STATISTICS IN MEDICINE
Statist. Med. 2009; **28**:3567–3579
Published online 16 September 2009 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/sim.3726

Multiple testing for bioequivalence with pharmacokinetic data in 2×2 crossover designs

Yuh-Ing Chen*,† and Chi-Shen Huang

Institute of Statistics, National Central University, Jhongli 32054, Taiwan

SUMMARY

To evaluate globally the average bioequivalence of a test drug to a reference drug in a pharmacokinetic (PK) study under a 2×2 crossover design, we consider directly comparing the associated drug concentration–time curves. Statistical models for the drug concentrations are suggested when the concentrations measured at different time points are distributed according to a generalized gamma distribution and the mean concentrations over time is described by a one-compartment PK model. A multiple test based on the supreme distance between the two curves over the time interval under study is then proposed for testing the equivalence of the two drug concentration–time curves. The results of a Monte Carlo study suggest that, comparative to the conventional univariate and bivariate tests, the proposed test is more powerful for detecting the global bioequivalence and superior on maintaining its level when the global bioequivalence is violated. The application of the proposed tests is finally illustrated by using the data in a PK study involving two brands of benzbromarone tablets for reducing the uric acid. Copyright © 2009 John Wiley & Sons, Ltd.

KEY WORDS: bioequivalence test; crossover design; generalized gamma distribution; multiple test; pharmacokinetic study

1. INTRODUCTION

In a pharmacokinetic (PK) study, to claim a test drug under study as a generic drug, proof of the bioequivalence between the test drug and a comparative reference drug is needed. To do so, some healthy volunteers are recruited and administered with the two drugs in a 2×2 crossover

*Correspondence to: Yuh-Ing Chen, Institute of Statistics, National Central University, Jhongli 32054, Taiwan.
†E-mail: ychen@stat.ncu.edu.tw

Contract/grant sponsor: National Science Council of Taiwan; contract/grant number: NSC 97-2118-M-008-003-MY2

Received 11 August 2008
Accepted 4 August 2009

design with a reasonable wash-out time period, where the volunteers in one sequence receive the reference (R) drug and then the test (T) drug in two different periods, while the volunteers in the other sequence take the drugs in reverse order in the two periods. After the drug is administered to each volunteer, the drug concentrations in blood or plasma at different time points are then measured, which is referred to as the drug concentration–time curve or profile. The average bioavailability parameters such as the area under the drug concentration–time curve (AUC) and the maximum drug concentration ($C_{\max}$) are conventionally of interest for assessing the bioequivalence of the test drug to the reference drug [1, 2]. For example, Berger and Hsu [3] suggested a nearly unbiased test for the equivalence of a certain bioavailability parameter. Multivariate bioequivalence tests [3, 4] also received extensive attention for the equivalence of several bioavailability parameters, simultaneously. In particular, Berger and Hsu [3] recommended an intersection–union test in which nearly unbiased tests for the equivalence of both the AUC and $C_{\max}$ are incorporated.

However, it is well known that the equivalence of one or some bioavailability parameters does not necessarily imply that the two drug concentration–time curves are equivalent [5]. Hence, Liao [6] compared the drug concentration–time curves directly when the two curves satisfy a functional linear model. Nevertheless, nonrejection of the null hypothesis in Liao [6] that the two curves are equal does not provide any significant evidence for supporting the bioequivalence of the two drugs. Therefore, testing procedures are still needed for testing the bioequivalence of the test and reference drugs in terms of the concentration–time curves.

Note that all of the aforementioned tests are constructed under the assumption of lognormal distribution for the drug concentrations or estimators of bioavailability parameters. However, it has been well recognized that the lognormal distribution is of little practical use in the PK study. Therefore, we consider herein the situation where the drug concentration is distributed according to a generalized gamma distribution [7], which includes some well-known right-skewed distributions such as gamma, Weibull and lognormal.

In fact, when the kinetics of the drug under study is fully understood, we may use some suitable compartment models [8] to describe the mean drug concentrations over time. Therefore, to make a parametric bioequivalence test, we consider, in this paper, a statistical model for the drug concentrations where the mean drug concentrations follow a one-compartment PK model and the error variable is distributed according to a generalized gamma distribution. Note that the statistical model is actually an extension of the work in Lindsey *et al.* [9] and Salway and Wakefield [10]. Based on the statistical model, we then construct a multiple test searching for the evidence for the equivalence of the two mean drug concentration–time curves over the time interval under study.

In Section 2, we propose a statistical model for the drug concentrations obtained from a 2×2 crossover study and show how to find the maximum likelihood estimators of the related parameters. We also discuss the goodness-of-fit test for the distributions and the associated problem of model selection. In Section 3, a parametric test based on the supreme distance between the two estimated drug concentration–time curves is then constructed for testing against the equivalence of the two curves. The results of a Monte Carlo study investigation of the level and power performances of the proposed test for a variety of PK models and generalized gamma distributions with different numbers of sampling time points and sample sizes are further presented and discussed in Section 4. In Section 5, the proposed test is implemented for the bioequivalence of two brands of benzbromarone tablets in the reduction of uric acid. Finally, in Section 6, conclusions and discussions are made concerning the application of the proposed model and bioequivalence test.

