm{SE}(m_T - m_R)]^2 \text{ and} \\ C_w &= \theta_s (N-2) s_{WR}^2 / \chi_{1-\alpha,N-2}^2. \end{aligned}$$

The final confidence limit for $(\mu_T - \mu_R)^2 - \theta_S \sigma_{WR}^2$ is:

$$CL = E_m - E_w + (L_m + L_w)^2, \tag{A.1}$$

where

$$\begin{aligned} L_m &= (C_m - E_m)^2 \text{ and} \\ L_w &= (C_w - E_w)^2. \end{aligned}$$

# Appendix B

## Alternative approach to obtain the profile likelihood for $\phi$ using `lme()`

### B.1 Computer code

```
bedata ← read.csv("/Users/xiaofengqi/Desktop/thesis2/thesis/original files/exam44impmod3.csv")
##lme approach
bedata$Rind ← ifelse(bedata$formula=="R", 1,0)
bedata$Tind ← ifelse(bedata$formula=="T", 1,0)
likelihood ← c()
x ← seq(-0.25,0.5, length.out=200)
for(i in 1:length(x)) {
  ds ← transform(bedata, newY = ifelse(formula=="R", log(AUC), (log(AUC)-x[i])))
  linfit ← lme(fixed=newY ~ sequence+as.factor(period), method="ML",
        random = list(~0+Rind+Tind|subject), data=ds, weights= varIdent(form= ~1|formula),
                control=lmeControl(maxIter=1000))
  likelihood ← c(likelihood,exp(logLik(linfit)[1]))
}
```

### B.2 Comparison of profile likelihoods obtained using `lme()` and `optim()`

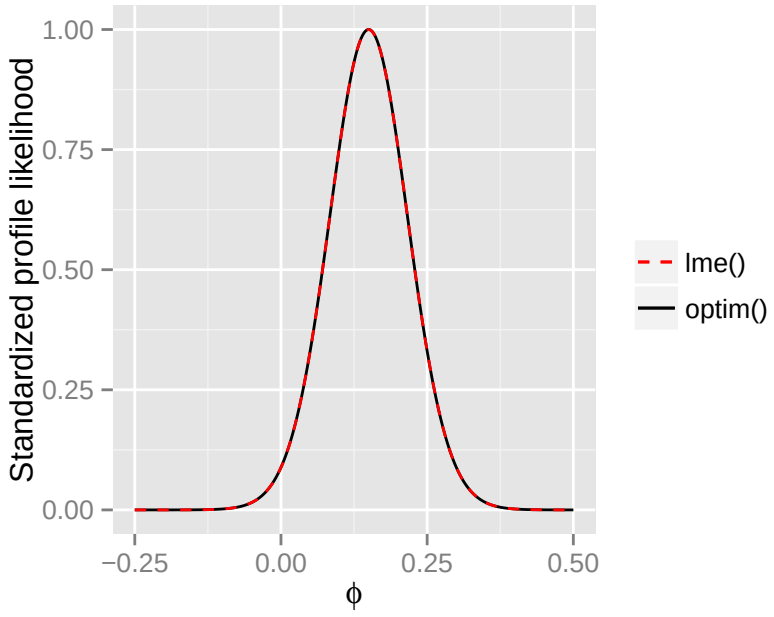


Figure B.1: Comparison of profile likelihoods obtained using `lme()` and `optim()`.

## Appendix C

### Simulation results for RSABEs without the point estimate constraint

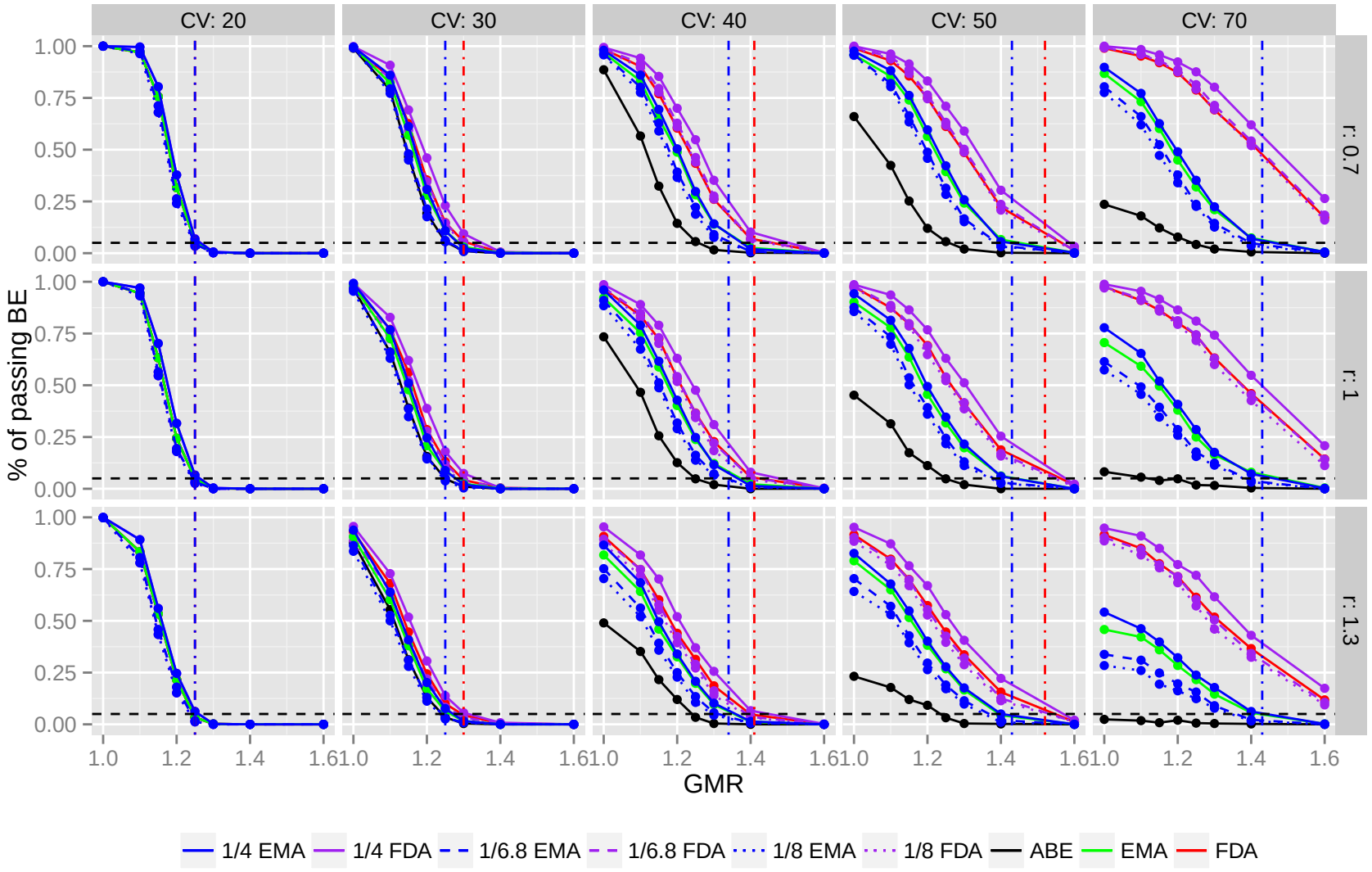


Figure C.1: Power curves ($n$=24) using different methods (without point estimate constraint) under different scenarios for $\sigma_{WR}$ and $\sigma_{WT}/\sigma_{WR}$.

