academicJournals

Vol. 7(16), pp. 852-864, 29 April, 2013
DOI 10.5897/AJPP12.341
ISSN 1996-0816 © 2013 Academic Journals
http://www.academicjournals.org/AJPP

African Journal of Pharmacy and Pharmacology

Full Length Research Paper

The comparison of multilevel models, method of moment and restricted maximum likelihood in assessing population bioequivalence

Zhuo-Zhi Shen[1], Min Yang[2]*, Qiao-Lan Liu[3], Samir Mehta[4] and Xiao-Song Li[1]

[1]Department of Health Statistics, West China School of Public Health, Sichuan University, Chengdu, Sichuan, 610041, China.
[2]Division of Psychiatry, School for Community Health Sciences, University of Nottingham, 9 Triumph Road, Nottingham, NG7 2TU, UK.
[3]Department of Health Education, West China School of Public Health, Sichuan University, Chengdu, Sichuan, 610041, China.
[4]Clinical Trials Unit, University of Nottingham, Queen's Medical Centre, Nottingham, NG7 2UH, UK.

Accepted 28 June, 2012

The determination of bioequivalence is very important in the pharmaceutical industries because the regulatory agencies like the 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 brand-name product. Up to date, there is a lack of widely accepted statistical procedure for assessing population bioequivalence. We propose multilevel models (MLMs) for evaluating and estimating parameters to assess population bioequivalence (PBE), and compare statistical properties of PBE estimators between MLMs and current approaches recommended by the FDA-the method of moment (MOM) and REML. The approach developed is illustrated using a real data set from the FDA. Statistical properties of MLM estimators are further explored using simulation studies as compared with MOM and restricted maximum likelihood (REML) estimators. The performance of MLM appeared to be much comparable to the existing REML procedure. The results suggest that MLM estimators that are fully comparable with REML estimators can be an adequate approach for assessing PBE. The MLMs approach proposed in the study provides an alternative and yet more flexible and powerful method than existing methods in assessing bioequivalence (BE) for complex study designs and data structures.

Key words: Population bioequivalence, multilevel models, simulation, estimation procedure, restricted iterative generalized least square (RIGLS), restricted maximum likelihood (REML), method of moments (MOM), food and drug administration (FDA).

INTRODUCTION

The goal of a bioequivalence study is to show that two formulations of a drug have similar bioavailability (Ashby, 2006). The determination of bioequivalence is very important in the pharmaceutical industries because the regulatory agencies like the United States 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 brand-name product. Thus, the design, performance and evaluation of bioequivalence studies have received major attention among the health authorities and pharmaceutical industry, as well as, statisticians (Ashby, 2006; Blume and Midha, 1993).

During the last few decades, the assessment of average

*Corresponding author. E-mail: minyang64@yahoo.cn.

bioequivalence (ABE), which emphasizes comparing the means of two drug formulations has been well established (Ashby, 2006; Chen et al., 2000, 2001). The major concern was then of course how population bioequivalence (PBE) and individual bioequivalence (IBE) should be statistically assessed (Chow and Liu, 2008; Davit et al., 2008). In contrast to ABE, PBE requires equivalence on both the averages as well as variances, whereas IBE not only requires equivalence on similarity of averages and variances, but also on the homogeneity of subject-by-formulation interaction. In 2001, The FDA guidance recommended the use of the method of moments (MOM) for variance component estimation, as well as, the restricted maximum likelihood (REML) estimation of data including random missing records based on a two-sequence, four-period (2 × 4) cross-over design (Carrasco and Jover, 2003).

During the last decade, although, ABE is now a relatively matured field and is theoretically well understood, there is still a lack of widely accepted statistical procedure for assessing PBE. Many researchers have argued that the current FDA's statistical procedures are unsatisfactory in terms of their statistical properties (Endrenyi et al., 2003; FDA, 2001, 2003; Ghosh and Gönen, 2008; Ghosh and Ntzoufras, 2005). In addition, complete data sets are not common in bioequivalence studies, and MOM estimation is known to be limited in such situations. Furthermore, little information regarding the statistical test procedure is provided with the exception of 2 × 4 cross-over design, while for PBE and IBE, McNally et al. (2003) examined the generalized *P*-value approach for making inferences concerning the FDA-recommended PBE and IBE criteria. Carrasco et al. (2003) introduced the structural equation model (SEM) approach for parameter estimation and constructing criterion to assess IBE. Dragalin et al. (2003) proposed the Kullback–Leibler divergence (KLD) approach for parameter estimation and defining criterion in evaluating PBE and IBE. In 2005, Bayesian methods of assessing IBE and PBE using the FDA criteria were proposed (Ghosh and Ntzoufras, 2005). However, none of these methods are robust and sensitive enough to be accepted as the standard approach by current FDA guidance. Moreover, since bioequivalence evaluations are often based on the logarithmic transformation of AUC and C_{max}, the FDA consider two drugs to be bioequivalent only if they are similar in both AUC and C_{max}. It would be better to consider a test which included several endpoints together. However, none of the proposed methods including MOM, REML, SEM or KLD gave examples in assessing ABE, PBE and IBE with multiple endpoints. Research in this area is almost blank.

The theory by Harvey (2010) suggested that bioequivalence is a natural field to be assessed using MLMs. In principle MLMs can easily disentangle multiple variance components as well as covariance required by PBE assessment. To our knowledge, there was no thoroughly study discussing its applications in PBE studies.

This paper serves two purposes: (1) to introduce multilevel models using restricted iterative generalized least square (RIGLS) algorithm in estimating parameters to assess PBE, and (2) to compare the statistical properties of PBE estimators between multilevel models and two current approaches recommended by FDA. Our concern is not on the inference of the PBE criterion but merely estimating the components in the current FDA-proposed PBE criterion.

