Memorial Sloan-Kettering Cancer Center
Memorial Sloan-Kettering Cancer Center, Dept. of Epidemiology
& Biostatistics Working Paper Series

Year 2006 *Paper* 8

Semiparametric Bayesian Modeling of Multivariate Average Bioequivalence

Pulak Ghosh Dr.* Mithat Gonen†

*Georgia State University, pghosh@mathstat.gsu.edu

†Memorial Sloan-Kettering Cancer Center, gonenm@mskcc.org

Semiparametric Bayesian Modeling of Multivariate Average Bioequivalence

Pulak Ghosh Dr. and Mithat Gonen

Abstract

Bioequivalence trials are usually conducted to compare two or more formulations of a drug. Simultaneous assessment of bioequivalence on multiple endpoints is called multivariate bioequivalence. Despite the fact that some tests for multivariate bioequivalence are suggested, current practice usually involves univariate bioequivalence assessments ignoring the correlations between the endpoints such as AUC and Cmax. In this paper we develop a semiparametric Bayesian test for bioequivalence under multiple endpoints. Specifically, we show how the correlation between the endpoints can be incorporated in the analysis and how this correlation affects the inference. Resulting estimates and posterior probabilities "borrow strength" from one another where the amount and direction of the strength borrowed are determined by the prior correlations. The method developed is illustrated using a real data set.

Semiparametric Bayesian Modeling of Multivariate Average Bioequivalence

Pulak Ghosh*and Mithat Gönen†

May 3, 2006

Abstract

Bioequivalence trials are usually conducted to compare two or more formulations of a drug. Simultaneous assessment of bioequivalence on multiple endpoints is called multivariate bioequivalence. Despite the fact that some tests for multivariate bioequivalence are suggested, current practice usually involves univariate bioequivalence assessments ignoring the correlations between the endpoints such as `AUC` and `Cmax`. In this paper we develop a semiparametric Bayesian test for bioequivalence under multiple endpoints. Specifically, we show how the correlation between the endpoints can be incorporated in the analysis and how this correlation affects the inference. Resulting estimates and posterior probabilities "borrow strength"' from one another where the amount and direction of the strength borrowed are deter-

*Department of Mathematics and Statistics, Georgia State University, Atlanta, GA, 30303-3083, USA; Email: pghosh@mathstat.gsu.edu

†Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, NY 10021;

mined by the prior correlations. The method developed is illustrated using a real data set.

Key words: Average bioequivalence; Crossover design; Gibbs sampling; Mixture of Dirichlet Process prior; Markov Chain Monte Carlo;

1 Introduction

The main objective of a bioequivalence study is to establish if different formulations of the same drug substance are equivalent. A bioequivalence study is thus an experiment to compare a test product (T) to a reference product (R). For an unapproved generic dosage form to be marketed and accepted as therapeutically effective to the innovator product, it must establish bioequivalence with the innovator product, *in vivo*. The determination of bioequivalence trials are of interest to many groups: pharmaceutical companies, insurance companies, prescribing doctors, pharamcists, regulatory authorities, etc. The availability of safe and effective generic drugs is beneficial to the public due to cost considerations. On the other hand, patients should be protected from the failure of treatment and toxicity of drugs via rigorous evaluation methods. In the pharmaceutical industries because the regulatory agencies like US Food and Drug Administration (FDA) allow a generic drug to be marketed only if its manufacturer can demonstrate that the generic product is bioequivalent to the innovator product. According to FDA regulations (2002) a valid statistical evaluation of bioequivalence trial is essential to guarantee the safety and efficacy of the generic drug products. FDA suggests the consideration of average bioequivalence (ABE) (Berger et. al., 1996) for establishing bioequivalence between two formulations, which requires the equivalence between the population means of the reference and test formulations.

A test formulation and a reference formulation are bioequivalent if the bioavailability of the two formulations, which are characteristics of the extent and rate of absorption, are sufficiently close. This concept relies on the fundamental assumption that two formulations are therapeutically equivalent if the bioavailabilities of the two formulations are the same. Bioequivalence is usually studied by administering drug formulation to a subject and measuring concentration of the drug in the blood at pre-specified sampling times. These concentration by time measurements are connected with a polygonal curve. The profiles of the concentration time

curves are compared by means of several pharmacokinetics (PK) parameters. These are, *the area under the blood concentration-time curve* (AUC), *maximum concentration* Cmax and *time to achieve maximum concentration* Tmax. Clearly, these bioavailability metrics are dependent on the formulation and the type of study.

While the AUC is the most accepted measure of absorption rate and bioequivalence is mostly established based on AUC measure, none of the above three measures can be deemed as universally superior to the other. One may be more appropriate for some drugs but not others. The measure Cmax is of importance because some drugs may need to reach a certain level of concentration in order to achieve the desired therapeutic effect. A drawback of Cmax is that it is highly correlated with the AUC, and thus is not a pure measure of absorption rate. Endrenyi, Fritsch and Yan (1991) suggested using the ratio Cmax/AUC instead of Cmax. However, this suggestion does not seem to have been adopted into common practice. Tmax may be a relevant measure for drugs such as analgesics and antibiotics that must reach the peak concentration as soon as possible. However, it may not be a good measure for drugs requiring multiple dosage before a therapeutic effect is observed. Tmax is independent of AUC but power from this measure may be limited because of its discrete nature (Kaniwa, Ogata, Aoyagi, Takeda and Uchiyama, 1989). United States FDA considers both AUC and Cmax as important metric of the rate of absorption.

Most bioequivalence analysis are based on any one of the PK parameters and thus while bioequivalence may be established for one PK parameter it may be bioinequivalent for another PK parameter. Because there is no single PK parameter that is universally accepted as the best, it would be best to consider a test of ABE which includes all the endpoints together. This set up is called multivariate ABE. FDA (1992) and EC-GCP (1993) consider two drugs to be bioequivalent only if they are similar in both AUC and Cmax. Westlake (1988) and Hauck et. al. (1995) have considered the problem of comparing AUC and Cmax simultaneously.

ABE has generated a lot of research in the last few decades, but there is only limited attempts

to address multivariate ABE (Berger and Hsu, 1996). Chinchilli et. al. (1997) proposed an approach to compare test and reference formulation based on two regions, an acceptable region based on reference profile and another region based on test profile. Wang et. al. (1999) developed an intersection union test and proved that their test is uniformly more powerful than the existing tests. Munk et. al. (1999) have also developed a equivariant confidence rule. Recently, Quan et. al. (2001), developed a simultaneous equivalence assessment of therapy and pain data.

All these methods considers that all the endpoints, e.g., `AUC` and `Cmax`, come from a multivariate normal distribution, resulting in correlated parameter estimates and test statistics. This has been the standard approach in frequentist and likelihood approaches to multivariate modeling. One shortcoming is that it cannot account for the correlation between parameters, something which can only be done in the Bayesian paradigm. Despite this fact, many Bayesian approaches to multivariate modeling uses standard independent priors.

There is a subtle difference between correlated parameters and correlated parameter estimates. The former induces "borrowing strength," a phenomenon that causes the estimates of the correlated parameters shrink towards a common point. Borrowing strength cannot be induced through correlated parameter estimates. This is studied in detail by Gönen, Westfall and Johnson (2003) and will be demonstrated in Section 4.