DOI: 10.1002/sim

2. STATISTICAL MODEL FOR DRUG CONCENTRATIONS

2.1. Statistical model

Let $Y_{ijk\ell}$ be the logarithm of the drug concentration of the jth subject in the ith sequence during period k at time t_ℓ, for $i,k=1,2$, $j=1,\ldots,n_i$, $\ell=1,\ldots,m$, that are observed under a 2×2 crossover design. For oral administered drugs, in particular, we consider the one-compartment model with first-order absorption and elimination:

$$\mu(k_a,k_e,V;t)=\frac{dk_a}{V(k_a-k_e)}(\mathrm{e}^{-k_e t}-\mathrm{e}^{-k_a t}) \tag{1}$$

where k_a and k_e are the absorption and elimination rate, V is the volume parameter, d is the dose level applied and t is the time point. Note that most of the variability in subjects is due to differences in the volume parameter, which can be regressed by some recorded covariates of the subjects [9]. Let $w_{j1},\ldots,w_{jr}$ be the covariates of subject j, $j=1,\ldots,n_i$, $i=1,2$. Then, we propose a statistical model for the logarithm of the drug concentration as given by

$$Y_{ijk\ell}=\log\mu(k_{ah},k_{eh},V_{jh};t_\ell)+\pi_k+\log\varepsilon_{ijk\ell},\quad i,k=1,2,\quad j=1,\ldots,n_i,\quad \ell=1,\ldots,m \tag{2}$$

where π_k is the effect of kth period with usual condition $\pi_1+\pi_2=0$, the $\varepsilon_{ijk\ell}$ are the associated error variables, and $\log(V_{jh})=\log(V_h)+\sum_{q=1}^{r}\alpha_q w_{jq}$ with $h=R$, if $i=k$, $h=T$, otherwise.

To make model (2) of wide application, we consider the error variables that are distributed according to the generalized gamma distribution [7], including a variety of right-skewed distributions such as gamma, Weibull and lognormal distributions. Note that the probability density function of the generalized gamma distribution [7] is

$$f(\varepsilon;\beta,\sigma,\lambda)=\begin{cases}\dfrac{|\lambda|[\lambda^{-2}(\mathrm{e}^{-\beta}\varepsilon)^{\lambda/\sigma}]^{\lambda^{-2}}\exp[-\lambda^{-2}(\mathrm{e}^{-\beta}\varepsilon)^{\lambda/\sigma}]}{\sigma\varepsilon\Gamma(\lambda^{-2})}, & \lambda\neq 0\\ \dfrac{\exp[-(\log\varepsilon-\beta)^2/(2\sigma^2)]}{\sqrt{2\pi}\sigma\varepsilon}, & \lambda=0\end{cases} \tag{3}$$

where β,σ and λ are the location, scale and shape parameters, respectively. The associated cumulative distribution function is then obtained as

$$F(\varepsilon;\beta,\sigma,\lambda)=\begin{cases}\Gamma[\lambda^{-2}(\mathrm{e}^{-\beta}\varepsilon)^{\lambda/\sigma};\lambda^{-2}], & \lambda>0\\ 1-\Gamma[\lambda^{-2}(\mathrm{e}^{-\beta}\varepsilon)^{\lambda/\sigma};\lambda^{-2}], & \lambda<0\\ \Phi[(\log\varepsilon-\beta)/\sigma], & \lambda=0\end{cases} \tag{4}$$

where $\Gamma(s;\gamma)=\int_0^s u^{\gamma-1}\mathrm{e}^{-u}\mathrm{d}u/\Gamma(\gamma)$ and $\Phi(\cdot)$ is the distribution function of a standard normal random variable. We denote, hereafter, $GG(\beta,\sigma,\lambda)$ for such a distribution. As usual, the error variable is assumed to have mean zero and, hence, we consider the situation with $E(\varepsilon_{ijk\ell})=1$. Therefore, β is a function of σ and λ as given by

$$g(\sigma,\lambda)=\log\Gamma(\lambda^{-2})-2\sigma(\log\lambda)/\lambda-\log\Gamma(\lambda^{-2}+\sigma/\lambda) \tag{5}$$

DOI: 10.1002/sim

2.2. Estimation and goodness-of-fit test for the models

Because of the detection ability of the equipment or machine, it occurs often that the drug concentration in blood or plasma below some level may not be detected or measured and, hence, the associated concentration data are subject to left-censorship. Let y_c be the detection limit and $\delta_{ijk\ell} = I(y_{ijk\ell} \geqslant y_c)$ the censoring index, where $I(A) = 1$ if A is true, $= 0$, otherwise. Since the left-censored data can have considerable impact on the estimation of the drug concentration–time curves [9], the likelihood function of the parameters given the data set, $\{y_{ijk\ell}, i, k = 1, 2, j = 1, \ldots, n_i, \ell = 1, \ldots, m\}$, is obtained as

$$L = L_{11} \times L_{12} \times L_{21} \times L_{22}$$

where L_{ik}, the likelihood function in sequence i during period k, for $i, k = 1, 2$, are given by

$$L_{ik} = \prod_{j=1}^{n_1} \prod_{\ell=1}^{m} \{f(\exp(y_{ijk\ell} - \log \mu(k_{aR}, k_{eR}, V_{Rj}; t_\ell) + (-1)^k \pi_1); g(\sigma, \lambda), \sigma, \lambda)\}^{\delta_{ijk\ell}}$$

$$\times \{F(\exp(y_c - \log \mu(k_{aR}, k_{eR}, V_{Rj}; t_\ell) + (-1)^k \pi_1); g(\sigma, \lambda), \sigma, \lambda)\}^{1-\delta_{ijk\ell}}$$

with $f(\varepsilon; \beta, \sigma, \lambda)$ and $F(\varepsilon; \beta, \sigma, \lambda)$ stated in (3) and (4), respectively. Note that the maximum likelihood estimates (MLEs) of the corresponding parameters can be obtained by using the general optimizing routine *optim* in R software and the variance–covariance matrix of the MLEs can be found by taking the inverse of the observed Fisher information matrix. Moreover, the MLEs in model (2) are, in fact, obtained based on all the $2m(n_1 + n_2)$, in total, observations. If this number is large enough, the approximated normal distributions for the MLEs would be reasonable [11].