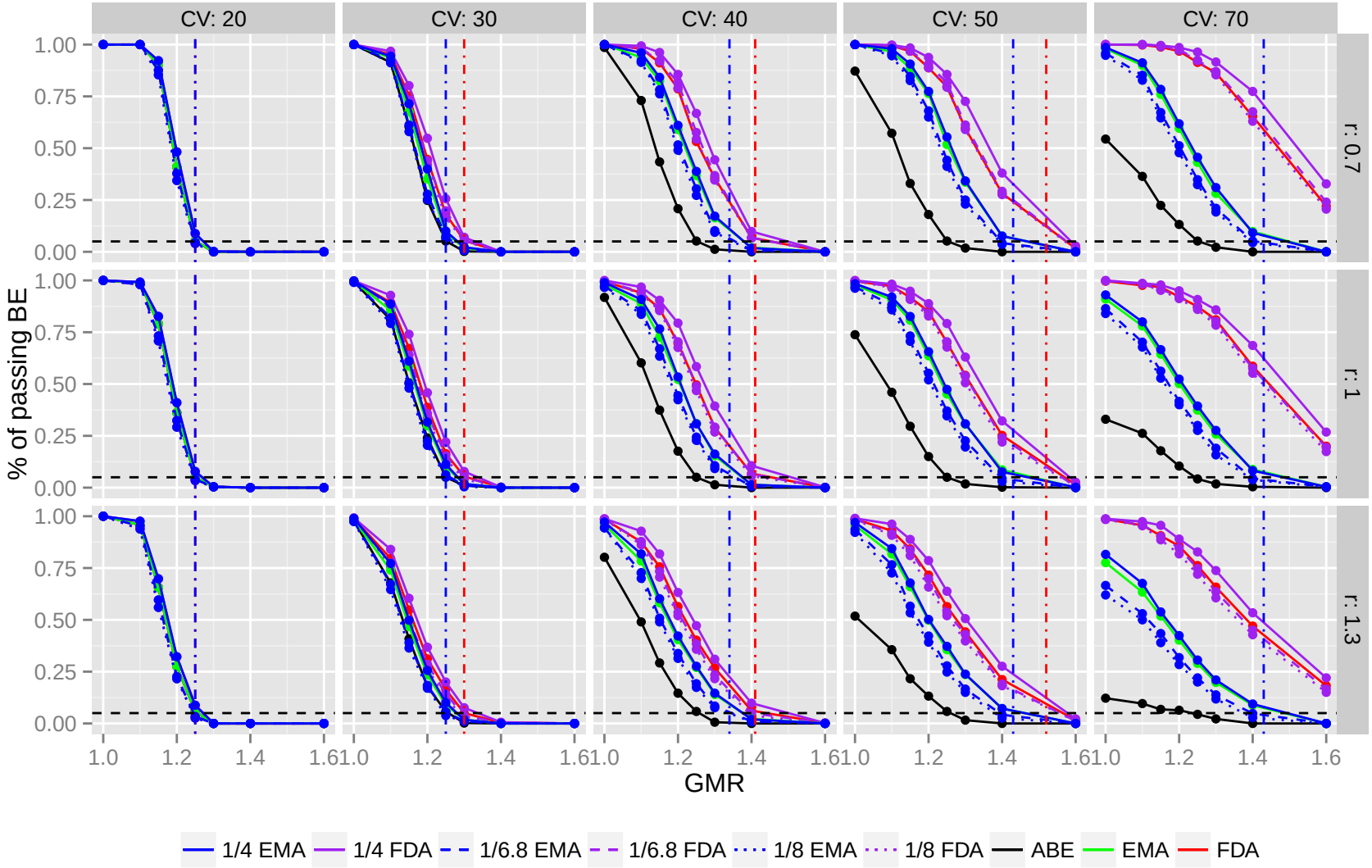


Figure C.2: Power curves ($n$=36) using different methods (without point estimate constraint) under different scenarios for $\sigma_{WR}$ and $\sigma_{WT}/\sigma_{WR}$.