Another drawback of the existing methods is the normality assumption. Normality assumption in a bioequivalence trial may not always hold and subsequent inferences can be misleading (Chow and Tse ,1990; Bolton 1991). There can be a great deal of variability in bioequivalence trials between subjects, and so we may need more flexibility than a parametric model would allow. For example, in the data set we analyze, which is discussed in more detail in section 4, there is a lot of diversity in the `AUC` and `Cmax` values. Histograms of the `AUC` and `Cmax` measures suggest nonnormality in their distributions as well as strong presence of outliers.

Thus, in this paper we extend the existing methods in two ways. First, we assume a multivariate prior on treatment effects that allows positive probability of null hypothesis for each endpoint. Specifically, we calculate how the posterior probability of bioequivalence depends on the prior correlation among hypothesis on each endpoint and also the joint prior probability that all hypothesis are true. Second, in this paper we develop a semiparametric Bayesian test for multivariate ABE. To our knowledge there is no Bayesian work in multivariate ABE in this direction. A Bayesian semiparametric method using a mixture of Dirichlet process (MDP) (MacEachern, 1994; Escober and West, 1995; MacEachern and Müeller, 1998) is thus proposed to relax the distributional assumption and to accommodate possible population heterogeneity. Mixture of DP by far are the most widely used nonparametric Bayesian model mainly because one can easily obtain posterior estimates using standard MCMC approaches such as Gibbs sampling and it is computationally tractable.

The plan of the paper is as follows. Section 2 introduces a semiparametric random effects model for multivariate ABE. In section 3, we describe the Bayesian approach. Section 4 describe the data and the results of the empirical analysis are presented. Section 5 draws conclusions.

2 A Semiparametric Model for Multivariate Bioequivalence

2.1 Model

We will consider a 2×2 crossover design for multivariate ABE endorsed by FDA (1992). In a two-sequence, two-period crossover design with multivariate responses, suppose p endpoints are measured for each of two treatments A and B. In this article the p characteristics are the pharmacokinetic parameters, such as `AUC`, `Cmax` and `Tmax`.

Let y_{ijkp} be the p characteristics on the j th formulation of the i th subject in the k th sequence. Then writing

$$\mathbf{y}_{ijk} = (y_{ijk1}, y_{ijk2}, \cdots, y_{ijkp})^T \tag{1}$$

the model for a multivariate 2×2 crossover design is given by

$$\boldsymbol{y}_{ijk} = \boldsymbol{\mu}_j + \boldsymbol{s}_{ik} + \boldsymbol{\pi}_{(j,k)} + \boldsymbol{\xi}_k + \boldsymbol{e}_{ijk} \tag{2}$$

Thus each $\mathbf{y}_{ijk}$ $(j = T, R;\ k = 1, 2;\ i = 1, 2, \cdots, n_k)$ is a $p \times 1$ response vector. It is important to note that $\mathbf{y}_{ijk}$ might have been subjected to a transform to improve suitbaility of model assumptions. For example in Section 5 we will use the log-transform for `AUC` and `Cmax`. Whether the endpoints are transformed or not,however, has no bearing on what follows, except for prior elicitation.

Here, $\boldsymbol{\mu}_j = (\mu_{j1}, \mu_{j2}, \cdots, \mu_{jp})^T$ is the treatment effect vector for formulation j, $\boldsymbol{s}_{ik} = (s_{ik1}, s_{ik2}, \cdots, s_{ikp})^T$ is the random effect of subject i in sequence k, $\boldsymbol{\pi}_{j,k} = (\pi_{(j,k)1}, \pi_{(j,k)2}, \cdots, \pi_{(j,k)p})^T$ is the fixed effect vector of period administering treatment j in sequence k, $\boldsymbol{\xi}_k = (\xi_{k1}, \xi_{k2}, \cdots, \xi_{kp})^T$ is the vector of fixed effect sequence, and $\mathbf{e}_{ijk} = (e_{ijk1}, e_{ijk2}, \cdots, e_{ijkp})^T$ is the error for subject i of sequence k for formulation j.

Note that the vector of direct effect for treatment j is $\boldsymbol{\mu}_j$. We have two treatments (T and R). We assume that the error terms are normally distributed around $\mathbf{0}$ but we leave the random subject effect distribution unspecified:

$$\boldsymbol{s}_{ik}|(.) \ \sim \ \mathbf{f}(.), \qquad \boldsymbol{e}_{ijk} \sim N_p(\mathbf{0}, \boldsymbol{\Sigma}) \tag{3}$$

We further assume that $\boldsymbol{s}_{ik}$ and $\boldsymbol{e}_{ijk}$ are independent and

$$\boldsymbol{\mu}_T + \boldsymbol{\mu}_R = 0, \quad \boldsymbol{\xi}_1 + \boldsymbol{\xi}_2 = 0, \quad \text{and} \ \ \boldsymbol{\pi}_1 + \boldsymbol{\pi}_2 = 0. \tag{4}$$

The usual distribution of the random effects $\mathbf{f}(.)$ is multivariate Gaussian. However, the choice of normal distribution may not be appropriate, if the data are skewed or contain outliers. To

guard against the influence of outliers and /or skewness, which can be quite influential in typical small-sample bioequivalence studies we replace the usual normality assumption for the random effect with a mixture of Dirichlet process (MDP) prior. The use of MDP prior is also motivated by the following considerations: (1) Bioequivalence endpoints are direct measures of human metabolism and they can exhibit substantial between-subject variability which may not be captured by the normal distribution, (2)The MDP model defines and entire class of distribution and multivariate Gaussian is a special case of it (as described in section later), (3) This generalization has the potential to make the inference robust to departures from a normal distribution while still having good performance if the actual distribution is normal.

The application of MDP for the random effect $\mathbf{s}_{ik}$ using Gibbs sampler has been pioneered by MacEachern (1994), Escobar and West (1995), MacEachern and Müeller (1998). In particular we assume $\boldsymbol{s}_{ik}|(.) \sim f(.)$ with $f(.|G) = \int N(.|\mu, \Omega) dG(\mu)$. Here, $N(.|\mu, G)$ denotes the density of normal distribution with mean μ and variance and covariance matrix Ω. The mixture model f can be equivalently written as a hierarchical model as

$$\boldsymbol{s}_{ik}|G \sim N(\boldsymbol{m}_i, \Omega) \tag{5}$$

$$\boldsymbol{m}_i \sim G \tag{6}$$

The model is completed with a prior probability model for the random distribution G. We assume $G \sim DP(\alpha, G_0)$, a Dirichlet process (DP) prior with precision parameter α and baseline distribution G_0. For the baseline distribution G_0 we assume a $N(0, \Delta)$ distribution. This is a standard specification for MDP priors.