MATERIALS AND METHODS

Current methods for PBE

The 2001 FDA guidance proposed the following aggregated, scaled moment-based one-sided null hypotheses as:

$$H_0 : \frac{(\mu_T - \mu_R)^2 + (\sigma_T^2 - \sigma_R^2)}{\max(\sigma_{T0}^2, \sigma_R^2)} \geq \theta_p$$

Note that μ_T and μ_R, σ_T^2 and σ_R^2 are population means and variances in the logarithm scale. For convenience, we define $\Delta = \mu_T - \mu_R$ as mean difference, $\sigma_T^2 = \sigma_{BT}^2 + \sigma_{WT}^2$ as total variance of T, $\sigma_R^2 = \sigma_{BR}^2 + \sigma_{WR}^2$ as total variance of the R. σ_{BR}^2 and σ_{BT}^2 are between-subject variances, σ_{WR}^2 and σ_{WT}^2 are within-subject variances respectively for the T and R. The expression is linearised as:

$$H_0 : \nu_{PBE} = \Delta^2 + \sigma_T^2 - (1 + \theta_p)\sigma_R^2 \geq 0$$

When $\hat{\sigma}_R^2 > 0.04$ or $H_0 : \nu_{C.PBE} = \Delta^2 + \sigma_T^2 - \sigma_R^2 - \theta_p \cdot \sigma_{T0}^2 \geq 0$, $\hat{\sigma}_R^2 \leq 0.04$. θ_p is the bioequivalent limit with a recommended value of 1.7448. If the upper bound of the 95% one-sided confidence interval of criterion for both ln (AUC) and ln (C_{max}) are below zero, PBE can be claimed. The measure of PBE is a mixture of the mean and variance of the ln (AUC) and the ln (C_{max}). In the guidance, FDA recommended MOM to estimate variance component for complete data, as well as, REML estimation for data including missing records. After the estimation of the mean difference and the variances has been completed, 95% one-sided upper confidence bound for a linearised form of the PBE criterion can be obtained. FDA indicated that the method for the upper confidence bound should be consistent with the method used for estimating the variances.

Multilevel models (MLMs)

Models and notations

When measurements are repeated on the same subjects, as the data from 2 × 4 replicated cross-over design, a 2-level hierarchical structure is established. The measurement replicates are regarded as level 1 units, and subjects as level 2 units. Therefore, multilevel models (MLMs), proposed by Harvey Goldstein could be a powerful tool to analyze BE data. For the j^{th} subject at i^{th} replicate, we can fit

Table 1. Variance components for PBE criteria in multilevel model.

Parameter	Drug R	Drug T
Between-subject (Level 2)	$\sigma^2_{u0}\ (\sigma^2_{BR})$	$\sigma^2_{u0}+2\sigma_{u01}+\sigma^2_{u1}\ (\sigma^2_{BT})$
Within-subject (Level 1)	$\sigma^2_{e4}\ (\sigma^2_{WR})$	$\sigma^2_{e5}\ (\sigma^2_{WT})$
Total	σ^2_R	σ^2_T

a two-level random slope model in which factors for sequence and period are included, and the variances for T and R groups are defined at level 1 as:

$$y_{ij}=\beta_{0j}+\beta_{1j}x_1+\beta_2\text{sequence}_j+\beta_3\text{period}_{ij}+e_{4ij}z_{2ij}+e_{5ij}z_{3ij}$$

$$\beta_{0j}=\beta_0+u_{0j},\quad \beta_{1j}=\beta_1+u_{1j}$$

$$\begin{bmatrix}u_{0j}\\u_{1j}\end{bmatrix} \sim N(0,\ \Omega_u):\quad \Omega_u=\begin{bmatrix}\sigma^2_{u0} & \\ \sigma_{u01} & \sigma^2_{u1}\end{bmatrix}$$

$$\begin{bmatrix}e_{4ij}\\e_{5ij}\end{bmatrix} \sim N(0,\ \Omega_e):\quad \Omega_e=\begin{bmatrix}\sigma^2_{e4} & \\ \sigma_{e45} & \sigma^2_{e5}\end{bmatrix}$$

X_1 = 1 for drug T and 0 for drug R.
$Z4ij$ = 1 for drug R and 0 for drug T.
$Z5ij$ = 1 for drug T and 0 for drug R.

The intercept term is denoted by β_0 which estimates the mean of drug R, and the slope β_1 estimates the difference of means between T and R, (that is, $\beta_1=\mu_T-\mu_R=\Delta$). The '$\beta$s' are always referred to as "fixed" parameters of the model, such as period and sequence effects. What makes the two-level model different from standard linear regression model are the additional random effect, '*u*s' and '*e*s'. The '*u*s' parameters denote the residual in level 2, and the '*e*s' denote the residual in level 1.

Variance components for PBE criteria

To obtain variance components of PBE as shown in Equation 2, we can specify the level 2 variance as function of x_1 as represented:

$$\text{var}(u_{0j}+u_{1j}x_1)=\sigma^2_{u0}+2\sigma_{u01}x_1+\sigma^2_{u1}x_1^2.$$

Thus, the level 2 variance of between subjects for drug R (x_1 = 0) is σ^2_{u0}, and $\sigma^2_{u0}+2\sigma_{u01}+\sigma^2_{u1}$ for drug T (x_1 = 1). Since a subject cannot take both test and reference drug in each measurement, so $\text{cov}(e_{4ij},e_{5ij})=0$. Hence, the level 1 variance (or within-subject) for drug R (Z_{4ij} = 0, Z_{5ij} = 1) is:

$$\sigma^2_{WR}=\sigma^2_{e4}$$

for drug T (Z_{4ij} = 0, Z_{5ij} = 1) is:

$$\sigma^2_{WT}=\sigma^2_{e5},$$

The total level 1 variance is:

$$\text{var}(e_{4ij}z_{4ij}+e_{5ij}z_{5ij})=\sigma^2_{e4}z^2_{4ij}+\sigma^2_{e5}z^2_{5ij}.$$

The total variance of R is:

$$\sigma^2_R=\sigma^2_{BR}+\sigma^2_{WR}=\sigma^2_{u0}+\sigma^2_{e4}.$$

The total variance of T is:

$$\sigma^2_T=\sigma^2_{BT}+\sigma^2_{WT}=\sigma^2_{u0}+2\sigma_{u01}+\sigma^2_{u1}+\sigma^2_{e5}.$$

The variance components in the model above are summarized in Table 1.

Parameter estimation methods in MLMs

Goldstein and his colleagues developed the MLwiN software package (Version 2.23, Centre for Multilevel Modeling, University of Bristol, UK) for multilevel modeling (Rashbash et al., 2008). MLwiN uses the iterated generalized least squares (IGLS) and restricted iterated generalized least squares (RIGLS) algorithms to estimate parameters (Goldstein, 1986, 1989). In considering the small size of samples in both the illustrative example and our simulation data, we used RIGLS algorithm for all multilevel models presented in the paper. It is noted that we also use the term MLMs and RIGLS exchangeable for convenience throughout the paper.

Table 2. Parameter estimations (SE) from fitting multilevel models.