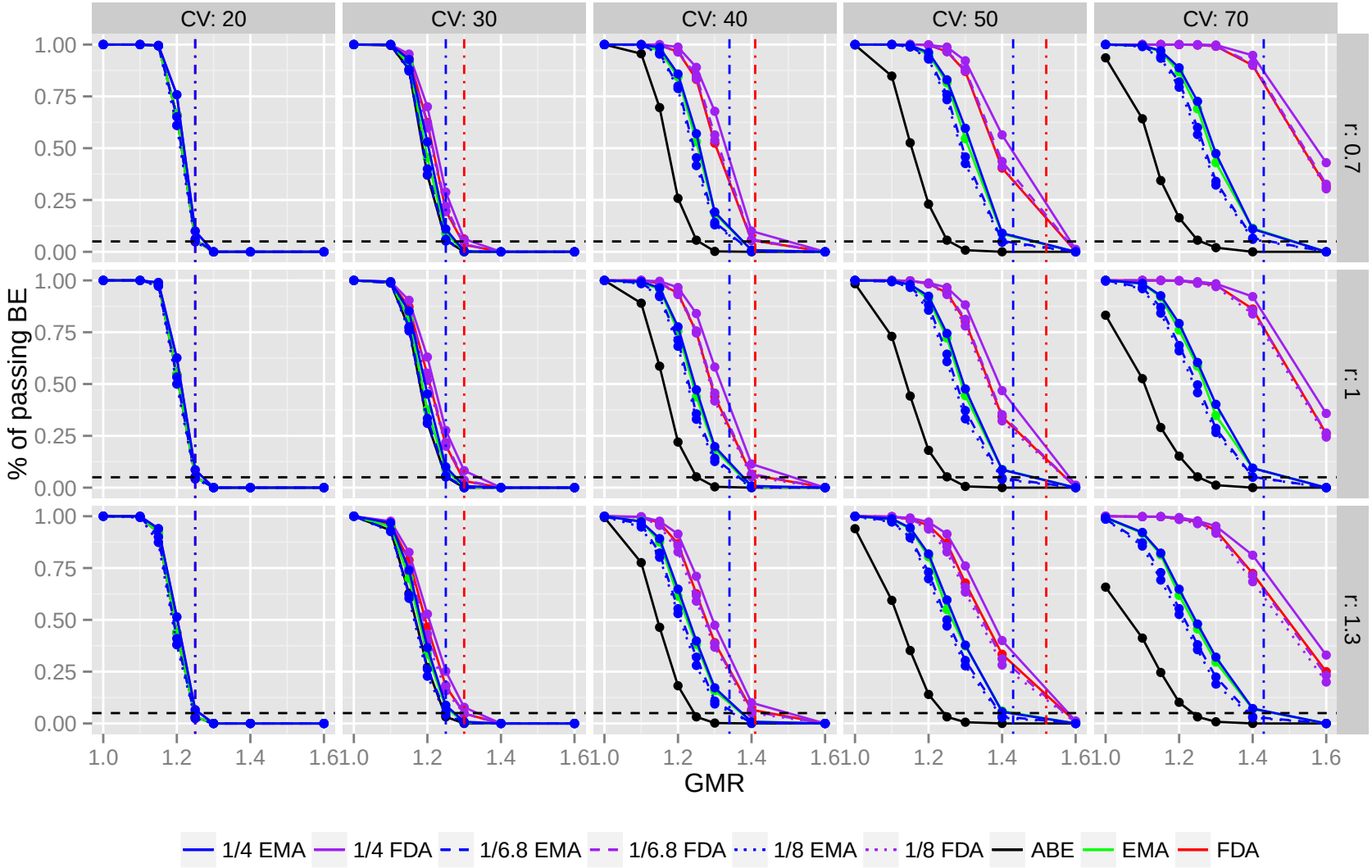


Figure C.3: Power curves ($n$=72) using different methods (without point estimate constraint) under different scenarios for $\sigma_{WR}$ and $\sigma_{WT}/\sigma_{WR}$.

# Appendix D

## Other computer code

### D.1 ABE, FDA RSABE and EMA RSABE

#### D.1.1 Moment estimators

```
bedata ← read.csv("C:/Users/dul/Desktop/LD/thesis/original files/exam44impmod3.csv")
bedata$auc ← bedata$AUC
bedata$id ← bedata$subject
bedata$seq ← bedata$sequence
## Moment estimators
# number of subject in each sequence
n1 ← length(bedata$seq[bedata$seq=="RTRT"])/length(unique(bedata$period)) ## number
#of subjects in seq 1
n2 ← length(bedata$seq[bedata$seq=="TRTR"])/length(unique(bedata$period)) ## number
#of subjects in seq 2

## intermediates
seq1sub ← subset(bedata, seq=="RTRT")
TRd1 ← c()
for (i in unique(seq1sub$id)){
  TRd1i ← with(seq1sub, mean(log(auc[id==i& formula=="T"]))−mean(log(auc[id==i&
                                                              formula=="R"])))
  TRd1 ← c(TRd1, TRd1i)
}

seq2sub ← subset(bedata, seq=="TRTR")
TRd2 ← c()
for (i in unique(seq2sub$id)){
  TRd2i ← with(seq2sub, mean(log(auc[id==i& formula=="T"]))−mean(log(auc[id==i&
                                                              formula=="R"])))
  TRd2 ← c(TRd2, TRd2i)
}

## estimate of phi
d1 ← mean(TRd1)
d2 ← mean(TRd2)
phihat ← (d1+d2)/2
```

```
## estimate of standard error for phihat
s2hat <- (var(TRd1)*(n1-1)+var(TRd2)*(n2-1))/(n1+n2-2)
sehat <- 1/2*sqrt(s2hat*(1/n1+1/n2))

## caculate the 90% confidence interval
phihatlw <- phihat -qt(0.95, df=n1+n2-2)*sehat
phihatup <- phihat +qt(0.95, df=n1+n2-2)*sehat

##exponenatiate the estimates to get the GMR estimate and its confidence intervals
ratiohat <- round(exp(phihat),2)
ratiolw <- round(exp(phihatlw),2)
ratioup <- round(exp(phihatup),2)

## estimate of sWR
Rd1 <- c()
for (i in unique(seq1sub$id)){
  Rd1i <- with(seq1sub, log(auc[id==i& period==1])-log(auc[id==i& period==3]))
  Rd1 <- c(Rd1, Rd1i)
}

Rd2 <- c()
for (i in unique(seq2sub$id)){
  Rd2i <- with(seq2sub, log(auc[id==i& period==2])-log(auc[id==i&period==4]))
  Rd2 <- c(Rd2, Rd2i)
}

sR2hat <- (var(Rd1)*(n1-1)+var(Rd2)*(n2-1))/(2*(n1+n2-2))
sRhat <- round(sqrt(sR2hat),3)

## estimate of sWT
Td1 <- c()
for (i in unique(seq1sub$id)){
  Td1i <- with(seq1sub, log(auc[id==i& period==2])-log(auc[id==i& period==4]))
  Td1 <- c(Td1, Td1i)
}

Td2 <- c()
for (i in unique(seq2sub$id)){
  Td2i <- with(seq2sub, log(auc[id==i& period==1])-log(auc[id==i&period==3]))
  Td2 <- c(Td2, Td2i)
}
```

```
sT2hat ← (var(Td1)*(n1-1)+var(Td2)*(n2-1))/(2*(n1+n2-2))
sThat ← sqrt(sT2hat)
```

### D.1.2 Random-effects model: `lme()`

```
#####################random intercept model
bedata$Rind ← ifelse(bedata$formula=="R", 1,0)
bedata$Tind ← ifelse(bedata$formula=="T", 1,0)

library(nlme)

###########
m ← lme(fixed=log(auc) ~ sequence+as.factor(period)+formula, method="REML",
        random=list(~0+Rind+Tind|id), data=bedata, weights= varIdent(form= ~1|formula))
#summary(m)
##########

##extract swr
swrrandom.auc ← summary(m)$sigma
##extract ratio of swt/swr
deviratio.auc ← 1/unique(varWeights(m$modelStruct))[2]

#library(mgcv)
#vari ← extract.lme.cov2(m, data=bedata)

##extract estimates
mtTable ← summary(m)$tTable
d ← nrow(mtTable)
est2 ← exp(mtTable[d,1])
cilow2 ← exp(mtTable[d,1]- qt(0.95,df=n1+n2-2)*mtTable[d,2])
cihigh2 ← exp(mtTable[d,1]+ qt(0.95,df=n1+n2-2)*mtTable[d,2])
```

### D.1.3 FDA RSABE

```
## using TOST results for AUC
theta_s ← (log(1.25)/0.25)^2
N ← n1+n2
Em ← phihat^2
##Em ← phihat^2-sehat^2
Ew  ← theta_s*sR2hat
Cm ← (abs(phihat) + qt(0.95, df=N-2)* sehat)^2
Cw ← theta_s * (N-2) * sR2hat/qchisq(0.95, df=N-2)
```

```
Lm ← (Cm–Em)^2
Lw ← (Cw–Ew)^2
CL ← Em–Ew+(Lm+Lw)^(1/2)



## using LME model results
##AUC
Ema ← mtTable[d,1]^2
Ew  ← theta_s*swrrandom.auc^2
Cma ← (abs(mtTable[d,1]) + qt(0.95, df=N–2)* mtTable[d,2])^2
Cw ← theta_s * (N–2) * swrrandom.auc^2 /qchisq(0.95, df=N–2)
Lma ← (Cma–Ema)^2
Lw ← (Cw–Ew)^2
CLa ← Ema–Ew+(Lma+Lw)^(1/2)
```