To investigate the goodness-of-fit of a specific distribution for the drug concentrations, we consider the likelihood ratio test for the null hypothesis that the error variables in model (2) are distributed according to such a distribution against an alternative hypothesis that the error variables are distributed according to the generalized gamma distribution. Denote the associated maximum likelihood functions as L_0 and L_{GG}, respectively. We then claim, at significance level α, that the assumption of the specific distribution is invalid if $-2\{\log L_0 - \log L_{GG}\} \geqslant \chi^2_{1,\alpha}$, where $\chi^2_{df,\alpha}$ is the upper αth percentile of a chi-squared distribution with degrees of freedom df. To further investigate the feasibility of the assumption of the generalized gamma distribution for the drug concentrations, we then suggest that one draws the quantile–quantile plot for the residuals under model (2). If the plot shows a tendency that is relatively close to the straight line with slope one through the origin, then the generalized gamma distribution is employed for modeling the drug concentrations.

To select covariates for use in model (2), we suggest that one fits the data with all possible models since the number of covariates involved in the PK study is usually not so large. Let M_j be the fitted model that involves v_j unknown parameters and the associated maximum likelihood functions are denoted by $L(M_j)$, $j = 1, 2, \ldots, m$. We then compute, for each model involved, the Akaike Information Criterion (AIC) [12], that is, $-2 \log L(M_j) + 2v_j$, $j = 1, 2, \ldots, m$. Finally, the model with the smallest AIC is recommended for fitting the data obtained from a 2×2 crossover study.

DOI: 10.1002/sim

3. MULTIPLE TEST

To conduct a bioequivalent test for the test and reference drugs under model (2), we suggest directly compare the two mean drug concentration–time curves. To this end, we consider testing the null hypothesis of departure

$$H_0: \left\{ \begin{array}{l} \mu(k_{aT}, k_{eT}, V_T; t)/\mu(k_{aR}, k_{eR}, V_R; t) \leqslant 0.80 \\ \text{or } \mu(k_{aT}, k_{eT}, V_T; t)/\mu(k_{aR}, k_{eR}, V_R; t) \geqslant 1.25 \text{ for some } t \in [t_1, t_m] \end{array} \right\}$$

against the alternative hypothesis of equivalence

$$H_1: \{0.80 < \mu(k_{aT}, k_{eT}, V_{Tj}; t)/\mu(k_{aR}, k_{eR}, V_{Rj}; t) < 1.25 \text{ for all } t \in [t_1, t_m]\}$$

where the mean drug concentration $\mu(\cdot)$ is given in (1) and the margins of relative efficiency as 0.80 and 1.25 are currently employed in FDA [2]. To perform a multiple test for the above hypotheses, we may construct the confidence set for the mean ratios over the time period under study. However, the bioequivalence holds if the largest discrepancy between the two concentration curves is equivalence. Let

$$D(t) = \log \mu(k_{aT}, k_{eT}, V_T; t) - \log \mu(k_{aR}, k_{eR}, V_R; t) \quad \text{for } t \in [t_1, t_m]$$

Therefore, we suggest to find the $100(1-\alpha)$ per cent upper confidence bound for $\sup_{t\in[t_1,t_m]} |D(t)|$, denoted by DU_α. We then reject H_0 and claim the bioequivalence at significance level α if $DU_\alpha < 0.223 (= \log 1.25)$. We term 0.223 hereafter to be the bioequivalence margin.

To find DU_α, the $100(1-\alpha)$ upper confidence bound for $\sup_{t\in[t_1,t_m]} |D(t)|$, let

$$\hat{D}(t) = \log \mu(\hat{k}_{aT}, \hat{k}_{eT}, \hat{V}_T; t) - \log \mu(\hat{k}_{aR}, \hat{k}_{eR}, \hat{V}_R; t) \quad \text{for } t \in [t_1, t_m]$$

where $\hat{k}_{ah}, \hat{k}_{eh}$ and $\hat{V}_h$ are the MLEs of k_{ah}, k_{eh} and V_h, respectively, for $h = T$ and R. Note that, given the MLEs, $\sup_{t\in[t_1,t_m]} |\hat{D}(t)|$ is a function of t. Hence, we can numerically obtain the value of t, say, t_S, such that $|\hat{D}(t_S)| = \sup_{t\in[t_1,t_m]} |\hat{D}(t)|$. We then suggest to use the parametric bootstrap procedure [13] to find the upper αth percentile of the sampling distribution of $\sup_{t\in[t_1,t_m]} |D(t)| - |\hat{D}(t_S)|$. The algorithm of the parametric bootstrap procedure is stated in the following:

1. Generate B random sets $\{\varepsilon^b_{ijk\ell}, i, k = 1, 2, j = 1, \ldots, n_i, \ell = 1, \ldots, m\}$ from $GG(g(\hat{\sigma}, \hat{\lambda}), \hat{\sigma}, \hat{\lambda})$ and let $y^b_{ijk\ell} = \mu(\hat{k}_{ah}, \hat{k}_{eh}, \hat{V}_h; t_\ell) \times \varepsilon^b_{ijk\ell}$ for $h = T$ and R, $b = 1, \ldots, B$.
2. Compute the MLEs $\hat{k}^b_{aR}, \hat{k}^b_{eR}, \hat{V}^b_R, \hat{k}^b_{aT}, \hat{k}^b_{eT}, \hat{V}^b_T$ under model (2) based on the bootstrapped data $\{y^b_{ijk\ell}, i, k = 1, 2, j = 1, \ldots, n_i, \ell = 1, \ldots, m\}$ and find $\xi_b = |\hat{D}(t_S)| - |\hat{D}^b(t^b_S)|, b = 1, \ldots, B$.
3. Let ξ_α be $B(1-\alpha)$th value in the ordered list of $\xi_b, b = 1, \ldots, B$. Then DU_α is $|\hat{D}(t_S)| + \xi_\alpha$.

4. A SIMULATION STUDY

4.1. Design of the simulation study

We conducted a Monte Carlo study to investigate the level and power performances of the proposed test, denoted by MT, relative to the nearly unbiased test for the equivalence of AUC and the

DOI: 10.1002/sim

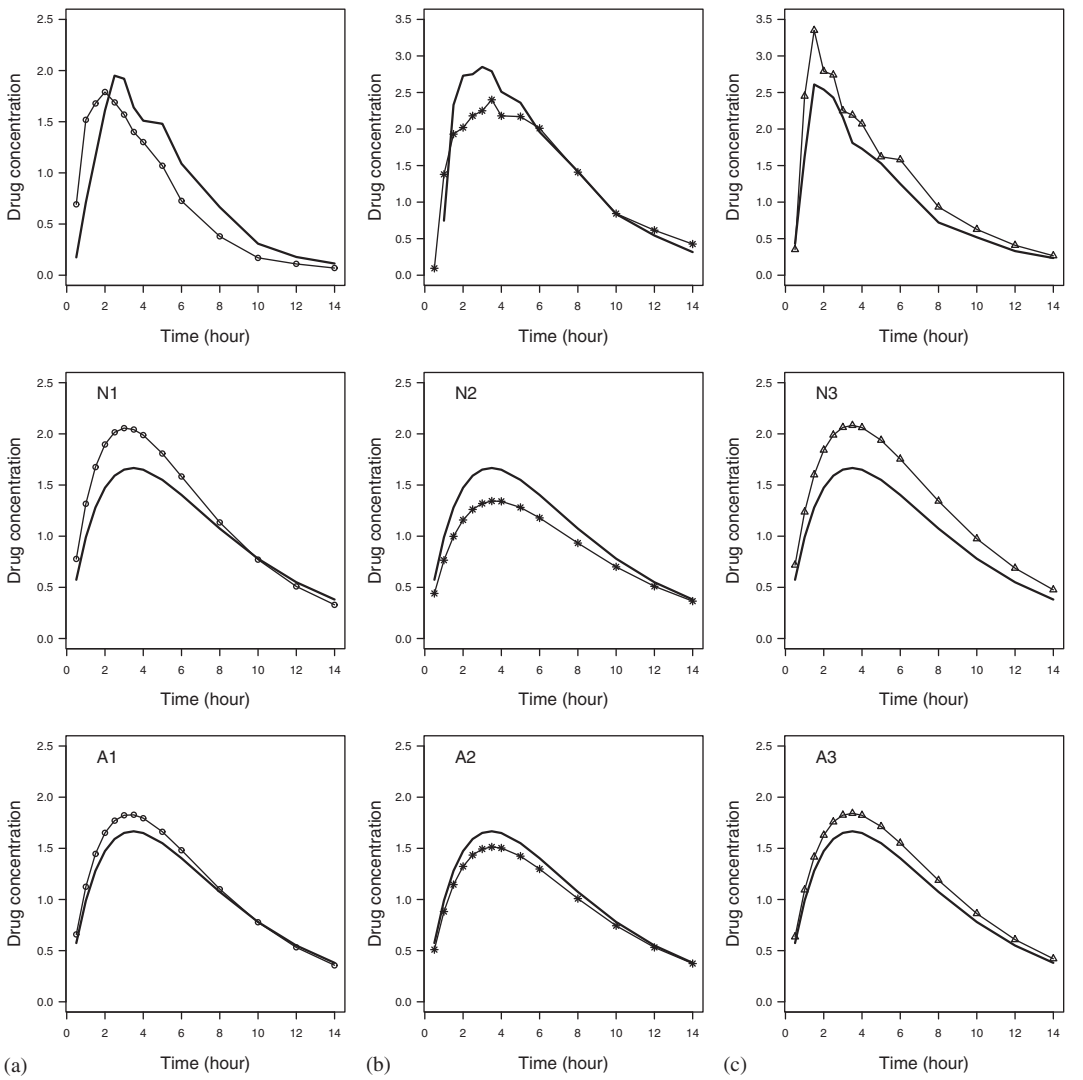

Figure 1. The three patterns of drug concentration–time curves recorded (top), under null (middle) and alternative (bottom) hypotheses.

bivariate test for the equivalence of both the AUC and $C_{\max}$ suggested in Berger and Hsu [3], denoted by UT and BT, respectively. Note that the BT is an intersection–union test incorporating with two nearly unbiased tests for the equivalence of AUC and $C_{\max}$, respectively. In the simulation study, we consider the situation where the two drugs are administered to $n=16$ or 24 volunteers in 2×2 crossover designs and the drug concentrations are measured at (i) 0.5, 2, 4, 6, 10, 14 h ($m=6$) or (ii) 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4, 5, 6, 8, 10, 12, 14 h ($m=14$) after the volunteers started to take the drug. Note that the former case is especially considered for the elderly persons whose drug concentrations are usually sampled in a relative few time points.