Parameter	ln (AUC)	ln (C_{max})
Fixed effect		
β_0 intercept	7.823(0.325)	5.894(0.370)
β_1 treatment	-0.043(0.060)	-0.109(0.074)
β_2 sequence	-0.186(0.204)	-0.302(0.231)
β_3 period	0.047(0.020)	0.056(0.027)
Random effect		
Level 2 (subject)		
σ^2_{u0} (σ^2_{BR})	0.366(0.093)	0.478(0.126)
σ^2_{u1}	0.053(0.034)	0.060 (0.053)
σ_{u01}	-0.009(0.039)	-0.035(0.058)
Level 1 (replicate)		
σ^2_{WR}	0.065(0.015)	0.119(0.028)
σ^2_{WT}	0.094(0.022)	0.165(0.038)

Table 3. Comparison of three methods in variance components estimates.

Variance component	ln (AUC)			ln (C_{max})		
	MOM	REML	MLM	MOM	REML	MLM
σ^2_{BT}	0.400(0.096)	0.400 (0.107)	0.401(0.105)	0.472 (0.113)	0.468(0.133)	0.468(0.129)
σ^2_{BR}	0.378(0.090)	0.366(0.095)	0.366(0.093)	0.515 (0.123)	0.478 (0.129)	0.478(0.126)
σ^2_{WT}	0.098 (0.024)	0.094(0.022)	0.094(0.022)	0.171 (0.041)	0.165(0.039)	0.165(0.038)
σ^2_{WR}	0.067 (0.016)	0.065 (0.015)	0.065(0.015)	0.126 (0.030)	0.119 (0.028)	0.119(0.028)
ν_{PBE}	-0.723	-0.686	-0.691	-1.106	-0.996	-1.019
95 percent upper bound	-0.360	-0.355	-0.431	-0.603	-0.524	-0.650

* The MLM's 95 percent one-sided upper bound was estimated using asymptotic normal procedure.

An illustrative example

To demonstrate the methods described in this paper, we acquired data set 17a for drug#17: Antihypertensive on the FDA website, http://www.fda.gov/drugs/scienceresearch/researchareas/biostatistics/ucm081434.htm. It is a 2 × 4 cross-overdesign (RTTR and TRRT) with 19 and 18 subjects per sequence. Thus, there are two replications for each drug. The two response variables are the natural logarithm of the reported value for AUC and C_{max}. We fitted two MLMs in MLwiN by using Equation 4 for ln(AUC) and ln(Cmax) respectively. The results obtained are presented in Table 2. Since total variance of the R group for both ln(AUC) and ln(Cmax) in this case were greater than 0.04, that is, $\hat{\sigma}^2_R = \hat{\sigma}^2_{BR} + \hat{\sigma}^2_{WR} > 0.04$, the reference–scaled (FDA, 2001) was used to test for the PBE criterion. We compared the scale calculated based on parameter estimates from the MOM, REML and RIGLS in MLM. As shown in Table 3, the estimators of variance components in REML and MLM are similar, while MOM provided slightly larger estimations .Using all the aforementioned methods, the upper bounds of the 95% one-sided confidence intervals for both ln(AUC) and ln(C_{max}) are less than 0, so PBE can be claimed.

Simulation study

Study design

The simulation study was designed to investigate the performance of RIGLS estimation in multilevel models in comparison to that of MOM and REML. We simulated log-transformed data for the two sequence (TRTR/RTRT), replicated design assuming a sample size

Table 4. Parameter settings for simulations (n=36, $\rho = 0.9$, 500 runs per scenario).

Scenario	Δ	$\sigma_{WT} = \sigma_{WR}$	$\sigma_{BT} = \sigma_{BR}$
1	0.05	0.3	0.3
2		0.3	0.6
3		0.5	0.5
4		0.5	1.0
5	0.223	0.3	0.3
6		0.3	0.6
7		0.5	0.5
8		0.5	1.0

of 36 subjects, 18 per sequence. This sample size corresponded to the illustrative example earlier mentioned, which also represented a moderate study size suggested in the FDA guidance.

Scenarios to be investigated

Scenarios were assumed based on drugs with typical characteristics and statistical distributions in practice and in previous studies. Six different parameters are to be defined to generate sample datasets for the simulation: Δ, σ_{WR}, σ_{WT}, σ_{BR}, σ_{BT} and ρ. Following other examples, we assumed same standard deviation for the two formulations, namely $\sigma_{WT} = \sigma_{WR}$ and $\sigma_{BT} = \sigma_{BR}$ (FDA, 2001; Endrenyi et al., 2000). We then considered the values for $\Delta = \mu_T - \mu_R$ =0.05 and ln1.25 (≈0.223), which corresponds to two prototypical situations in ABE studies, and were also used in the simulation research of PBE in the GlaxoSmithKline technical report by Patterson and Jones (2002).

The log-transformed parameters σ_{WR} and σ_{WT} can be calculated from the within-subject coefficients of variation (CV) on the original scale using the equation:

$$\sigma_{WR} = \sigma_{WT} = \sqrt{\ln(1 + CV^2)}$$

A CV of 30% is generally considered the threshold for a highly variable drug or drug product (HVD). In 2003 to 2005, the FDA's Office of Generic Drugs (OGD) reviewed 1010 acceptable bioequivalence studies of 524 different drug products. Among them, the highest CV was 55% (Davit et al., 2005). Since the topic of BE for highly variable drugs is one that has been intensely debated in many recent articles, conferences and meetings nationally and internationally (Haidar et al., 2008), our simulations were conducted for moderate high and very high variability products with CV of 30 and 55% respectively, which derives the values for σ_{WR} and σ_{WT} approximately 0.3 and 0.5.

In many studies, σ_{BR} and σ_{BT} are generally larger than σ_{WR} and σ_{WT} (Riviere and Papich, 2009; Willavize et al., 2006), we considered scenarios including cases where the σ_{BR} and σ_{BT} ranged from 1 to 2 times greater than the σ_{WR} and σ_{WT}.This range was also used by the sample size determination tables for PBE given in the Appendix of the FDA guidance document, and it also appeared to correspond to some retrospective analyses of AUC in replicate cross-over bioequivalence studies.

Finally, all scenarios considered a strong correlation ρ between R and T at 0.9. The reason for examining cases with such seemingly high correlation is that responses within a subject to different formulations of the same drug are still likely to have a strong degree of correlation even if the formulations are only borderline bioequivalent (Shao et al., 2000).