#### D.1.4 EMA RSABE for Cmax

```
### EMEA approach for Cmax
if (sRhatc<0.294){
  EMAlow ← 0.8
  EMAup ← 1.25
} else if(sRhatc<0.472){
  EMAlow ← round(exp(−0.76*sRhatc),4)
  EMAup ← round(exp(0.76*sRhatc),4)
  }else {
  EMAlow ← 0.6984
  EMAup ← 1.4319
  }
```

### D.2 Profile likelihoods using the `optim()` function

```
###change dataset to vector or matrix
no.p ← length(unique(bedata$period)) ## number of period
no.seq ← length(unique(bedata$sequence)) ## number of sequence

n1 ← length(bedata$seq[bedata$sequence=="RTRT"])/no.p ## number of subject in seq 1
n2 ← length(bedata$seq[bedata$sequence=="TRTR"])/no.p ## number of subject in seq 2

Y ← log(bedata$AUC) ## data(outcome)

ids ← bedata$subject
id1 ← unique(bedata$subject[bedata$sequence=="RTRT"]) ## unique id's in sequence 1
id2 ← unique(bedata$subject[bedata$sequence=="TRTR"]) ## unique id's in sequence 2
```

```
## design matrix for 2x4 design
Xi1 <- matrix(c(1,0,0,0,0,0,1,1,0,0,0,1,1,0,1,0,0,0,1,0,0,1,0,1,0,1), nrow=4, byrow=TRUE)
## design matrix of (mu, P2, P3,P4, S,Phi) for each subject in seq 1
Xi2 <- matrix(c(1,0,0,0,1,1,1,1,0,0,1,0,1,0,1,0,1,1,1,0,0,1,1,0), nrow=4, byrow=TRUE)
## design matrix of (mu, P2,P3,P4,S,Phi) for each subject in seq 2
```

### D.2.1 Profile likelihood for $\phi$

```
######## profile likelihood for phi using optim() function
x <- seq(-0.25,0.5, length.out=300)
maxL <- c() ## profile likelihood vector
j <- c() ## record of the converaged phi point

for (i in 1:length(x)) {
  ##theta= c(mu,P2, P3, P4, S, phi, sigmaST, sigmaSR, sigmaT,sigmaR,rho),
  ##sigmaST=log(sigmaST^2), sigmaSR=log(sigmaST^2), sigmaT=log(sigmaT^2),
  #sigmaR=log(sigmaR^2)
  loglik2by4phi <- function(theta, phi=x[i]){
      ## parameters beta
      beta <- c(theta[1],theta[2],theta[3],theta[4],theta[5],phi)
      ## parameters (mu, P2,P3,P4,S, phi)

      # construct the variance-covariance matrix
      vi1 <- vi2 <- matrix(NA, nrow=4, ncol=4)

      vi1[1,1] <- vi1[3,3] <- vi2[2,2] <- vi2[4,4] <- exp(theta[8])+ exp(theta[10])
      vi1[2,2] <- vi1[4,4] <- vi2[1,1] <- vi2[3,3] <- exp(theta[7])+ exp(theta[9])

      vi1[1,3] <- vi1[3,1] <- vi2[2,4] <- vi2[4,2] <- exp(theta[8])

      vi1[2,4] <- vi1[4,2] <- vi2[1,3] <- vi2[3,1] <- exp(theta[7])


      vi1[1,2] <- vi1[2,1] <- vi1[1,4] <- vi1[4,1] <- vi1[2,3] <- vi1[3,2] <-
        vi1[3,4] <- vi1[4,3] <-  theta[11]*sqrt(exp(theta[7])*exp(theta[8]))

      vi2[1,2] <- vi2[2,1] <- vi2[1,4] <- vi2[4,1] <- vi2[2,3] <- vi2[3,2] <-
        vi2[3,4] <- vi2[4,3] <-  theta[11]*sqrt(exp(theta[7])*exp(theta[8]))

      ## log likelihood function for all subjects
      l <- 0  ## sum of log likelihood
```

```
    for (i in 1: n1){
    yi ← Y[ids==id1[i]]
    l ← l-no.p/2*log(2*pi)-log(det(vi1))/2-1/2*t(yi-Xi1%*%beta)%*%
      solve(vi1)%*%(yi-Xi1%*%beta)
    } ## sum of log likelihood for subjects in seq 1

    for(i in 1: n2){
    yi ← Y[ids==id2[i]]
    l ← l-no.p/2*log(2*pi)-log(det(vi2))/2-1/2*t(yi-Xi2%*%beta)%*%
      solve(vi2)%*%(yi-Xi2%*%beta)
    } ## sum of log likelihood for subjects in seq 2
  return(-l) ## negative log likelihood for minimization by optim().
  }

  ## theta: initial values for the nuisance parameters

  theta ← c(5.8, 0.107, 0.116, 0.254, -0.08, 0, -1.35, -1.43,-1.63 ,-2.18, 0.78)

  ## optimization to find the maxi likelihood at fixed phi
  op ← optim(theta, loglik2by4phi, method="CG", control=list(maxit=2000))
  con ← op$convergence
  if (con==0){
    j ← c(j,i)
    maxL← c(maxL,exp(-op$value))
  }
}
```

### D.2.2 Profile likelihood for $\sigma_{TT}/\sigma_{TR}$

```
#### profile likelihood for alpha=log((sigma.sT^2+sigma_T^2)/(sigma.sR^2+sigma_R^2))###
alphas← seq(0.7, 0.9, length.out=50)

maxL ← c() ## profile likelihood vector
j ← c() ## record of the converaged phi point

for (i in 1:length(alphas)) {

  ##theta= c(mu,P2, P3, P4, S, phi, alpha, sigmasR, sigmaT, sigmaR, rho), s=log(sigma.s^2),
  #alpha=log((sigmaST^2+sigmaT^2)/(sigmaSR^2+sigmaR^2))
  loglik2by4var1 ← function(theta, alpha=alphas[i]){
```

```
    ## parameters beta
    beta ← c(theta[1],theta[2],theta[3],theta[4],theta[5], theta[6])
    ## parameters (mu, P2,P3,P4, S,phi)
    # construct the variance-covariance matrix
    vi1 ← vi2 ← matrix(NA, nrow=4, ncol=4)

    vi1[1,1] ← vi1[3,3] ← vi2[2,2] ← vi2[4,4] ← exp(theta[8])+exp(theta[10])
    vi1[2,2] ← vi1[4,4] ← vi2[1,1] ← vi2[3,3] ←
      (exp(theta[8])+exp(theta[10]))*exp(alpha)

    vi1[1,3] ← vi1[3,1] ← vi2[2,4] ← vi2[4,2] ← exp(theta[8])
    vi1[2,4] ← vi1[4,2] ← vi2[1,3] ← vi2[3,1] ← (exp(theta[8])+exp(theta[10]))
    *exp(alpha)-exp(theta[9])

    vi1[1,2] ← vi1[2,1] ← vi1[1,4] ← vi1[4,1] ← vi1[2,3] ← vi1[3,2] ←
      vi1[3,4] ← vi1[4,3] ← theta[11]*sqrt(exp(theta[8]))*sqrt((exp(theta[8])+
                                  exp(theta[10]))*exp(alpha)-exp(theta[9]))

    vi2[1,2] ← vi2[2,1] ← vi2[1,4] ← vi2[4,1] ← vi2[2,3] ← vi2[3,2] ←
      vi2[3,4] ← vi2[4,3] ← theta[11]*sqrt(exp(theta[8]))*sqrt((exp(theta[8])+
                                  exp(theta[10]))*exp(alpha)-exp(theta[9]))



    ## log likelihood function for all subjects
    l ← 0  ## sum of log likelihood

    for (i in 1: n1){
      yi ← Y[ids==id1[i]]
      l ← l-no.p/2*log(2*pi)-log(det(vi1))/2-1/2*t(yi-Xi1%*%beta)%*%
        solve(vi1)%*%(yi-Xi1%*%beta)
    } ## sum of log likelihood for subjects in seq 1

    for(i in 1: n2){
      yi ← Y[ids==id2[i]]
      l ← l-no.p/2*log(2*pi)-log(det(vi2))/2-1/2*t(yi-Xi2%*%beta)%*%
        solve(vi2)%*%(yi-Xi2%*%beta)
    } ## sum of log likelihood for subjects in seq 2

  return(-l) ## negative log likelihood for minimization by optim().
}
```

```
  ## theta: initial values for the nuisance parameters
  theta ← c(5.75, 0.105, 0.104, 0.244, −0.08, 0.15, 0, −1.71,−1.53 ,−2.27, 0.78)
  ###optimization at fixed alpha
  op ← optim(theta, loglik2by4var1,control=list(maxit=2000), method="CG")
  con ← op$convergence
  if (con==0){
    j ← c(j,i)
    maxL← c(maxL,exp(−op$value))
  }
}
```