MDP prior assumes that the prior distribution G itself is uncertain drawn from a Dirchlet process. The parameters of a Dirichlet process are G_0 a probability measure, and α, a positive scalar assigning mass to the real line. The parameter G_0 is a distribution that approximates the true nonparametric shape of G. The concentration parameter α reflects our prior belief

about how similar G is to G_0. Large values of α lead to a G that is very close to G_0. Small values of α allow G to deviate more from G_0 and put most of its probability mass on just a few atoms. MDP prior allows for heterogeneity, outliers and skewness as desired. It also includes a multivariate normal prior as a limiting case for $\alpha \to \infty$. One can assume a preassigned value for α or may specify a prior distribution on α to capture the uncertainty and sample α from posterior distribution using Gibbs sampler (Brown and Ibrahim, 2003; Escober and West, 1995).

There are several ways to implement a MDP prior. Following Sethuraman (1994) one way to generate the DPP prior is to regard the infinite dimensional parameter G as an infinite mixture. Recent research, however, has focussed on using the constructive definition of the Dirichlet process to produce MCMC algorithms (Ishwaran and James, 2002; Ishwaran and Zarepour, 2002). This includes a finite approximation for MDP (Ishwaran and Zarepour, 2002) can be used within the software `WinBUGS` (Spiegelhalter et. al., 2003) to implement Gibbs sampling. In fact, for our data analysis we use the finite approximation to MDP to implement the Gibbs sampling in `WinBUGS`. This finite approximation for MDP can be done by introducing latent variables $\boldsymbol{I} = (I_1, J_2, \cdots, I_n)$ which indicate the group membership for the unobserved variables m_I along with a probability vector $\boldsymbol{w} = (w_1, w_2, \cdots, w_L)^T$. Thus, the model can be written as:

$$
\begin{aligned}
\boldsymbol{s}_{ik}|\boldsymbol{m}_i, \Omega &\sim N(\boldsymbol{m}_{I_i}, \Omega) && (7)\\
I_i|\boldsymbol{w} &\sim \text{Multinomial}(\{1, 2, \cdots, L\}, \boldsymbol{w}) && (8)\\
\boldsymbol{m}_l|G &\sim G, \quad l = 1, 2, \cdots, L && (9)\\
G|\alpha, G_0 &\sim \text{DP}(\alpha G_0) && (10)\\
G_0|\Delta &\sim N(0, \Delta) && (11)\\
\boldsymbol{w} &\sim \text{Dirichlet}(\frac{\alpha}{L}, \frac{\alpha}{L}, \cdots, \frac{\alpha}{L}) && (12)\\
\Delta &\sim IW(q, \Phi) && (13)\\
\Omega &\sim IW(\xi, R) && (14)\\
\alpha &\sim \text{Gamma}(a_\alpha, b_\alpha). && (15)
\end{aligned}
$$

Here $IW(n, A)$ denotes a inverse-Wishart distribution with scalar parameter n and matrix parameter A, and Gamma(a, b) denotes a Gamma distribution with shape parameter a and scale parameter b, parameterized such that the expected value is a/b.

2.2 Hypotheses of Interest

We consider the ABE assessment in terms of difference of the averages in the treatment effects. Let $\boldsymbol{\theta} = \boldsymbol{\mu}_T - \boldsymbol{\mu}_R$. Then $\boldsymbol{\theta} = \{(\theta_l)\}$ $(l = 1, \cdots, p)$ are the true mean differences between the two treatments for the p endpoints. Then in order to assess ABE on each end point we test the following hypothesis:

$$
H_{0l} : \theta_l \leq \delta_l \quad \text{or} \quad \theta_l \geq \gamma_l \quad \text{vs.} \quad H_{1l} : \delta_l < \theta_l < \gamma_l, \quad , l = 1, \cdots, p. \tag{16}
$$

where δ_l, γ_l are pre-specified equivalence bounds set up by regulatory agencies like FDA. ABE is established on all the p endpoints simultaneously if the following hypothesis is correct.

$$
H_1 : \delta_1 < \theta_1 < \gamma_1 \quad \text{and } \delta_2 < \theta_2 < \gamma_2 \cdots \quad \text{and } \delta_p < \theta_p < \gamma_p \tag{17}
$$

For convenience, we assume $\delta_1 = \delta_2 = \cdots = \delta_p = \delta$ and $\gamma_1 = \gamma_2, \cdots, \gamma_p = \gamma$. FDA guidelines recommend $\delta = \log(0.80)$ and $\gamma = \log(1.25)$ and that is what we will use in Section 4.

3 Bayesian Inference

3.1 Prior Distribution

We will start by placing a multivariate normal prior on $\boldsymbol{\mu}_T$ with mean $\mathbf{b}_1$ and variance $\mathbf{B}_1$. Since $\boldsymbol{\mu}_T$, is the parameter of interest the choice of $\mathbf{b}_1$ and $\mathbf{B}_1$ is of great importance to inference and we will explain in detail in Section 4, in the context of an example, how traditionally available information in a bioequivalence trial can be used to specify these prior parameters.

In a similar vein, the prior for $\boldsymbol{\xi}_1$ is multivariate normal with mean $\mathbf{b}_2$ and variance $\mathbf{B}_2$ and $\boldsymbol{\pi}_1$ is also given a multivariate prior with mean $\mathbf{b}_3$ and variance $\mathbf{B}_3$. While the model is developed for the generic prior, in most cases the variance-covariance matrices $\mathbf{B}_2$ and $\mathbf{B}_3$ can be taken a diagonal matrix. These fixed-effect covariates, though as they may be correlated, are not the focus of investigation and it is unlikely that the analysts will have sufficient information to elicit a covariance matrix with non-zero off-diagonals. For the same reason we will take the prior of $\boldsymbol{\xi}_1$ and $\boldsymbol{\pi}_1$ to be independent of one another. The prior for the within-subject variance, $\boldsymbol{\Sigma}$ are assumed to be of conjugate form. The following summarizes the prior structure we impose on this problem:

$$\begin{aligned}
\boldsymbol{\mu}_T &\sim N_p(\mathbf{b}_1, \mathbf{B}_1) && (18)\\
\boldsymbol{\xi}_1 &\sim N_p(\mathbf{b}_2, \mathbf{B}_2) && (19)\\
\boldsymbol{\pi}_1 &\sim N_p(\mathbf{b}_3, \mathbf{B}_3) && (20)\\
\boldsymbol{\Sigma}^{-1} &\sim W(\eta, R) && (21)\\
& && (22)
\end{aligned}$$

with the additional constraint that $\boldsymbol{\mu}_T \perp \boldsymbol{\xi}_1 \perp \boldsymbol{\pi}_1 \perp \boldsymbol{\Omega} \perp \boldsymbol{\Sigma}$.

Although this prior specification is sufficient to derive the posterior distribution it is instructive to examine the implications on $\boldsymbol{\theta}$ and the hypotheses in (16,17). Since $\boldsymbol{\theta} = 2\boldsymbol{\mu}_T$ the implied prior for $\boldsymbol{\theta}$ is

$$\boldsymbol{\theta} \sim N_p(2\mathbf{b}_1, 4\mathbf{B}_1).$$

which can be used to express the prior probability of the individual hypotheses

$$P(H_{0l}) = P(I(\theta^{(l)} \leq \delta_l \cup \theta^{(l)} \geq \gamma_l))$$
$$= 1 - \Phi\left(\frac{\gamma_l - (2\mathbf{b}_1)^{(l)}}{(4\mathbf{B}_1)^{(ll)}}\right) + \Phi\left(\frac{\delta_l - (2\mathbf{b}_1)^{(l)}}{(4\mathbf{B}_1)^{(ll)}}\right)$$

where the superscript indexes the elements of the corresponding vector or the matrix.