In summary, given the consideration of six parameters, eight scenarios were generated (Table 4). The chosen parameter values were considered to be representative of typical bioequivalence data and these are quite natural and common in practice.

Data generation and data analysis

For each scenario, 500 samples were simulated using SAS (Version 9.1.3, Cary, NC). We used SAS to perform MOM and REML estimation procedures and fitted MLMs using RIGLS procedure in MLwiN. Performances of different estimation procedures were evaluated using the following quantities that are frequently regarded as benchmark accuracy and precision measures; bias, mean square error (MSE) and coverage of 95% confidence interval.

SIMULATION RESULTS

The summarized results of parameter estimates, bias, MSE and CI coverage from 500 run samples were shown in Table 5 and Figures 1 to 3.

Parameter estimates

From Table 5, we observed that all three estimation procedures yielded same values of Δ for all scenarios. The estimates of variance components are close to their true values from all three methods. In nearly every

Table 5. Parameter estimates of MOM, REML and MLM (RIGLS) from 500 runs.

Scenario	Parameter settings	MOM		REML		MLM	
		Mean estimation	Standard error	Mean estimation	Standard error	Mean estimation	Standard error
1	$\Delta = 0.05$	0.052	0.046	0.052	0.046	0.052	0.046
	$\sigma^2_{BR} = 0.09$	0.095	0.034	0.092	0.033	0.093	0.032
	$\sigma^2_{BT} = 0.09$	0.092	0.036	0.090	0.035	0.092	0.033
	$\sigma^2_{WR} = 0.09$	0.091	0.021	0.088	0.020	0.087	0.019
	$\sigma^2_{WT} = 0.09$	0.093	0.022	0.090	0.021	0.089	0.019
2	$\Delta = 0.05$	0.052	0.055	0.052	0.055	0.052	0.055
	$\sigma^2_{BR} = 0.36$	0.375	0.101	0.364	0.098	0.364	0.098
	$\sigma^2_{BT} = 0.36$	0.369	0.103	0.359	0.100	0.359	0.100
	$\sigma^2_{WR} = 0.09$	0.091	0.021	0.088	0.020	0.088	0.020
	$\sigma^2_{WT} = 0.09$	0.093	0.022	0.090	0.021	0.090	0.020
3	$\Delta = 0.05$	0.054	0.076	0.054	0.076	0.054	0.076
	$\sigma^2_{BR} = 0.25$	0.264	0.095	0.257	0.093	0.259	0.088
	$\sigma^2_{BT} = 0.25$	0.256	0.099	0.249	0.096	0.255	0.091
	$\sigma^2_{WR} = 0.25$	0.252	0.058	0.244	0.056	0.241	0.053
	$\sigma^2_{WT} = 0.25$	0.259	0.060	0.251	0.057	0.246	0.054
4	$\Delta = 0.05$	0.053	0.092	0.053	0.092	0.053	0.092
	$\sigma^2_{BR} = 1$	1.041	0.279	1.012	0.272	1.011	0.271
	$\sigma^2_{BT} = 1$	1.026	0.286	0.997	0.278	0.998	0.278
	$\sigma^2_{WR} = 0.25$	0.252	0.058	0.244	0.056	0.244	0.056
	$\sigma^2_{WT} = 0.25$	0.259	0.060	0.251	0.057	0.250	0.057
5	$\Delta = 0.223$	0.225	0.046	0.225	0.046	0.225	0.046
	$\sigma^2_{BR} = 0.09$	0.095	0.034	0.092	0.033	0.093	0.032
	$\sigma^2_{BT} = 0.09$	0.092	0.036	0.090	0.035	0.092	0.033
	$\sigma^2_{WR} = 0.09$	0.091	0.021	0.088	0.020	0.087	0.019
	$\sigma^2_{WT} = 0.09$	0.093	0.022	0.090	0.021	0.089	0.019
6	$\Delta = 0.223$	0.225	0.055	0.225	0.055	0.225	0.055
	$\sigma^2_{BR} = 0.36$	0.375	0.101	0.364	0.098	0.364	0.098
	$\sigma^2_{BT} = 0.36$	0.369	0.103	0.359	0.100	0.359	0.100
	$\sigma^2_{WR} = 0.09$	0.091	0.021	0.088	0.020	0.088	0.020
	$\sigma^2_{WT} = 0.09$	0.093	0.022	0.090	0.021	0.090	0.020
7	$\Delta = 0.223$	0.227	0.076	0.227	0.076	0.227	0.076
	$\sigma^2_{BR} = 0.25$	0.264	0.095	0.257	0.093	0.259	0.088
	$\sigma^2_{BT} = 0.25$	0.256	0.099	0.249	0.096	0.255	0.091
	$\sigma^2_{WR} = 0.25$	0.252	0.058	0.244	0.056	0.241	0.053
	$\sigma^2_{WT} = 0.25$	0.259	0.060	0.251	0.057	0.246	0.054

Table 5. Contd.

	$\Delta = 0.223$	0.226	0.092	0.226	0.092	0.226	0.092
	$\sigma^2_{BR} = 1$	1.041	0.279	1.012	0.272	1.011	0.271
8	$\sigma^2_{BT} = 1$	1.026	0.286	0.997	0.278	0.998	0.278
	$\sigma^2_{WR} = 0.25$	0.252	0.058	0.244	0.056	0.244	0.056
	$\sigma^2_{WT} = 0.25$	0.259	0.060	0.251	0.057	**0.250**	**0.057**

corresponding estimates from REML and RIGLS (we use the term exchangeable with MLMs), while the latter two showed similar results as expected because RIGLS provides REML estimators. In cases of scenarios 1, 3, 5 and 7 where the between-subject variances are the same as within-subject variances, the between-subject variances estimated from MLMs are slightly larger than that of REML, while the within-subject variances from MLM are slightly smaller. This trend is not evident in scenarios 2, 4, 6 and 8.

Assessment of bias

Figure 1 graphically summarizes the bias of variance components versus their true population values. It can be scenario, estimates from the MOM were greater than seen that the biases for all three methods are rather small. However, MOM has the largest biases for both between-subject variances and within-subject variances, that is consistent with the previous simulation results (Patterson and Jones, 2002), whilst REML and RIGLS showed similar bias. In Figure 1, we can see that the bias for MOM is consistently larger (in magnitude), and increases steeply when the variance increases. This suggests that compared to RIGLS and REML, the MOM seems more prone to overestimate as the drug effects become more highly variable. RIGLS and REML estimates demonstrated less bias than MOM with similar patterns in bias as drug effects change in the simulation.