DOI: 10.1002/sim

We choose in the simulation study three pairs of the drug concentration–time curves for the reference and test drugs, which are found in the real data presented in the next section (Figure 1). All the concentration–time curves under the null or alternative hypothesis are designed to have the patterns that are similar to the three stated above. To study the level performance of the proposed test, we consider the settings with nonbioequivalence of the two drugs where the curves have $(k_{aR}, k_{eR}, V_R) = (0.4, 0.2, 15)$ but different values of (k_{aT}, k_{eT}, V_T):

N1. $(k_{aT}, k_{eT}, V_T) = (0.4, 0.2, 12)$;
N2. $(k_{aT}, k_{eT}, V_T) = (0.44, 0.24, 11.52)$;
N3. $(k_{aT}, k_{eT}, V_T) = (0.38, 0.18, 19.13)$.

Note that the cases N1–N3 are the null hypotheses under consideration, which represent nonbioequivalence. Therefore, the level of the test is the probability of claiming bioequivalence when the two drugs are, in fact, not bioequivalent. Moreover, to study the power performance of the proposed test, we consider the settings where the two drugs are actually bioequivalent with $(k_{aR}, k_{eR}, V_R) = (0.4, 0.2, 15)$, again, but with a variety of (k_{aT}, k_{eT}, V_T):

A1. $(k_{aT}, k_{eT}, V_T) = (0.4, 0.2, 13.57)$;
A2. $(k_{aT}, k_{eT}, V_T) = (0.42, 0.22, 13.33)$;
A3. $(k_{aT}, k_{eT}, V_T) = (0.39, 0.19, 16.73)$.

Therefore, in this simulation study, the power of the test is the probability of claiming bioequivalence under cases A1–A3 where the two drug concentration–time curves are equivalent. In particular, cases N1 and A1 indicate the situation that the curve of the test drug is higher and lower than that of the reference drug in earlier and later time periods, respectively. Both the cases N2 and A2 show that the two curves are depart in earlier time period but are close to each other in later time period. On the other hand, cases N3 and A3 show that the curve of the test drug is higher than the curve of the reference drug for whole time period. The mean drug concentration–time curves under N1–N3 and A1–A3 are shown in Figure 1 and the associated logarithm of the ratio of the mean drug concentrations are further given in Figure 2.

Note that the mean of the error variable $\varepsilon_{ijk\ell}$ is 1. Since the sample variance of the data under study in next section is 0.2, we consider a variety of generalized gamma distributions with mean 1 and variance 0.2. Therefore, we take $\beta = -\log 1.2/2$ and $\sigma^2 = \log 1.2$ for lognormal distribution. The gamma distribution under study has shape parameter 5 and scale parameter 0.2 and the Weibull distribution has shape parameter 2.4 and scale parameter 1.1, and the generalized gamma distribution has $\sigma^2 = 0.09$ and $\lambda = 2$. For simplicity, we assume that there is no period effect and subject variation.

For each of the settings under study, 2000 replicates were used to obtain the estimated level or power under the nominal level $\alpha = 0.10$. The standard deviation of the level estimate can then be approximated by 0.007 ($\approx \sqrt{0.1 \times 0.9/2000}$). Note that we obtain the necessary upper confidence bound in the proposed test based on the number of bootstrap samples $B = 2000$. The estimated levels and powers for the settings under study are finally given in Tables I and II, respectively.

4.2. Results of the simulation study

From the results in Table I, we find that the proposed test MT is generally superior to either the UT or BT for maintaining its level when the global bioequivalence is violated. The type I error rate of the test UT, testing for the bioequivalence in terms of AUC only, is far beyond its level under

DOI: 10.1002/sim

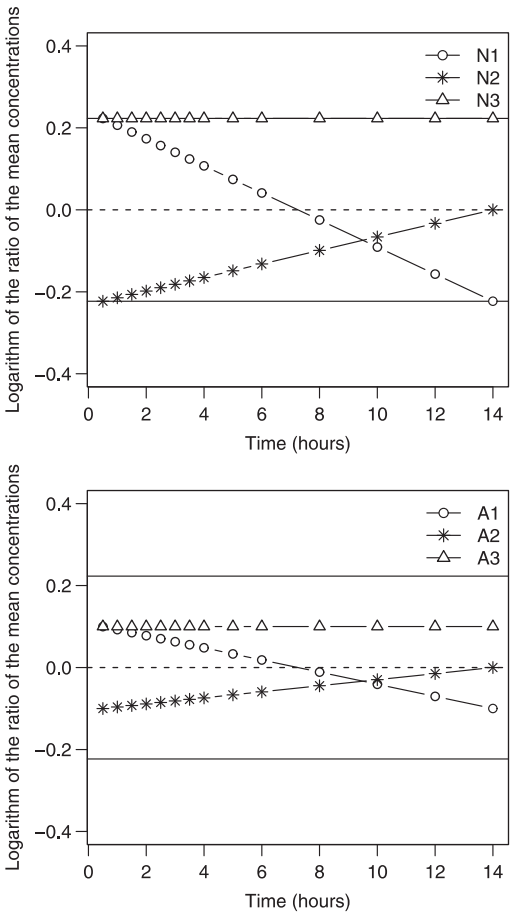

Figure 2. Logarithm of the ratio of mean drug concentrations for level (top) and power (bottom) study.