Furthermore one can examine the *joint* prior distribution of the hypotheses H_{0l} which form a multivariate array of Bernoulli variates with tetrachoric correlation that is a function of $\mathbf{b}_1$ and $\mathbf{B}_1$. Their joint distribution can be assessed in terms of the prior of $\boldsymbol{\mu}_T$ as

$$P(\{H_{0l}\}_l = \{r_l\}_l) = \int \cdots \int_{\Upsilon_{\{r_l\}_l}} \phi_k(\boldsymbol{\mu}_T; \mathbf{b}_1, \mathbf{B}_1)\, d\boldsymbol{\mu}_T$$

where $\phi_k(.;.,.)$ is the multivariate normal density, r_l is a 0-1 index indicating whether $H_{(0l)}$ is true or not and $\Upsilon_{\{r_l\}_l}$ is the appropriate subset of $\Re^p$ on which the joint distribution is to be computed. Specifically, let

$$\Upsilon_{\{r_l\}_l} = \bigcup_{\{r_l\}_l} \Upsilon^{r_l}$$

where

$$\Upsilon^{r_l} = \begin{cases} \delta_l < \boldsymbol{\mu}_T^{(l)} < \gamma_l, & r_l = 1 \\ \boldsymbol{\mu}_T^{(l)} > \gamma_l \cup \boldsymbol{\mu}_T^{(l)} < \delta_l, & r_l = 0 \end{cases} \tag{23}$$

We will discuss the practical implications of the choice of prior parameters in the context of an example in section 4. It is important note here that diagonal elements of $\mathbf{B}_1$ plays an

important role. They represent the correlation between the endpoints *a priori* and they will allow one endpoint to "borrow strength" from the other. This cannot be accomplished in a frequentist framework. Traditional mixed-effect models allow for modeling the correlation between test statistics, but not between endpoints themselves and thus they cannot induce borrowing strength. For a more detailed elucidation of this point along with other examples, see Gönen, Westfall and Johnson (2003).

3.2 Gibbs Sampling

The posterior distributions are analytically intractable and thus computations are done via Monte Carlo approximations with the help of the MCMC method. The Gibbs sampler is one of the most widely used MCMC methods, and is implemented in the software package `WinBUGS`. It works by drawing samples from the posterior distribution of the parameters in such a way that sample draws are made from the conditional posterior distributions of univariate parameters given the most recent draws of the other parameters. Thus, what is required for the Gibbs sampler to work is the ability to sample from the full conditional posterior distribution of the parameters. The conditional distribution of all the parameters are obtained from the joint distribution of all the parameters. We omit the explicit expression of the conditional distribution as `WinBUGS` does not require their explicit specification.

3.3 Model Assessment

Since we used a MDP distribution for random effects instead of a usual Gaussian distribution, it is interesting two compare the two models. Thus, we compare the following two models:

Model 1: This is the model with the random effect $\mathbf{s}_{ik}$ following a multivariate normal distribution, i.e., $\boldsymbol{s}_{ik} \sim N_p(0, \Sigma)$.

Model 2: This is the model we considered, i.e., $\mathbf{s}_{ik}$ have a MDP distribution.

We examine two statistics for comparing these models, the Deviance Information Criterion (DIC; Spiegelhalter et. al., 2002) and Conditional Predictive Ordinate (CPO; Gelfand, Dey, and Chang, 1992). DIC is the Bayesian equivalent of the AIC , particularly suitable for hierarchical models. The DIC is given by

$$\text{DIC} = \hat{D} + 2p_D$$

where $\hat{D}$ is the deviance evaluated at the posterior mean $\bar{\theta}$, and p_D is the effective number of parameters in the model, which is given by $p_D = \bar{D} - \hat{D}$, where $\bar{D}$ is the posterior mean deviance. DIC can be computed based on the MCMC sample, and is reported automatically by `WinBugs`. The interpretation of DIC is similar to that of the AIC, as a single-number summary of the relative fit between the model and the "true model" generating the data, from the perspective of prediction, conditional on the clusters in the hierarchy, e.g., the subjects in the study. The smaller the DIC the better the fit, and, in analogy with AIC, a difference larger than 10 is overwhelming evidence in favor of the better model (Burnham et. al., 2002).

We also calculate the conditional predictive ordinate (CPO) (Gelfand, Dey, and Chang, 1992) for each observation. Chen, Shao, and Ibrahim (2000, chapter, 10) show in detail how to obtain Monte Carlo estimates of the CPO statistic. We compare these two different models using the *log pseudo marginal likelihood* (LPML). Define $\widehat{CPO}_i$ to be the Monte Carlo estimate of the ith subject's CPO statistic. Models with greater LPML $= \sum \log(\widehat{CPO}_i)$ values will indicate a better fit.

4 Data Analysis

4.1 Data

In this section we illustrate the above methods using a data set from Clayton and Leslie (1981). Two erythromycin formulations were compared using the two-sequence, two-period, two-treatment crossover design involving 18 subjects. The two formulations were, A : Erythromicin stearate [Erthrocin (R), 500mg, ovalid tablets, 6316, Abbott australasia Private Limited] and B : Erythromycin base [Eryc (R), 2×250mg capsules, containing enteric coated pellets]. The primary concern was to compare the bioavailability of formulations A, B when each was administered immediately after food. Generally speaking, a stearate formulation will exhibit more rapid absorption as compared to a base formulation. Nine subjects were randomized to a formulation sequence AB while another nine subjects were randomized to sequence BA. A one week washout period separated the two periods. Blood samples were taken immediately before drug administration and then at 0.5, 1, 1.5, 2, 3, 4, 6 and 8 hours after drug administration. The primary variables of interest `AUC` and `Cmax` was calculated from 0 to 8 hours.

Figures 1 and 2 give the histograms for period differences for the two endpoints. Once can identify a few outliers by visual inspection. The distribution of the endpoints for each treatment are given in Figure 3 assuming no period differences and hint that there may be a treatment difference. We also observe that, even after the log-transformation, there is strong evidence against the normality assumption. Finally, Figure 4 is the scatter plot between the two endpoints for the two treatments, making clear the strong correlation between the endpoints.

4.2 Prior Elicitation

Since there are only two endpoints `AUC` and `Cmax`, we set $p = 2$. We choose a relatively weak prior for the parameters other than the treatment effect. In particular, we take $\boldsymbol{\xi}_2 \sim N_2\{(0,0)', \text{diag}(0.001, 0.001)\}$, $\boldsymbol{\pi}_2 \sim N_2\{(0,0)', \text{diag}(0.001, 0.001)\}$,

$$\boldsymbol{\Omega} \sim IW\left(2, \begin{pmatrix} 0.1 & 0 \\ 0 & 0.1 \end{pmatrix}\right), \boldsymbol{\Sigma} \sim IW\left(2, \begin{pmatrix} 0.1 & 0 \\ 0 & 0.1 \end{pmatrix}\right), \boldsymbol{\Delta} \sim IW\left(2, \begin{pmatrix} 0.1 & 0 \\ 0 & 0.1 \end{pmatrix}\right)$$

A Gamma$(0.01, 0.01)$ is assumed for α. We tried various values of L and found $L = 5$ works very well.