### D.2.3 Profile likelihood for $\sigma_{WT}/\sigma_{WR}$

```
#### profile likelihood for alpha=log(sigma.T^2/sigma_R^2)###
alphas← seq(−0.55, 1.6, length.out=300)
maxL ← c() ## profile likelihood vector
j ← c() ## record of the converaged phi point

for (i in 1:length(alphas)) {
  ##theta= c(mu,P2, P3, P4, S, phi, sigmasT sigmasR, alpha, sigmaR, rho),
  #s=log(sigma.s^2), alpha=log(sigma.T^2/sigma_R^2)
  loglik2by4var2 ← function(theta, alpha=alphas[i]){

      ## parameters beta
      beta ← c(theta[1],theta[2],theta[3],theta[4],theta[5], theta[6])
      ## parameters (mu, P2,P3,P4, S,phi)
      # construct the variance-covariance matrix
      vi1 ← vi2 ← matrix(NA, nrow=4, ncol=4)

      vi1[1,1] ← vi1[3,3] ← vi2[2,2] ← vi2[4,4] ← exp(theta[8])+exp(theta[10])
      vi1[2,2] ← vi1[4,4] ← vi2[1,1] ← vi2[3,3] ←
        exp(theta[7])+exp(theta[10])∗exp(alpha)

      vi1[1,3] ← vi1[3,1] ← vi2[2,4] ← vi2[4,2] ← exp(theta[8])
      vi1[2,4] ← vi1[4,2] ← vi2[1,3] ← vi2[3,1] ← exp(theta[7])

      vi1[1,2] ← vi1[2,1] ← vi1[1,4] ← vi1[4,1] ← vi1[2,3] ← vi1[3,2] ←
        vi1[3,4] ← vi1[4,3] ← theta[11]∗sqrt(exp(theta[8]))∗sqrt(exp(theta[7]))

      vi2[1,2] ← vi2[2,1] ← vi2[1,4] ← vi2[4,1] ← vi2[2,3] ← vi2[3,2] ← vi2[3,4]
      ← vi2[4,3] ← theta[11]∗sqrt(exp(theta[8]))∗sqrt(exp(theta[7]))
```

```
    ## log likelihood function for all subjects
    l ← 0  ## sum of log likelihood

    for (i in 1: n1){
      yi ← Y[ids==id1[i]]
      l ← l−no.p/2∗log(2∗pi)−log(det(vi1))/2−1/2∗t(yi−Xi1%∗%beta)%∗%
        solve(vi1)%∗%(yi−Xi1%∗%beta)
    } ## sum of log likelihood for subjects in seq 1

    for(i in 1: n2){
      yi ← Y[ids==id2[i]]
      l ← l−no.p/2∗log(2∗pi)−log(det(vi2))/2−1/2∗t(yi−Xi2%∗%beta)%∗%
        solve(vi2)%∗%(yi−Xi2%∗%beta)
    } ## sum of log likelihood for subjects in seq 2

  return(−l) ## negative log likelihood for minimization by optim().
}

## theta: initial values for the nuisance parameters

  theta ← c(5.71, 0.116, 0.183, 0.312, −0.08, 0.15, −1.26, −1.71, 0 , −1.5, 0.88)

if(!is.null(tryCatch(op ← optim(theta, loglik2by4var2,control=list(maxit=2000),
                          method="CG"), error = function(e) {}))) {
  con ← op$convergence
if (con==0){
  j ← c(j,i)
  maxL← c(maxL,exp(−op$value))
  }
}
}
```