MSE

The MSE from three methods are quite small, suggesting good performance of all methods. Overall, the differences observed in MSE could be by chance in the simulation sample and ignorable in practice. MLM estimates are very close to REML with slightly better accuracy or less bias in some parameter estimates in certain scenarios based on the simulation samples.

Coverage of 95% CI

Three methods demonstrated differences in the CI coverage and those differences seemed grouped by scenarios 1, 3, 5 and 7, and scenarios 2, 4, 6 and 8. So we drew Figures 2 and 3 for these situations separately. Regarding the four variance components, although, the values are different among Figures 2 and 3, they showed the same patterns. The coverage of all the four variance components are constant, completely unaffected by the increasing drug variability. MOM gains the largest coverage, while MLM and REML showed similar results which are close to the nominated 95% for σ^2_{BR}, σ^2_{BT} and σ^2_{WR}, while for σ^2_{WT}, MLM has the same coverage with MOM, which is slightly lower than REML.

In summary, for the assessment of PBE, we can still conclude that the MLM and REML showed similar performance characteristics, and the latter two performed better than MOM.

DISCUSSION

This paper served two purposes; to propose multilevel models (MLMs) using the RIGLS algorithm for evaluating and estimating parameters to assess PBE, and to compare statistical properties of PBE estimators between MLMs and current approaches recommended by the FDA. To our knowledge, this paper is the first to present findings from simulation using the RIGLS algorithm in the software package MLwiN to analyze bioequivalence data. The results indicate that the performance of MLM (RIGLS) appeared to be much comparable to the existing REML procedure, because RIGLS is equal to REML in large sample (Goldstein, 1989). These two procedures are frequently indistinguishable and often provide better performance than MOM.

However, MLM exhibits three advantages over the existing methods. First of all, it has the ability to decompose the total variance into components for complex data structure, that is, more than three levels of hierarchy, more than two treatment and more than two periods.

The second advantage is its easy extension for assessing not only PBE, but also IBE and ABE. IBE is a criterion for deciding if a patient who is currently being treated with R can be switched to T. When IBE is considered, it assesses an aggregate measure involving the means and variance of T and R, as well as the subject-by-formulation interaction. The key difference from PBE is

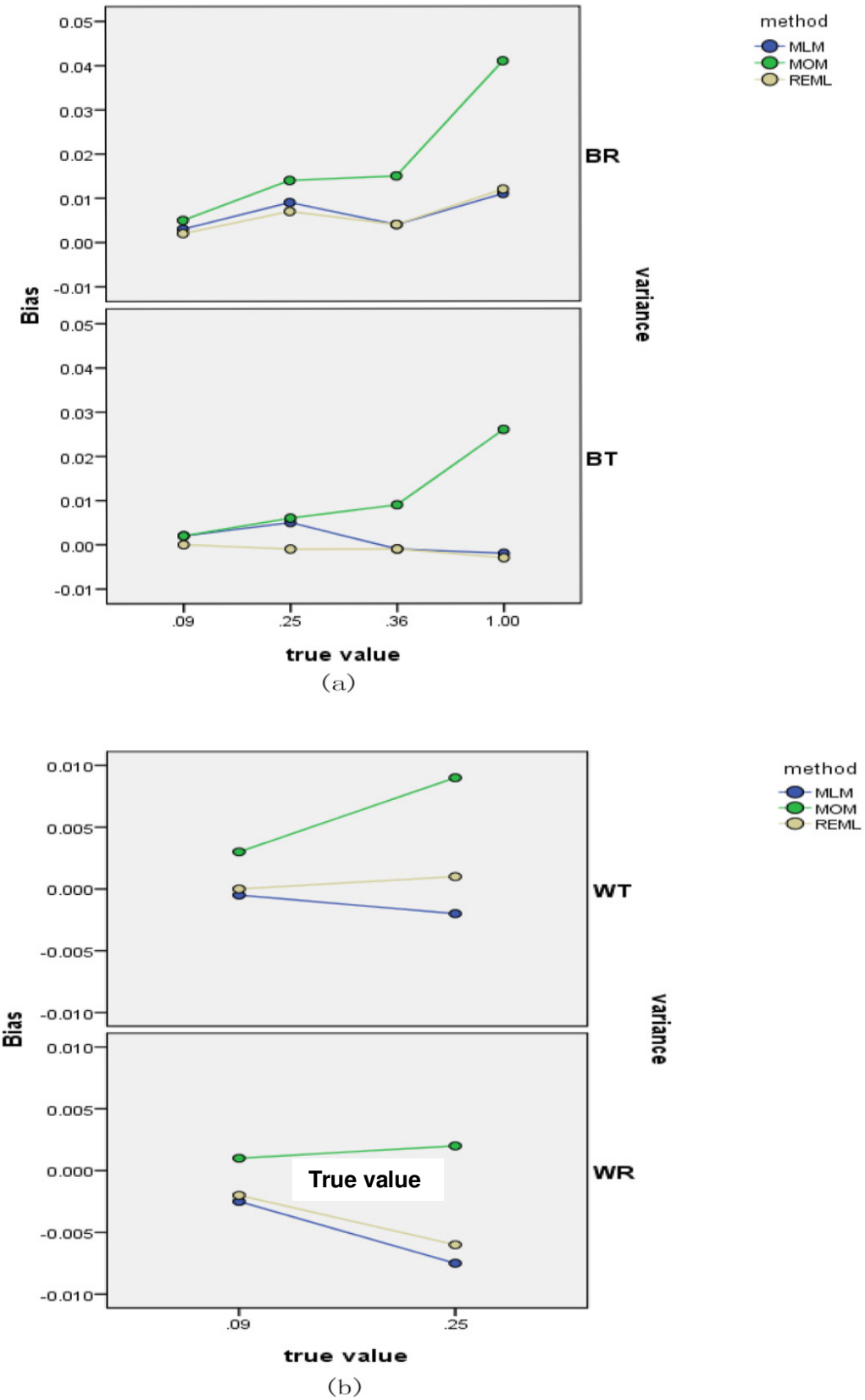


Figure 1. The mean bias in the estimates of four variance components by MLM, MOM and REML. BT stands for σ^2_{BT} , BR for σ^2_{BR} , WT for σ^2_{WT} ,and WR for σ^2_{WR} .