The choice of $\boldsymbol{b}_1$ is usually facilitated by the specific null hypothesis. It is commonly accepted that a clinical trial is justified only if the prior probability of the null and alternative hypotheses are comparable, otherwise it would be unethical to subject humans to experimentation. In addition, if the two products are not bioequivalent there is no a priori reason to anticipate the direction of bioinequivalence for either of the endpoints, at least in this example. These considerations suggest $\boldsymbol{b}_1 = (0, 0)$. The diagonal of $\boldsymbol{B}_1$ can be elicited using the following constraint:

$$P(H_{0l}) = P(\theta_l \leq \delta_l \quad \text{or} \quad \theta_l \geq \gamma_l) = 0.5$$

Using the logarithmic scale equivalent of FDA mandated bioequivalence limits of $\delta = \log(0.80)$ and $\gamma = \log(1.25)$, setting

$$\boldsymbol{B}_1 = \begin{bmatrix} 0.055 & \sigma_{12} \\ \sigma_{12} & 0.055 \end{bmatrix}$$

ensures that prior probability for each of the bioequivalence hypothesis is 0.5.

Eliciting a value for σ_{12}, or ρ, the implied prior correlation is more difficult. One can presumably research the literature for all relevant multivariate average bioequivalence studies and record whether the hypothesis of bioequivalence is rejected for `AUC` and `Cmax` and use the empirical correlation. If there is available a rich literature on the subject with detailed reporting

of individual endpoints, then such an undertaking with the collaboration of PK experts would be preferable.

An alternative which we used here is to try different values of ρ and produce a sensitivity analysis. There is no reason to consider $\rho < 0$; the nature of the endpoints dictate that if there is any correlation it will be positive. To choose an upper bound for ρ for the sensitivity analysis it is instructive to consider what the correlation between a pair of binary variables will be if they are concordant for a certain percentage of the time. In this exercise the binary variables represent the two elements of the hypotheses of multivariate bioequivalence. A simple calculation yields that if two binary variables are concordant 75% of the time, then their correlation will be approximately 50%. This assumes marginal probabilities of 50% for each binary variable, the scenario which yields the maximum correlation. It seems reasonable for this example, to take $\rho = 0.5$ as the upper bound of the range for sensitivity analysis. However, if one is interested in more than 75% concordance the same logic can be applied to find an upper bound of ρ.

4.3 Results

We compared the semiparametric model (Model 2) with the simple parametric model (Model 1) using the deviance information criterion (DIC, Spiegelhalter *et al.* , 2002). The LPML are -311.45 and -259.44 for model 1 and 2 respectively. Also, the DIC for the two models are 251.16 and 221.99 respectively. Since, the semiparametric model has the largest LPML and smalles DIC, we conclude that for this data set the model with MDP is the better fitting model.

Thus, we report here the results based on semiparametric model. The posterior estimates of the fixed effects parameters are given (for $\rho = 0.3$) in Table 1. Treatment effect under `AUC` is quite high while it is small in `Cmax`. The treatment effects are also significant in both the endpoints as the 95% posterior interval does not contain 0. The estimate of the variance-

covariance matrix Σ is found to be $\hat{\Sigma} = (8.4, 2.521, 2.521, 0.913)$. This is a reflection of the high variability in `AUC`.

The Bayesian hypothesis test requires calculating the posterior probability of the hypotheses described in (17). Thus, the posterior probability of ABE for single endpoint, e.g., lth endpoint can be computed using the following equation:

$$\begin{aligned} \Pr[\text{ABE/Data}] &= \Pr[\log(0.8) < \theta_l < \log(1.25)/\text{Data}] \\ &= \frac{1}{m}\sum_{c=1}^{m} I[\log(0.8) < \theta_{l_c} < \log(1.25)/\text{Data}] \end{aligned}$$

where, $(\theta_{l_c} : c = 1, 2, \cdots, m)$ is a sample from the observed posterior density of θ_l, $I(.)$ denotes the indicator function, and $m = 10,000$ is the number of iterations.

In a similar way, we can evaluate the posterior probability of ABE on multiple endpoints simultaneously, as follows:

$$\begin{aligned} \Pr[\text{Joint ABE/Data}] &= \Pr[\log(0.8) < \theta_1 < \log(1.25), \cdots, \log(0.8) < \theta_p < \log(1.25)/\text{Data}] \\ &= \frac{1}{m}\sum_{c=1}^{m} I[\log(0.8) < \theta_{1_c} < \log(1.25), \cdots, \log(0.8) < \theta_{p_c} < \log(1.25)/\text{Data}] \end{aligned}$$

Table 2 presents the posterior probability of ABE on each of the two endpoints as well as the joint posterior probability for various values of ρ, the correlation corresponding to six different specifications of σ_{12}. Although the development in the previous sections has formulated $\boldsymbol{B}_1$ in terms of the covariance parameters we find the correlation more useful in terms of presentation.

The "borrowing strength" effect, briefly discussed in Section 3.1, for the treatment effect under AUC (tAUC), Cmax (tCmax) and the posterior probability of bioequivalence (pAUC, pCmax) is clear. When the two endpoints are deemed independent there is a substantial difference with respect to AUC, but a moderate one for Cmax. This is also reflected in the posterior

probabilities. We note that the frequentist threshold of 0.05 plays no role in interpreting posterior probabilities and we think the relevant threshold is 0.5, suggesting that one should retain the hypothesis with the higher posterior probability. Therefore an analyst assuming independence might conclude that the two treatments are bioequivalent for Cmax, but not for AUC. We note that using $\rho = 0$ is for comparative purposes only. We think any reasonable Bayesian analysis of this model should use $\rho > 0$.

As one increase the correlation, treatment effects shrink towards a common point and so do the posterior probabilities. Despite this shrinkage they remain substantially different even at high levels of correlation suggesting that the treatment difference is more pronounced for `AUC` than it is for `Cmax`. Notably, when one introduces a correlation the bioequivalence hypothesis is no longer retained for `Cmax`. The fact that the two treatments are so strongly bioinequivalent for `AUC`, along with a moderate prior correlation between `AUC` and `Cmax`, reduces the posterior probability of bioequivalence from 0.610 to 0.418. We also observe that posterior probability of multivariate ABE increases with increasing ρ. As ρ gets near 1, posterior probabilities of the two individual hypotheses and the multivariate ABE all converge to the same point.