### D.3 Simulations

```
library(MASS)
library(nlme)
### function to generate data for full replicate cross-over (2x4)
## no squence effect, no period effect, random effect for i (R=T)
### sigmaBR=sigmaBT=sigmaS=0.2, sigmaWR
ABEdata ← function(n1, n2, mu, phi, sigmaS, sigmaWT, sigmaWR){
  ## generate data for sequence 1: RTRT
  random1 ← rnorm(n1, mean=0, sd=sigmaS)
```

```
  random1 ← rep(random1, each=4)
  fixed1 ← rep(c(mu, mu+phi, mu, mu+phi), n1)
  sigma1 ← matrix(c(sigmaWR^2, 0, 0, sigmaWT^2),2,2)
  error10 ← mvrnorm(2*n1, rep(0,2), sigma1)
  error1 ← c()
  k ← 2*n1
  for(i in 1:k){
    error1 ← c(error1, error10[i,1], error10[i,2])
  }
  Y1 ← fixed1 + random1 + error1

  ## generata data for sequence 2:TRTR
  random2 ← rnorm(n2, mean=0, sd=sigmaS)
  random2 ← rep(random2, each=4)
  fixed2 ← rep(c(mu+phi, mu, mu+phi, mu), n2)
  sigma2 ← matrix(c(sigmaWT^2, 0, 0, sigmaWR^2),2,2)
  error20 ← mvrnorm(2*n2, rep(0,2), sigma2)
  error2 ← c()
  k ← 2*n2
  for(i in 1:k){
    error2 ← c(error2, error20[i,1], error20[i,2])
  }

  Y2 ← fixed2 + random2 + error2

  ##### response Y
  Y ← c(Y1, Y2)

  #### design matrix (no period, no sequence effect)
  formula1 ← rep(c(0,1,0,1), n1) ## 1-test, 0-reference
  formula2 ← rep(c(1,0,1,0), n2)
  formula ← c(formula1, formula2)
  period ← rep(seq(1:4), n1+n2)
  sequence ← c(rep(1, 4*n1), rep(2,4*n2))
  subject ← rep(seq(1:(n1+n2)), each=4)
  #####
  data ← as.data.frame(cbind(subject,Y, formula, period, sequence))
  #data$Rind ← ifelse(data$formula==0, 1,0)
  #data$Tind ← ifelse(data$formula==1, 1,0)
  return(data)
}
```

```
#### function to get the LIs and compare with FDA and EMA RSABE limits
spintervals ← function(sWR, data, xmin, xmax){
  ### use  lme() to get the profile likelihood
  x ← seq(xmin, xmax, length.out=200) ## grid of x
  likelihood ← c()
  j ← c()
  for (i in 1:length(x)) {
    phi ← x[i]
    ds ← transform(data, newY = ifelse(formula==0, Y, (Y-phi))) ##new data

    if(!is.null(tryCatch(fit ← lme(fixed=newY ~ 1 , method="ML",random = ~1|subject,
        data=ds, weights= varIdent(form= ~1|formula), control=lmeControl(maxIter=2000)),
                         error = function(e) {}))) {
      likelihood ← c(likelihood,exp(logLik(fit)[1])) ## maximum likelihood at each x(phi)
      j ← c(j, i) ## record x at which no error
    }
  }


  if(length(j)>1){

    # print(likelihood)
    ##caculate 1/8th 1/6.8th and 1/4 LIs
    lik.norm ← likelihood/max(likelihood) ## normalize the likelihood

    phi.x ← x[j]

    #plot(phi.x, lik.norm)

    phi8.x ← phi.x[lik.norm ≥1/8]  ##phi's support by 1/8
    phi6.8.x ← phi.x[lik.norm ≥1/6.8] ##phi's support by 1/6.8
    phi4.5.x ← phi.x[lik.norm ≥1/4] ##phi's support by 1/4

    phi.max ← max(phi.x[lik.norm==max(lik.norm)])
    ## maximum likelihood estimator for phi
    phi8.low ← min(phi8.x)
    phi8.up ←  max(phi8.x)  ## 1/8 th interval

    phi6.8.low ← min(phi6.8.x)
```

```
phi6.8.up ← max(phi6.8.x)   ## 1/6.8 th interval

phi4.5.low ← min(phi4.5.x)
phi4.5.up ← max(phi4.5.x) ## 1/4.5 th interval

### wheter those interals are within limits
sp8f ← sp6f ← sp4f ← 0 ## indicator whether interval is within the FDA limit
sp8e ← sp6e ← sp4e ← 0 ## indicator whether interval is within the EMA limit

sp8f1 ← sp6f1 ← sp4f1 ← 0 ## indicator whether BE limit + point constraint
sp8e1 ← sp6e1 ← sp4e1 ← 0  ## indicator whether BE limit + point constraint


if(sWR<0.294){
  if(phi8.low ≥ −0.223 & phi8.up ≤ 0.223){
    sp8f ← sp6f ← sp4f ←sp8e ← sp6e ← sp4e ← 1
    sp8f1 ← sp6f1 ← sp4f1 ←sp8e1 ← sp6e1 ← sp4e1 ← 1

  }else if(phi6.8.low ≥ −0.223 & phi6.8.up ≤ 0.223){
    sp6f ← sp4f ← sp6e ← sp4e ← 1
    sp6f1 ← sp4f1 ← sp6e1 ← sp4e1 ← 1

  } else if(phi4.5.low ≥ −0.223 & phi4.5.up ≤ 0.223){
    sp4f ← sp4e ← 1
    sp4f1 ← sp4e1 ← 1
  }
}

###FDA limits when sWR ≥ 0.294 no constraint
if(sWR ≥ 0.294){
  if(phi8.low ≥ −sWR∗0.8926 & phi8.up ≤ sWR∗0.8926){
    sp8f ← sp6f ← sp4f ← 1
  } else if(phi6.8.low ≥ − sWR∗0.8926& phi6.8.up ≤ sWR∗0.8926){
    sp6f ← sp4f ← 1
  } else if(phi4.5.low ≥ −sWR∗0.8926& phi4.5.up ≤ sWR∗0.8926){
    sp4f ← 1
  }
}

###FDA limits when sWR ≥ 0.294 with constraint
if(sWR ≥ 0.294){
```

```
  if(phi8.low ≥ -sWR*0.8926 & phi8.up ≤ sWR*0.8926 & phi.max ≤ 0.223&phi.max ≥ -0.223){
    sp8f1 ← sp6f1 ← sp4f1 ← 1
  } else if(phi6.8.low ≥ - sWR*0.8926& phi6.8.up ≤ sWR*0.8926 & phi.max ≤ 0.223&
              phi.max ≥ -0.223){
    sp6f1 ← sp4f1 ← 1
  } else if(phi4.5.low ≥ -sWR*0.8926& phi4.5.up ≤ sWR*0.8926 & phi.max ≤ 0.223&
              phi.max ≥ -0.223){
    sp4f1 ← 1
  }
}



###EMA limits when sWR ≥ 0.294 and <0.472 no constraint
if(sWR ≥ 0.294 & sWR<0.472) {

  if(phi8.low ≥ -sWR*0.76 & phi8.up ≤ sWR*0.76){
    sp8e ← sp6e ← sp4e ← 1