the three null hypotheses considered in the simulation study. The test BT for the equivalence of AUC and $C_{\max}$, simultaneously, is not able to maintain its level, especially, for the case when the two drug concentration–time curves are crossing. This result is not surprising since both the AUC and $C_{\max}$ cannot successfully reflect the difference between the two crossing curves. In this case, for the similar reason, the level of MT with six sampling time points ($m=6$) tends to be larger than the specified nominal level, especially when the sample size is small as 16 under lognormal distribution.

The results in Table II indicate that all the three tests are more powerful for sampling time points $m=14$ than $m=6$. This is because that the AUC, $C_{\max}$ and model (2) can be more accurately estimated. Note that the UT test is not able to hold its level. Therefore, we simply compare the power performance between the BT and MT tests. For $n=16$ volunteers involved, ST is more powerful than BT under the three alternative hypotheses considered in the simulation study. When the number of volunteers increases up to 24, both the MT and BT are competitive for lognormal or

DOI: 10.1002/sim

Table I. Estimated level for drug concentrations measured from n individuals at m time points under $\alpha = 0.10$.

			(i) $m=6$			(ii) $m=14$		
Distribution	n	Null hypothesis	UT	BT	MT	UT	BT	MT
Lognormal	16	N1	0.495	0.150	0.138	0.528	0.085	0.082
		N2	0.248	0.079	0.081	0.270	0.048	0.024
		N3	0.113	0.049	0.058	0.075	0.024	0.044
	24	N1	0.603	0.152	0.106	0.795	0.132	0.091
		N2	0.313	0.098	0.093	0.465	0.104	0.042
		N3	0.091	0.050	0.040	0.092	0.036	0.045
Gamma	16	N1	0.463	0.149	0.115	0.760	0.125	0.091
		N2	0.274	0.094	0.102	0.387	0.079	0.035
		N3	0.090	0.029	0.029	0.095	0.022	0.041
	24	N1	0.581	0.169	0.107	0.872	0.147	0.073
		N2	0.314	0.113	0.090	0.458	0.088	0.024
		N3	0.112	0.054	0.024	0.097	0.017	0.038
Weibull	16	N1	0.467	0.120	0.109	0.727	0.123	0.089
		N2	0.264	0.088	0.099	0.361	0.069	0.107
		N3	0.092	0.039	0.040	0.104	0.026	0.082
	24	N1	0.577	0.167	0.106	0.834	0.146	0.070
		N2	0.312	0.112	0.103	0.450	0.091	0.099
		N3	0.107	0.042	0.050	0.086	0.026	0.072
Generalized gamma	16	N1	0.431	0.129	0.069	0.700	0.135	0.067
		N2	0.249	0.091	0.080	0.390	0.072	0.053
		N3	0.095	0.047	0.065	0.107	0.031	0.059
	24	N1	0.570	0.168	0.031	0.854	0.162	0.057
		N2	0.330	0.105	0.071	0.495	0.108	0.047
		N3	0.103	0.042	0.049	0.083	0.017	0.056

gamma distribution. However, for the Weibull and generalized gamma distributions under study, the power of MT is higher than that of the BT.

5. DATA ANALYSIS

We illustrate the use of the proposed test for the bioequivalence between two brands of benzbromarone tablets, where the test drug Euricon and the reference drug Urinorm are manufactured by two different pharmaceutical companies, respectively. Note that the benzbromarone, a well-known uricosuric agent, reduces serum uric acid concentrations probably by blocking tubular reabsorption. In this 2×2 crossover study [14], 16 healthy adult volunteers were randomly allocated to two treatment sequences. In sequence 1, eight volunteers were orally administered with one tablet of 50 mg of Urinorm and then, after one week, one tablet of 50 mg of Euricon. On the other hand, the other eight volunteers in sequence 2 receive the two drugs in reverse order in two periods. The blood samples were taken and the benzbromarone concentration was measured 14 times at 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4, 5, 6, 8, 10, 12, 14 h after the drug been administered. Some covariates associated with the volunteers are also recorded, including age (20–34 years), body weight

DOI: 10.1002/sim

Table II. Estimated power for drug concentrations measured from n individuals at m time points under $\alpha = 0.10$.