5 Discussion

Multivariate response occurs frequently in clinical trials using crossover designs. Bioequivalence, or perhaps more to the point, bioavalibility, is essentially a multivariate endpoint and any empirical assessment of bioequivalence, just as in any Bayesian evaluation of multiple endpoints, has to take into account the prior correlation between endpoints. While there is substantial literature on the Bayesian analysis of multivariate endpoints using independent priors we believe that they have no place in the assessment of bioequivalence and one should expect the bioequivalence endpoints to bear a moderate prior correlation. In this article we showed how such an analysis can be carried out for average bioequivalence, including the analysis of a real

data set, commenting extensively on prior elicitation and sensitivity analysis. One should note that only in a Bayesian framework it is possible to formally incorporate the correlation between endpoints. It is clear that a non-Bayesian model can only incorporate the correlation between the observed test statistics and cannot induce a borrowing strength of the type observed in Section 4. It is also clear that specifying a diagonal $\mathbf{B}_1$ leads to a similar problem and undermines the strength of a Bayesian approach. Berger (2003), citing Gönen, Westfall and Johnson (2003), points out the importance of carefully calibrating the prior probabilities of multiple hypotheses.

The magnitude of the prior correlation is difficult to elicit, but a sensitivity analysis like in Table 1 can produce convincing conclusions that can sometimes contradict from traditional Bayesian analysis assuming prior independence. This has important implications for practitioners in the pharmaceutical industry who operate in a tightly-regulated environment. We have developed a semiparametric Bayesian approach for assessing bioequivalence on multiple endpoints. Our method can be generalized easily to more general cases. As FDA becomes more accepting of Bayesian analysis, the practice of agreeing on a prior (or a few of them) before the clinical trial is slowly emerging and it certainly should be considered in the case of bioequivalence.

References

[1] Berger, J. O., (2003), Rejoinder to "Fisher Lecture: Could Fisher, Jeffreys and Neyman have agreed on testing (with discussion). *Statistical Science* **18**, 1–32.

[2] Berger, R. L., and Hsu, J. C. (1996), Bioequivalence trials, intersection union tests and equivalence confidence sets (with discussion). *Statistical Science* **11**, 283–319.

[3] Bolton, S. (1991). Outliers-examples and opinions. Presented at *Bioavailability/Bioequivalence: Pharmacokinetic and Statistical Considerations* sponsored by Drug Information Association, August, 1991, Bethesda, Maryland.

[4] Brown, E. R., and Ibrahim, J. G. (2003). A Bayesian Semiparametric Joint Hierarchical Model for Longitudinal and Survival Data. *Biometrics* **59**, 221-228.

[5] Burnham, K. P. and Anderson, D. R. (2002). *Model Selection and Multivariate Inference: A Practical Information- Theoretic Approach*, 2nd ed. Springer: New York, 2002.

[6] Chen, M.-H., Shao, Q.-M., and Ibrahim, J. G. (2000). *Monte Carlo Methods in Bayesian Computation.* New York: Springer.

[7] Chinchilli, V. M. and Elswick, R. K. (1997). The Multivariate assessment of bioequivalence. *Journal of Biopharmaceutical Statistics* **7**, 113-123.

[8] Chow, S. C. and Tse, S. K. (1990). Outlier detection in bioavailability/bioequivalence studies. *Statistics in Medicine* **9** , 549-558.

[9] Clayton, D. and Leslie, A. (1981), The bioavailability of erythromycin stearate versus enteric-coated erythromycin base when taken immediately before and after food. *Journal of International Medical Research* **9**, 470-477.

[10] EC-GCP (1993). Biostatistical Methodology in Clinical Trials in Applications for Marketing Authorization for Medical Products. CPMP Working Party on Efficacy of Medical Products, Commission of the European Communities, Brusseles. (Draft guideline edition.)

[11] Endrenyi, L., Fritsch, S., Yan, W. (1991), (C_{max})/AUC is a clearer measure that (C_{max}) for absorption rates in investigations of bioequivalence. *Int J Clin Pharmacol Ther toxicol* **29(10)**, 394-399.

[12] Escober, M. D. and West, M. (1995). Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association* **90** , 577-580.

[13] Food and Drug Administration (1992). Bioavailability and bioequivalence requirements. In U. S. Code of Federal Regulations **21**, Chapter 320. U. S. Government Printing Office, Wsahington, D. C.

[14] Food and Drug Administration (2002). Bioavailability and Bioequivalence studies for orally administered drug products-general considerations. U. S. Department of Health and Human Services, FDA, Center of Drug Evaluation and Research (CDER), Rockville, Maryland. (http://www.fda.gov/cder/guidance).

[15] Gelfand, A. E., Dey, D. K., and Chang, H. (1992). Model determination using predictive distributions with implementation via sampling-based methods (with discussion). In *Bayesian Statistics 4*, J. M. Bernardo, J. O. Berger, A. P. Dawid, and A.F.M. Smith (eds). Oxford: Oxford University Press.

[16] Gönen, M, Westfall, P. H, and Johnson, W. Bayesian multiple testing for two-sample multivariate endpoints. *Biometrics* **59**, 76-82.

[17] Hauck, W. W., Hyslop, T., Anderson, S., Bois, F. Y. and Tozer, T. N. (1995). Statistical and regulatory considerations for multiple measures in bioequivalence testing. *Clinical Research and Regulatory Affairs* **12**, 249-265.

[18] Ishwaran, H. and James, L. (2002). Dirichlet process computing in finite normal mixtures: smoothing and prior information. *Journal of Computational and Graphical statistics* , 508-532.

[19] Ishwaran, H. and Zarepour, M. (2002). Dirichlet prior sieves in finite normal mixtures. *Statistica Sinica* **12** , 941-963.

[20] Kaniwa, N., Ogata, H., Aoyagi, N., Takeda, Y., Uchiyama, M. (1989), Power analysis of moment analysis parameter in bioequivalence tests. *Journal of Pharmaceutical Sciences*, **78(12)**, 1020-1024.

[21] MacEachern, S. N. (1994). Estimating normal means with a conjugate style Dirichlet process prior. *Communications in Statistics: Simulation and Computation* **23** , 727-741.

[22] MacEachern, S. N. and Müeller, p. (1998). Estimating mixtures of Dirichlet process models. *Journal of Computational and Graphical Statistics* **7** , 223-338.

[23] Munk, A. and Pfluger, R. (1999), $1-\alpha$ equivariant confidence rules for convex alternatives are $\alpha/2$ level tests- with applications to the multivariate assessment of bioequivalence. *Journal of the American Statistical Association* **94**, 1311-1319.

[24] Quan, h., Bolognese, J., and Yuan, W. (2001). Assessment of equivalence on multiple endpoints. *Statistics in Medicine* **20**, 3159-3173.

[25] Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica* **4** , 639-650.

[26] Spiegelhalter, D.J., Best, N.G., Carlin, B.P. and van der Linde, A. (2002). Bayesian Measures of Model Complexity and Fit (with discussion). *Journal of the Royal Statistical Society*, B, **64**, 583–639.

[27] Spiegelhalter, D., Thomas, A., Best, N. and Lunn, D. (2003). "WinBUGS User Manual, Version 1.4", MRC Biostatistics Unit, Institute of Public Health and Department of Epidemiology & Public Health, Imperial College School of Medicine, available at `http://www.mrc-bsu.cam.ac.uk/bugs`.

[28] Wang, W., Hwang, J. T. G. and Dasgupta, A. (1999). Statistical tests for multivariate bioequivalence.*Biometrika* **2** , 395-402.