  } else if(phi6.8.low ≥ -sWR*0.76 & phi6.8.up ≤ sWR*0.76){sp6e ← sp4e ← 1
  } else if(phi4.5.low ≥ -sWR*0.76& phi4.5.up ≤ sWR*0.76){
    sp4e ← 1
  }
}else if(sWR ≥ 0.472){

  if(phi8.low ≥ log(0.6984) & phi8.up ≤ log(1.4319)){
    sp8e ← sp6e ← sp4e ← 1
  } else if(phi6.8.low ≥ log(0.6984)& phi6.8.up ≤ log(1.4319)){
    sp6e ← sp4e ← 1
  } else if(phi4.5.low ≥ log(0.6984)& phi4.5.up ≤ log(1.4319)){sp4e ← 1}

}

###EMA limits when sWR ≥ 0.294 and <0.472 with constraint
if(sWR ≥ 0.294 & sWR<0.472) {

  if(phi8.low ≥ -sWR*0.76 & phi8.up ≤ sWR*0.76 & phi.max ≤ 0.223&phi.max ≥ -0.223){
    sp8e1 ← sp6e1 ← sp4e1 ← 1
  } else if(phi6.8.low ≥ -sWR*0.76 & phi6.8.up ≤ sWR*0.76 & phi.max ≤ 0.223&
              phi.max ≥ -0.223)
    {sp6e1 ← sp4e1 ← 1
  } else if(phi4.5.low ≥ -sWR*0.76& phi4.5.up ≤ sWR*0.76 & phi.max ≤ 0.223&
```

```
                    phi.max >= -0.223){
        sp4e1 <- 1
      }
    }else if(sWR >= 0.472){

      if(phi8.low >= log(0.6984) & phi8.up <= log(1.4319) & phi.max <= 0.223&phi.max >= -0.223){
        sp8e1 <- sp6e1 <- sp4e1 <- 1
      } else if(phi6.8.low >= log(0.6984)& phi6.8.up <= log(1.4319) & phi.max <= 0.223&
                    phi.max >= -0.223){
        sp6e1 <- sp4e1 <- 1
      } else if(phi4.5.low >= log(0.6984)& phi4.5.up <= log(1.4319) & phi.max <= 0.223&
                    phi.max >= -0.223)
      {sp4e1<- 1}

    }

    return(c(sp8f, sp6f, sp4f, sp8f1, sp6f1, sp4f1,sp8e, sp6e, sp4e,sp8e1, sp6e1, sp4e1))
    ## indicator of interval within the limit and the maximum likelihood
  }else{
    return(c(NA, NA, NA, NA,NA, NA, NA,NA, NA,NA, NA, NA))}
}

##############function for FDA and EMA approaches
FDA2EMEA <- function(data){

  # number of subject in each sequence
  n1 <- length(data$sequence[data$sequence==1])/length(unique(data$period))
  ## number of subjects in seq 1
  n2 <- length(data$sequence[data$sequence==2])/length(unique(data$period))
  ## number of subjects in seq 2

  ## intermidates
  seq1sub <- subset(data, sequence==1)
  TRd1 <- c()
  for (i in unique(seq1sub$subject)){
    TRd1i <- with(seq1sub, mean(Y[subject==i& formula==1])-mean(Y[subject==i& formula==0]))
    TRd1 <- c(TRd1, TRd1i)
  }

  seq2sub <- subset(data, sequence==2)
  TRd2 <- c()
```

```
for (i in unique(seq2sub$subject)){
  TRd2i <- with(seq2sub, mean(Y[subject==i& formula==1])-mean(Y[subject==i& formula==0]))
  TRd2 <- c(TRd2, TRd2i)
}

## estimate phi
d1 <- mean(TRd1)
d2 <- mean(TRd2)
phihat <- (d1+d2)/2

## estimate of standard error of phihat
s2hat <- (var(TRd1)*(n1-1)+var(TRd2)*(n2-1))/(n1+n2-2)
sehat <- 1/2*sqrt(s2hat*(1/n1+1/n2))

## caculate the 90% confidence interval
phihatlw <- phihat -qt(0.95, df=n1+n2-2)*sehat
phihatup <- phihat +qt(0.95, df=n1+n2-2)*sehat

##exponentiate the estimates
ratiohat <- exp(phihat)
ratiolw <- exp(phihatlw)
ratioup <- exp(phihatup)

## estimate of sWR
Rd1 <- c()
for (i in unique(seq1sub$subject)){
  Rd1i <- with(seq1sub, Y[subject==i& period==1]-Y[subject==i& period==3])
  Rd1 <- c(Rd1, Rd1i)
}

Rd2 <- c()
for (i in unique(seq2sub$subject)){
  Rd2i <- with(seq2sub, Y[subject==i& period==2]-Y[subject==i&period==4])
  Rd2 <- c(Rd2, Rd2i)
}

sR2hat <- (var(Rd1)*(n1-1)+var(Rd2)*(n2-1))/(2*(n1+n2-2))
sRhat <- sqrt(sR2hat) ## sWR

##########FDA's conventional ABE
ABE <- 0  ## bioequivalence indicator
```

```
  if(ratiolw >= 0.8 & ratioup <= 1.25){ABE <- 1}

  ##FDA RSABE (2 step) approach
  SABE <- 0 ## bioequivalence indicator for FDA RSABE
  SABE0 <- 0 ## bioequivalence indicator for FDA RSABE without constraint
  EMEA <- 0 ## for EMA RSABE
  EMEA0 <- 0 ## EMA without constrain

  if (sRhat<0.294){
    if(ratiolw >= 0.8 & ratioup <= 1.25){SABE <-SABE0 <- 1}  ## unscaled ABE
  }else{
    ## if sWR larger than 0.294 scaled ABE
    theta_s <- (log(1.25)/0.25)^2
    N <- n1+n2
    Ema <- phihat^2
    Ew  <- theta_s*sR2hat
    Cma <- (abs(phihat) + qt(0.95, df=N-2)* sehat)^2
    Cw <- theta_s * (N-2) * sR2hat /qchisq(0.95, df=N-2)
    Lma <- (Cma-Ema)^2
    Lw <- (Cw-Ew)^2
    CLa <- Ema-Ew+(Lma+Lw)^(1/2)
    if(CLa <= 0) SABE0 <- 1
    if(CLa <= 0 & ratiohat <= 1.25 & ratiohat >= 0.8) SABE <- 1
  }

  ### EMEA approach
  if (sRhat<0.294){
    if(ratiolw >= 0.8 & ratioup <= 1.25){EMEA <-EMEA0 <-1}  ## unscaled ABE
  }else if (sRhat<0.472) {
    ## if swr larger than 0.294 but smaller than 0.472
    if(ratiolw >= exp(-0.76*sRhat) & ratioup <= exp(0.76*sRhat)) EMEA0 <- 1
    if(ratiolw >= exp(-0.76*sRhat) & ratioup <= exp(0.76*sRhat)&ratiohat <= 1.25
       & ratiohat >= 0.8 ) EMEA <- 1
  }else if (sRhat >= 0.472) {
    #f swr larger than 0.472
    if (ratiolw >= 0.6984 & ratioup <= 1.4319) EMEA0 <- 1
    if (ratiolw >= 0.6984 & ratioup <= 1.4319 & ratiohat <= 1.25 & ratiohat >= 0.8) EMEA <- 1
  }

  return(c(sRhat, ABE, SABE0, SABE, EMEA0, EMEA))
}
```