			(i) $m=6$			(ii) $m=14$		
Distribution	n	Alternative hypothesis	UT	BT	MT	UT	BT	MT
Lognormal	16	A1	0.807	0.392	0.409	0.987	0.523	0.605
		A2	0.727	0.347	0.424	0.900	0.460	0.515
		A3	0.633	0.321	0.337	0.894	0.452	0.477
	24	A1	0.931	0.566	0.533	0.928	0.634	0.662
		A2	0.873	0.518	0.470	0.929	0.606	0.585
		A3	0.779	0.487	0.424	0.906	0.573	0.538
Gamma	16	A1	0.834	0.415	0.495	0.980	0.577	0.663
		A2	0.734	0.405	0.480	0.947	0.548	0.587
		A3	0.632	0.366	0.382	0.882	0.531	0.563
	24	A1	0.915	0.587	0.532	1.000	0.721	0.749
		A2	0.839	0.559	0.562	0.989	0.686	0.612
		A3	0.789	0.511	0.418	0.962	0.699	0.675
Weibull	16	A1	0.786	0.456	0.698	0.972	0.720	0.796
		A2	0.735	0.447	0.679	0.937	0.689	0.828
		A3	0.598	0.368	0.596	0.894	0.617	0.802
	24	A1	0.903	0.622	0.785	0.997	0.860	0.873
		A2	0.834	0.565	0.761	0.990	0.855	0.890
		A3	0.765	0.537	0.684	0.953	0.778	0.862
Generalized	16	A1	0.765	0.502	0.726	0.974	0.861	0.851
gamma	A2	0.680	0.459	0.756	0.941	0.806	0.936	
		A3	0.573	0.402	0.673	0.858	0.739	0.899
	24	A1	0.885	0.699	0.756	0.997	0.948	0.901
		A2	0.833	0.652	0.802	0.990	0.949	0.957
		A3	0.727	0.576	0.781	0.954	0.899	0.960

(42–73 kg) and body height (155.5–173.5 cm). The recorded drug concentrations for volunteers receiving Urinorm and Euricon, respectively, are shown in Figure 3.

To investigate whether the benzbromarone concentrations are distributed according to some specific distribution, we fitted the data using model (2) with the three covariates and computed the likelihood ratio statistics as 512.8, 58.7 and 34.3 for lognormal, gamma and Weibull distributions. Note that, comparing the upper 5th percentile of a chi-square distribution with 1 degree of freedom, the three distributions are not applicable to the current data set. Therefore, we choose to fit the data under a generalized gamma distribution.

To select covariates for use in model (2), we fitted the data with all possible models and found the reduced model that gives the smallest value of AIC ($=675.9$) contains only age and body weight. The associated quantile–quantile plot, in Figure 4, of the residuals for fitting the selected model shows a tendency that is relatively close to the straight line through the origin with slope 1. This, again, confirms that the selected model together with a generalized gamma distribution is reasonable for fitting into the data under study.

Based on $B=2000$ number of bootstrapped samples, we estimated the 90 per cent upper confidence bound for $\sup_{t\in[t_1,t_m]}|D(t)|$ to be 0.061 ($|\hat{D}(t_S)|=0.046$ and $\xi_{0.1}=0.015$), which is smaller than 0.223, the bioequivalence margin as stated in Section 3. Therefore, the proposed test claims, at significant level 0.10, the drugs Euricon and Urinorm are bioequivalent in the reduction

DOI: 10.1002/sim

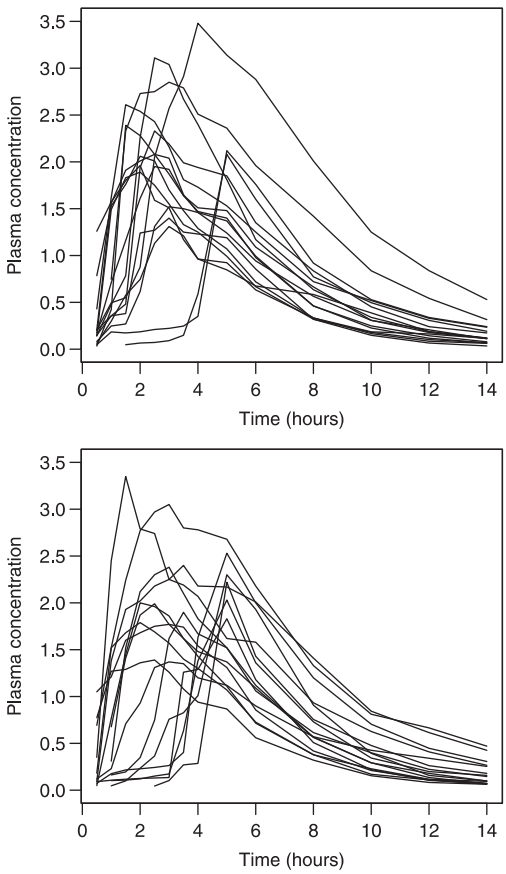

Figure 3. Drug concentrations observed for Urinorm (top) and Euricon (bottom) drugs.

of uric acid. Note that the fitted concentration–time curves are rather close to each other for the volunteer at any age with any body weight. Therefore, the univariate test based on AUC only and the bivariate test based on both AUC and $C_{\max}$ also lead to the conclusion of the bioequivalence of the two drugs.

For old people with possible sampling time points at, for instance, 0.5, 2, 4, 6, 10 and 14 h ($m=6$) after the drug been administered, the related 90 per cent upper confidence bound for the supreme distance is 0.089 ($|\hat{D}(t_S)|=0.072$ and $\xi_{0.1}=0.017$). Therefore, the test also concludes the bioequivalence between the two drugs. Note that both the univariate and bivariate tests reject, at significance level 0.10. Hence, all the three tests, again, reach the same conclusion of bioequivalence.