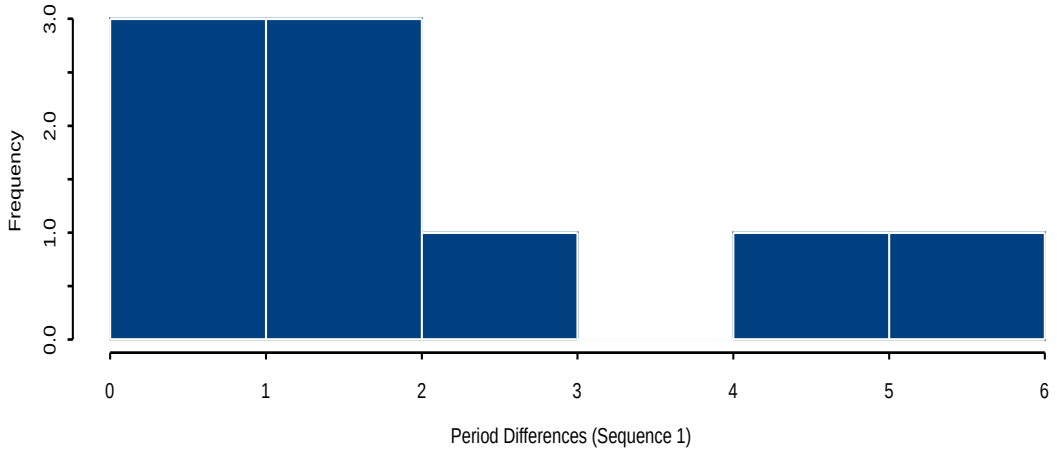
[29] Westlake, W. J. (1988), Bioavailability and bioequivalence of pharmaceutical Formulations. *In Biopharmaceutical Statistics for Drug Development. Karl Peace, editor. New York: Marcel Dekker, Inc.*.

Parameter	AUC	Cmax
Sequence	1.085	-0.0812
Period	1.129	0.4402
Treatment	2.027	0.790

Table 1: Posterior Parameter Estimates of the fixed effects for the two endpoints

ρ	tAUC	pAUC	tCmax	pCmax	p(AUC and Cmax)
0	3.107	0.026	0.64	0.610	0.081
0.1	2.877	0.027	0.664	0.528	0.096
0.2	2.161	0.057	0.785	0.498	0.232
0.3	2.027	0.059	0.790	0.452	0.241
0.4	1.821	0.062	0.880	0.430	0.252
0.5	1.790	0.065	0.884	0.418	0.279

Table 2: Treatment Effect and Posterior Probability of ABE for AUC (tAUC, pAUC), Cmax (tCmax, pCmax) and AUC and Cmax simultaneously, for a variety of the values of the prior correlation between the endpoints



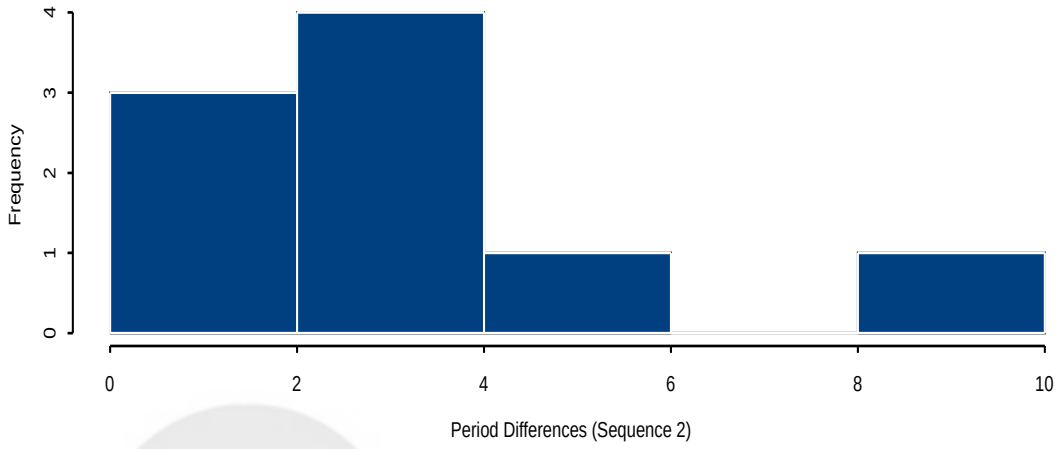
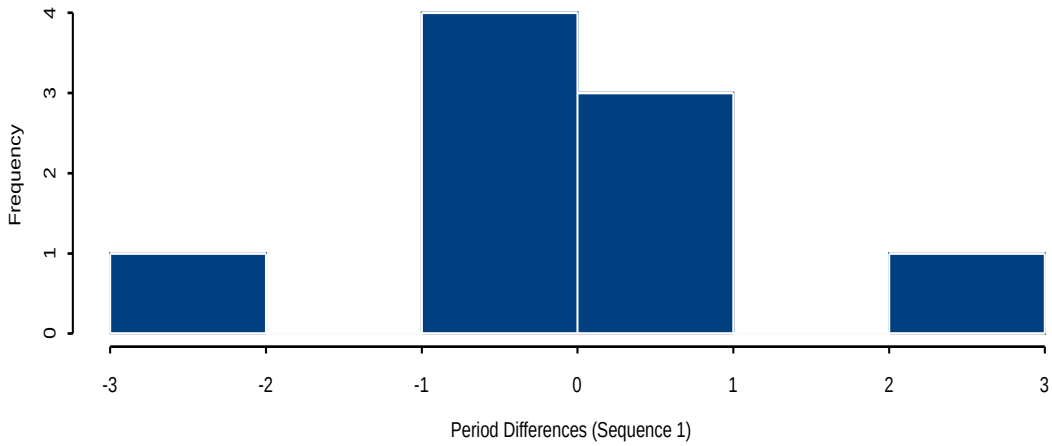