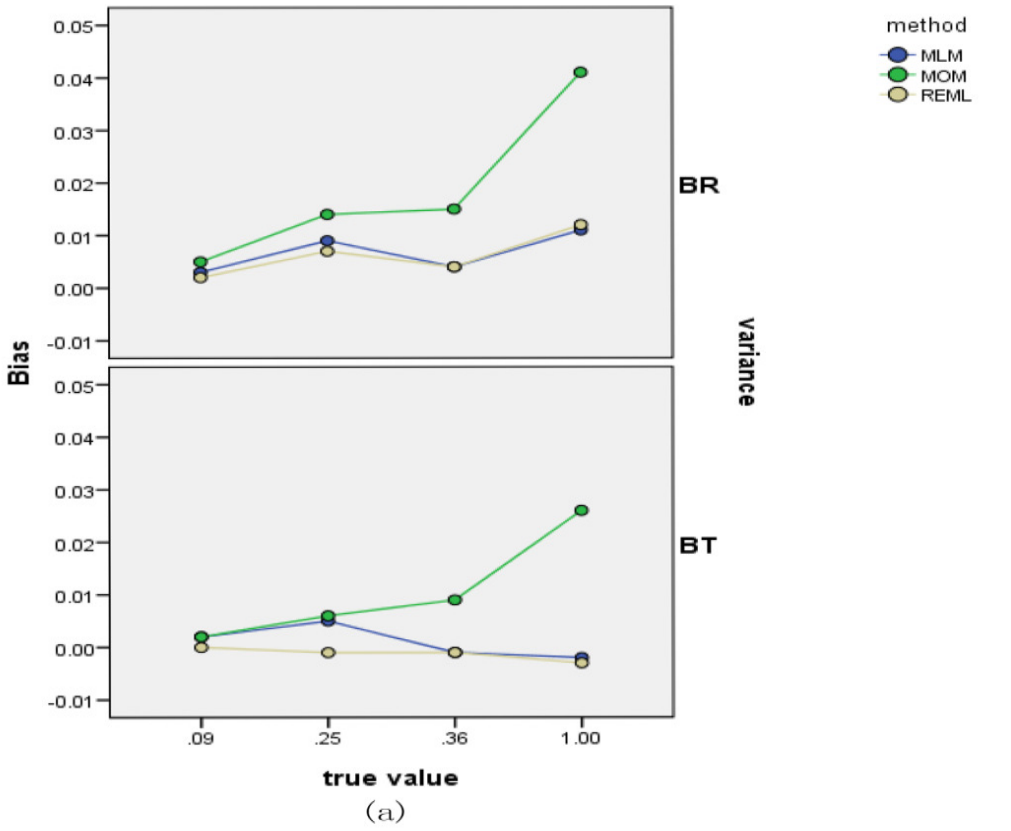


Figure 1. The mean bias in the estimates of four variance components by MLM, MOM and REML. BT stands for σ^2_{BT}, BR for σ^2_{BR}, WT for σ^2_{WT}, and WR for σ^2_{WR}.