## BIBLIOGRAPHY


Berger, R. L. and Hsu, J. C. "Bioequivalence Trials, Intersection-Union Tests and Equivalence Confidence Sets." *Statistical Science*, 11(4):283–319 (1996).

Blume, J. and Peipert, J. "What your statistician never told you about P-values." *Journal of Americian Association of Gynecologic Laparoscopists*, 10:439–444 (2003).

Blume, J. D. "Likelihood methods for measuring statistical evidence." *Stat. Med.*, 21:2563–2599 (2002).

—. "How often likelihood ratios are misleading in sequential trials." *Commun. Stat. Theory Methods*, 37:1193–1206 (2008).

Brown, L. D., Hwang, J. G., and Munk, A. "An Unbiased Test for the Bioequivalence Problem." *The Annals of Statistics*, 25(6):2345–2367 (1997).

Choi, L., Caffo, B., and Rohde, C. "A survery of the likelihood approach to bioequivalence trials." *Statistics in Medicine*, 27:4874–4894 (2008).

Chow, S.-C., Endrenyi, L., Chi, E., yang, L.-Y., and Tothfalusi, L. "Statistical Issues in Bioavailability/Bioequivalence Studies." *J Bioequiv Availab*, S1 (2011).

Davit, B. M., Chen, M.-L., Conner, D. P., Haidar, S. H., Kim, S., Lee, C. H., Lionberger, R. A., Makhlouf, F. T., patrick E. Nwakama, patel, D. T., Schuirmann, D. J., and lawrence X. Yu. "Implementation of a Reference-Scaled Average Bioequivalence Approach for Highly Variable Generic Drug Products by the US Food and Drug Administration." *The AAPS journal*, 14(4):915–924 (2012).

Davit, B. M., Conner, D. P., Fabian-Fritsch, B., Haidar, S. H., Jiang, X., Patel, D. T., Seo, P. R. H., Suh, K., Thompson, C. L., and X.Yu, L. "Highly Variable Drugs: Observations from Bioequivalence Data Submitted to the FDA for New Generic Drug Applications." *The AAPS Journal*, 10(1):148–156 (2008).

Davit, B. M., Nwatama, P. E., Buehler, G. J., PConner, D., Haidar, S. H., Patel, D. T., Yang, Y., Yu, L. X., and Woodcock, J. "Comparing Generic and Innovator Drugs: A review of 12 years of bioequivalence Data from the United States Food and Drug Administration." *The Annals of Pharmacotherapy*, 43:1583–1597 (2009).

de Souza, R. M., Achcar, J. A., and Martinez, E. Z. "Use of Bayesian methods for multivariate bioequivalence measures." *Journal of Biopharmaceutical Statistics*, 19:42–66 (2009).

Dentali, F., Donadini, M. P., Clark, N., Crowther, M. A., Garcia, D., Hylek, E., Witt, D. M., and Ageno, W. "Brand Name versus Generic Warfarin: A Systematic Review of the Literature." *Pharmacotherapy*, 31(4):386–393 (2011).

EMA, 2010. "Guideline on the Investigation of Bioequivalence." www.ema.europa.eu/pdfs/human/qwp/140198enrev1fin.pdf (2010).

FDA, 2001. "Guidance for Industry:Statistical Approaches to Establishing Bioequivalene." http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm070244.pdf (2001).

FDA, 2011. "Draft Guideline on Progesterone." http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM209294.pdf (2011).

FDA, 2012. "Approved drug products with therapeutic equivalence evalutions, 32 edition." http://www.fda.gov/downloads/Drugs/DevelopmentApprovalProcess/ucm071436.pdf (2012).

Ghosh, P. and Gonen, M. "Bayesian modeling of multivariate average bioequivalence." *Statistics in Medicine*, 27:2402–2419 (2008).

Hacking, I. *Logic of statistical inference*. New York: Cambridge University Press (1965).

Haidar, S. H., Davit, B., Chen, M.-L., Conner, D., Lee, L., H.LI, Q., Lionberger, R., Markhlouf, F., Patel, D., Schuirmann, D. J., and X.Yu, L. "Bioequivalence Approaches for Highly Variable Drugs and Drug Products." *Pharmaceutical Research*, 25(1):237–241 (2008a).

Haidar, S. H., Markhlouf, F., Schuirmann, D. J., Hyslop, T., Davit, B., Conner, D., and X.Yu, L. "Evaluation of a Scaling Approach for the Bioequivalence of Highly Variable Drugs." *The AAPS Journal*, 10(3):450–454 (2008b).

Hauck, W. W. and Anderson, S. "A new statistical procedure for testing equivalence in two-group comparative bioavailability trials." *Journal of Pharmacokinetics and Biopharmaceutics*, 12(1):83–91 (1984).

Hoenig, J. M. and Heisey, D. M. "The abuse of power: the pervasive fallacy of power calculations for data analysis." *The American Statistician*, 55(1):19–24 (2001).

Howe, W. "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(347):789–794 (1974).

Hyslop, T., Hsuan, F., and Holder, D. J. "A small sample confidence interval appraoch to assess individual bioequivalence." *Statistics in Medicine*, 19:2885–2897 (2000).

Karalis, V., Symillides, M., and Macheras, P. "Bioequivalence of Highly Variable Drugs: A Comparison of the Newly Proposed Regulatory Approaches by FDA and EMA." *Pharmaceutical Research*, 29:1066–1077 (2012).

Liu, J.-P. and Chow, S.-C. "Bioequivalence Trials, Intersection-Union Tests and Equivalence Confidence Set: Comment." *Statistical Science*, 11(4):306–312 (1996).

Munk, A. and Pfluger, R. "1-alpha Equivariant Confidence Rules for Convex Alternatives are alpha/2 - level Tests-with Applications to the Multivariate Assessmetn of Bioequivalence." *Journal of the American Statistical Association*, 94(448):1311–1319 (1999).

Patterson, S. and Jones, B. *Bioequivalence and Statistics in Clinical Pharmacology*. Boca, London and New York: Chapman Hall/CRC Press (2005).

Patterson, S. D. and Jones, B. "Viewpoint: observations on scaled average bioequivalence." *Pharmaceutical Statistics*, 11:1–7 (2011).

Qcana, J., O., M. P. S., Sanchez, A., and Carrasco, J. L. "On equivalence and bioequivalence testing." *SORT*, 32(2):151–176 (2008).

Royall, R. *Statistical Evidence: A likelihood paradigm*. New York: Chapman Hall (1997).

—. "On the probability of observing misleading statistical evidence." *Journal of American Statistical Association*, 95:760–768 (2000).

Schuirmann, D. J. "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*, 15(6):657–680 (1987).

Shaw, S. J. and Hartman, A. L. "The Controversy over Generic Antiepileptic Drugs." *J Pediatr Pharmacol Ther*, 15(2):81–93 (2010).

Talati, R., Scholle, J. M., Phung, O. P., Baker, E. L., Baker, W. L., Ashaye, A., Kluger, J., Coleman, C. I., and White, C. M. "Efficacy and Safety of Innovator versus Generic Drugs in Patients with Epilepsy: A Systematic Review." *Pharmacotherapy*, 32(4):314–322 (2012).

Tamboli, A. M., Todkar, P., Zope, P., and Sayyad, F. "An Overview on Bioequivalence: Regulatory Consideration for generic Drug Products." *J Bioequiv Availab*, 2(4):086–092 (2010).

Tothfalus, L., Endrenyi, L., Midha, K. K., Rawson, M. J., and Hubbard, J. W. "Evaluation of the Bioequivalence of Highly-Variable Drugs and Drug Products." *Pharmaceutical Research*, 18(6):728–733 (2001).

Wang, S.-J. and Blume, J. D. "An evidential approach to non-inferiority Clinical Trials." *Pharmaceutical Statistics*, 10(5):440–447 (2011).

Westlake, W. "Use of confidence intervals in analysis of comparative bioavailability trials." *Journal of Pharmaceutical Science*, 61:1340–1341 (1972).

Zhang, Z. "Interpreting Statistical Evidence with Empirical Likelihood Functions." *Biometrical Journal*, 51(4):710–720 (2009).

Zhang, Z. and Zhang, B. "A Likelihood Paradigm for Clinical Trials." *Journal of Statistical Theory and Practice*, 7(2):157–177 (2013).