Figure 1: Histogram of Period differences under AUC



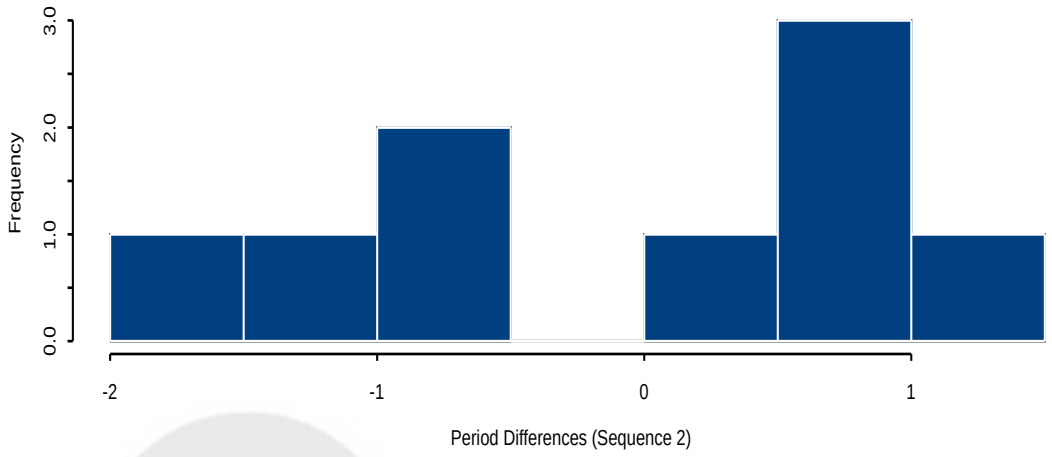


Figure 2: Histogram of Period differences under CMAX

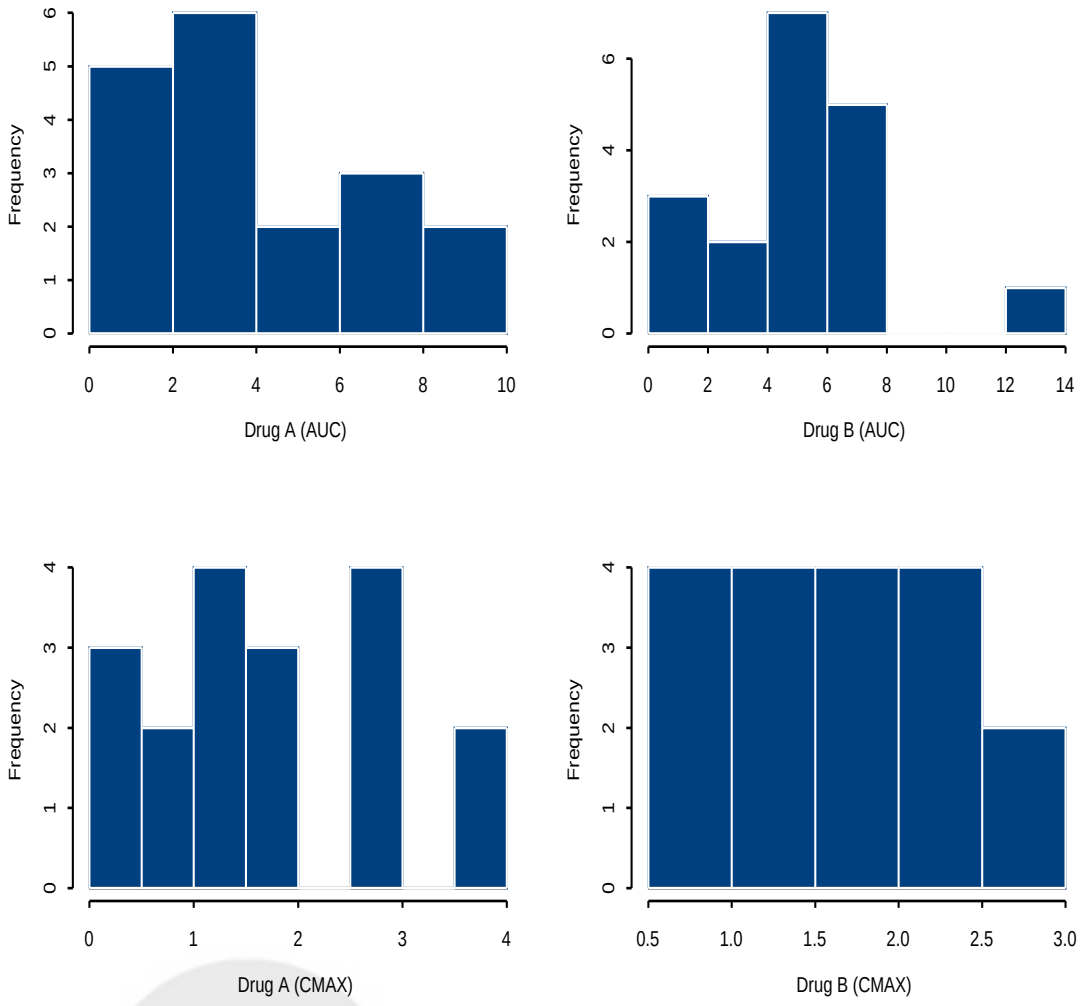
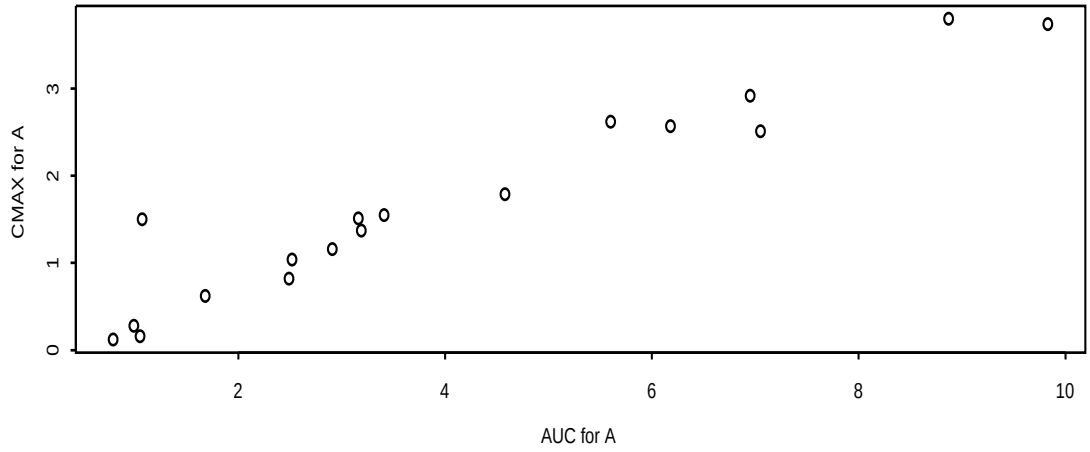


Figure 3: Histogram of two formulation ignoring period effects



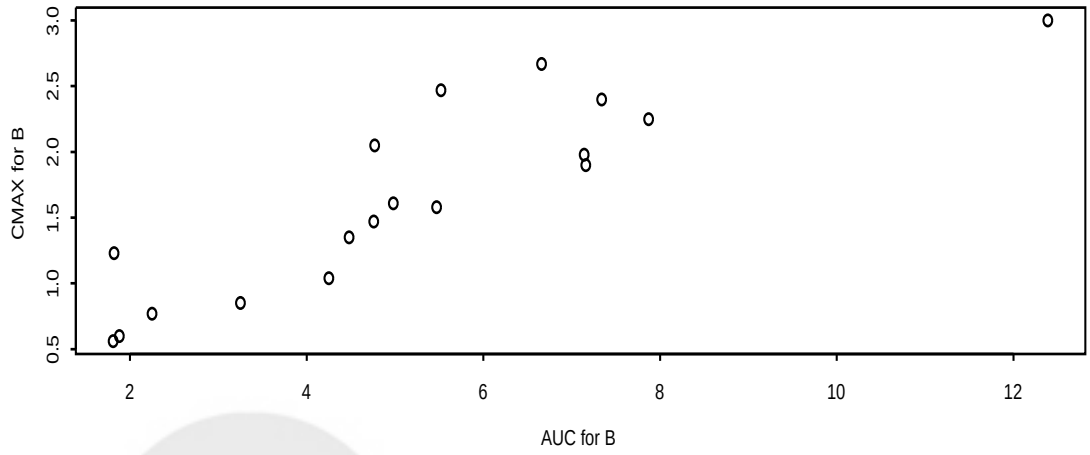


Figure 4: Scatter plot of the two formulations