Since lognormal distribution is usually used in practice, we also analyze the data set under such a distribution when sampling time points are either 6 or 14. For $m=6$, we observe $|\hat{D}(t_S)|=0.267$ with $\xi_{0.1}=0.141$, and the 90 per cent upper confidence bound for $\sup_{t\in[t_1,t_m]}|D(t)|$ is 0.408. In addition, for $m=14$, we have $|\hat{D}(t_S)|=0.309$ and $\xi_{0.1}=0.112$, and hence the 90 per cent upper confidence bound for $\sup_{t\in[t_1,t_m]}|D(t)|$ is 0.421. Comparing the bioequivalence margin of 0.223,

DOI: 10.1002/sim

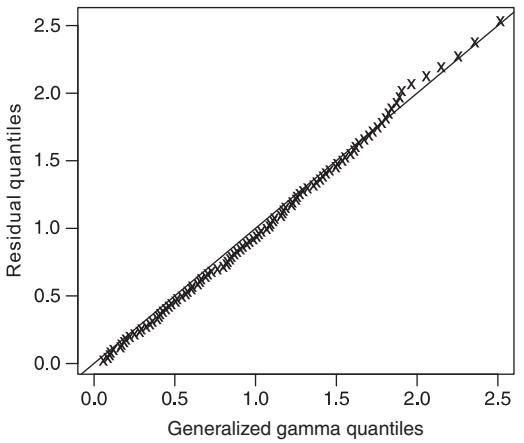

Figure 4. Quantile–quantile plot of the residuals of model (2) under the generalized gamma distribution.

the test does not reject the null hypothesis for $m=6$ or 14. Therefore, applying model (2) with lognormal error variables to the data set, the proposed test does not support, at significant level 0.10, the bioequivalence between Euricon and Urinorm in reducing uric acid.

6. DISCUSSION AND CONCLUSION

In this paper, we propose a statistical model for the drug concentrations in a PK study under 2×2 crossover designs. This statistical model includes a one-compartment PK model for the mean concentrations over time that allows for covariate-related between-subject variation, a fixed period effect and an error variable that is distributed according to a generalized gamma distribution. Note that the one-compartment PK model is usually used for orally taken drug. If the drug is administered subcutaneously or by intravenous bolus or infusion, other feasible PK models can be used. Moreover, the generalized gamma distribution is an extensive family that contains the most commonly used distributions. Therefore, the proposed statistical model would be of wide application in most of the practical PK studies.

Both the univariate and bivariate tests for the equality of the bioavailability parameters fail to maintain their levels and, hence, tend to erroneously conclude the bioequivalence between the test and reference drugs when the two drug concentration–time curves are, in fact, not equivalent. On the other hand, the multiple test proposed in this paper provides with a global investigation in the bioequivalence study that not only reasonably maintains its level but also has a better power performance than the bivariate test. Moreover, note that the proposed multiple test can be adapted to a variety of models and distributions involved in the PK study. For example, when any suitable compartment model and any reasonable distribution for the error variable are considered, the proposed test is still applicable only if we remedy the parametric bootstrap procedure stated in Section 3 accordingly.

Our simulation study shows that the number of the sampling time points would majorly affect the power performance of the proposed test, although the test tends to be more conservative on holding its level with more sampling time points. Therefore, by adding more sampling time points

DOI: 10.1002/sim

so that the appropriate model under study can be more accurately estimated would be a reasonable way to have a better chance to conclude the bioequivalence of a generic drug to the reference drug.

ACKNOWLEDGEMENTS

The authors thank the associate editor and two referees for many constructive suggestions which greatly improved the presentation of this paper. This work was supported by the National Science Council of Taiwan under grant number NSC 97-2118-M-008-003-MY2.

REFERENCES

1. Chow SC, Liu JP. *Design and Analysis of Bioavailability and Bioequivalence Studies*. Marcel Dekker: New York, 2000.
2. FDA Guidance. *Statistical Approaches to Establishing Bioequivalence*. U.S. Food and Drug Administration: U.S.A., 2001.
3. Berger RL, Hsu JC. Bioequivalence trials, intersection-union tests and equivalence confidence sets. *Statistical Science* 1996; **11**:283–319.
4. Wang W, Hwang JTG, Dasgupta A. Statistical tests for multivariate bioequivalence. *Biometrika* 1999; **86**:395–402.
5. Westlake WJ. Bioavailability and bioequivalence of pharmaceutical formulations. In *Biopharmaceutical Statistics for Drug Development*, Peace KE (ed.). Marcel Dekker: New York, 1988.
6. Liao JZ. Comparing the concentration curves directly in a pharmacokinetics, bioavailability/bioequivalence study. *Statistics in Medicine* 2005; **24**:883–891.
7. Stacy EW. A generalization of the gamma distribution. *Annals of Mathematical Statistics* 1962; **33**:1187–1192.
8. Wagner JG. *Pharmacokinetics for the Pharmaceutical Scientist*. Technomic: Pennsylvania, 1993.
9. Lindsey JK, Byrom WD, Wang J, Jarvis P, Jones B. Generalized nonlinear models for pharmacokinetic data. *Biometrics* 2000; **56**:30–36.
10. Salway R, Wakefield J. Gamma generalized linear models for pharmacokinetic data. *Biometrics* 2007; **64**:620–626.
11. Prentice RL. A log gamma model and its maximum likelihood estimation. *Biometrika* 1974; **61**(3):539–544.
12. Akaike H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 1974; **19**:716–723.
13. Efron B, Tibshirani R. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science* 1986; **1**:54–75.
14. Bioequivalence test for EURICON versus URINORM. Study no. MB-87107. *Technical Report*, Mithra Bioindustry Company, Ltd. Taipei, Taiwan, 2001.

DOI: 10.1002/sim