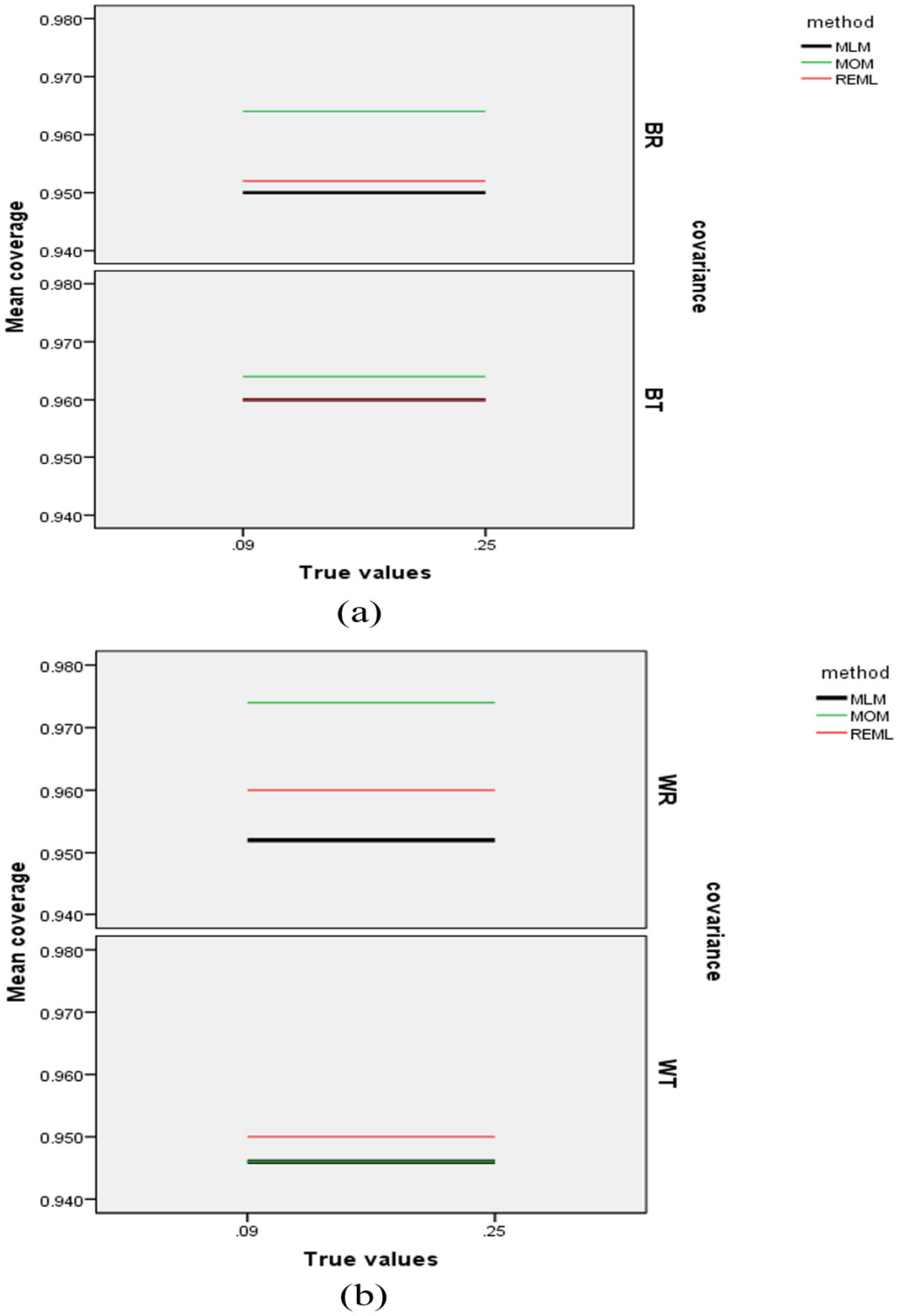
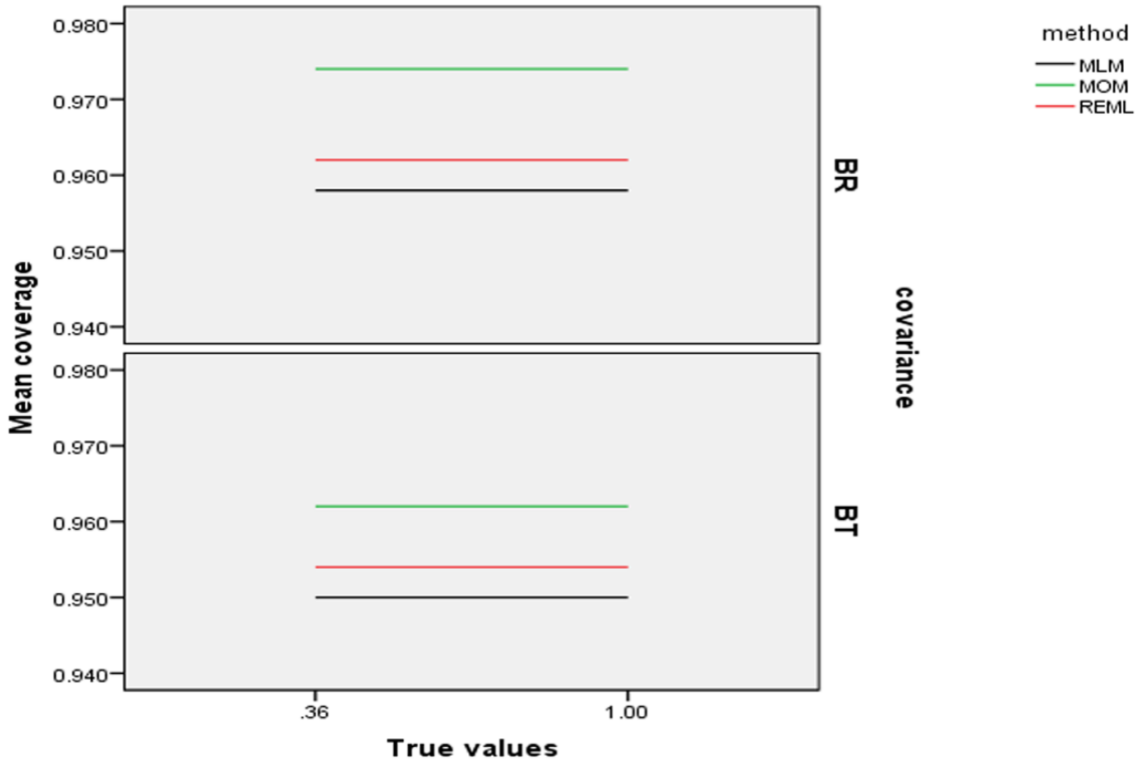
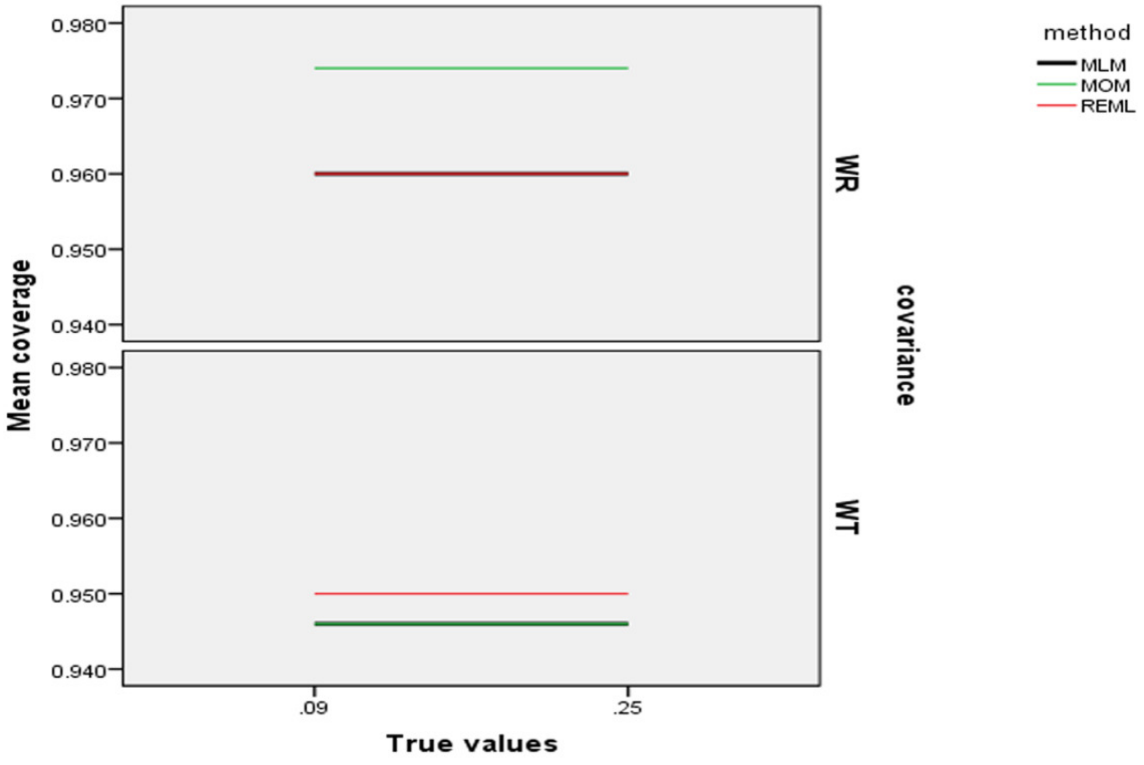


Figure 2. Coverages of 95% CI for four variances components according to MLM, MOM and REML on Scenario 1, 3, 5, 7.



(a)



(b)

Figure 3. Coverages of 95% CI for four variances components according to MLM, MOM and REML on Scenario 2, 4, 6, 8.

that IBE assumed individual difference in responding to formulation. Such difference can be captured and measured by the subject-by-formulation interaction σ_D^2 :

$$\sigma_D^2 = \mathrm{var}(\eta_{jT} - \eta_{jR}) = \sigma_{BT}^2 + \sigma_{BR}^2 - 2\rho\sigma_{BT}\sigma_{BR},$$

Where η_{iT} and η_{iR} are the random subject effect for subject j on T and R especially.

Multilevel model is readily expanded for assessing IBE. For instance, consider a two-level random slope model as an example, similar to the one we mentioned in simulation result. In Equation 4, the random effect u_{1j} is the slope residual in subject *j*, which measured the subject-specific increment to the treatment effect. It is easy to prove that $u_{1j} = \eta_{jT} - \eta_{jR}$, $\mathrm{var}(u_{1j}) = \mathrm{var}(\eta_{jT} - \eta_{jR}) = \sigma_D^2$. Therefore the key ingredient in the IBE criteria σ_D^2 can be expressed as σ_{u1}^2, which is a variance component in multilevel model. The extension of MLMs for assessing IBE is straightforward. A preliminary examination of MLMs in BE, including assessing IBE by Shen et al. (2009), has been published most recently. As the study only illustrated the MLMs approach by applying it on real example, further simulation studies to demonstrate the performance of such model in assessing IBE will be helpful.

The third advantage is the flexibility to cover multiple endpoints for simultaneous bioequivalence assessment. It should be noted that, in the example of FDA dataset illustrated in this paper, both AUC and C_{max} met the criterion for establishment of PBE. However, in practice, it is not uncommon to pass AUC but fail C_{max}. In this case, the regulatory authorities and researchers proposed several alternative measures (Wang et al., 1999; Ghosh and Gönen, 2008; Chen et al., 2001). One proposal was that for highly correlated AUC and C_{max}, one should obtain a combined estimate of the drug effects by considering the two outcomes simultaneously. In a multilevel model framework, data with multiple outcomes at certain time period within-subject can be viewed as a 3-level structure: the outcomes measured on each occasion are considered nested at the lowest level, within replicated measurement (now at level 2) within subject (now at level 3). MLMs approach can simply link the marginal models of AUC and C_{max} through a variance-covariance structure of the two at level 1 to form a simultaneous model to assess either ABE or PBE. Further simulation study should be conducted to examine performance of multivariate MLMs and the statistical features of the parameter estimates.

This research offers many exciting new directions for future research. A number of issues in the proposed MLMs approach are the remaining debatable. First of all, the 95% upper bound of the linearised PBE criterion in this study was calculated based on asymptotically normal assumption. The standard error for each variance component and function of a particular variance provided directly by the software MLwiN of RIGLS algorithm is based on asymptotic properties and may be unreliable when sample size is small. The advantage of MLM's capacity to assess IBE or multivariate outcomes would be penalized by more complex variance-covariance structure, more complex criterion, hence more uncertainty in the upper bound of the criterion. Recently FDA becomes more accepting of Bayesian (Li and Xu, 2011; Ashby, 2006). The use of the Bayesian approach allows us to obtain credible intervals and density plots for both random effect variances and standard deviations. The MLwiN was built in tools for MCMC and bootstrapping modeling. Hence, further research should explore those models within the MLMs framework in estimating upper bound of the PBE and its extendibility for IBE or multivariate MLMs criterion. Other methodological study in order to establish the MLMs approach in the BE field could be the rate of type 1 error over the effect size, sample size determination and the power of test. Further investigations in those areas are also suggested.

ACKNOWLEDGEMENTS

Professor Min Yang was supported by a grant from the China Medical Board in New York (00-722). Zhuo-Zhi Shen was supported by a scholarship grant from the China Scholarship Council (2010624083).

REFERENCES

Ashby D (2006). Bayesian statistics in medicine: a 25 year review. Stat. Med. 25:3589-3631.

Blume HH, Midha KK (1993). Bio-International '92, Conference on Bioavailability, Bioequivalence and Pharmacokinetic Studies. Pharm. Res. 10:1806-1811.

Carrasco JL, Jover L (2003). Assessing individual bioequivalence using the structural equation model. Stat. Med. 22:901-912.

Chen ML, Shah V, Patnaik R, Adams W, Hussain A, Conner D, Mehta M, Malinowski H, Lazor J, Huang SM (2001). Bioavailability and bioequivalence: an FDA regulatory overview. Pharm. Res. 18:1645-1650.

Chen ML, Patnaik R, Hauck W, Schuirmann DJ, Hyslop T, Williams R (2000). An individual bioequivalence criterion: regulatory considerations. Stat. Med. 19:2821-2842.

Chow SC, Liu J (2008). Design and analysis of bioavailability and bioequivalence studies. Chapman & Hall/CRC. Print ISBN: 978-1-58488-668-6. eBook ISBN: 978-1-4200-1167-8. DOI: 10.1201/9781420011678.fmatt

Davit BM, Conner DP, Fabian-Fritsch B, Haidar SH, Jiang X, Patel DT, Seo PR, Suh K, Thompson CL, Yu LX (2008). Highly Variable Drugs: Observations from Bioequivalence Data Submitted to the FDA for New Generic Drug Applications. AAPS J. 10(1):148-156.

Dragalin V, Fedorov V, Patterson S, Jones B (2003). Kullback–Leibler divergence for evaluating bioequivalence. Stat. Med. 22:913-930.

Endrenyi L, Taback N, Tothfalusi L (2000). Properties of the estimated variance component for subject-by-formulation interaction in studies of individual bioequivalence. Stat. Med. 19:2867-2878.

FDA (2001). Guidance for Industry: Statistical approaches to establishing bioequivalence. Rockville, MD: US Food and Drug Administration Center for Drug Evaluation and Research.

FDA (2003). Guidance for Industry: Bioavailability and Bioequivalence

Studies for orally administered drug products - General considerations Rockville, MD: US Food and Drug Administration Center for Drug Evaluation and Research.
Ghosh P, Ntzoufras I (2005). Testing Population and Individual Bioequivalence: A Hierarchical Bayesian Approach. Technical report, Department of Statistics, Athens University of Economics and Business.
Ghosh P, Gönen M (2008). Bayesian modeling of multivariate average bioequivalence. Stat. Med. 27:2402-2419.
Goldstein H (1986). Multilevel mixed linear model analysis using iterative generalized least squares. Biometrika 73:43-56.
Goldstein H (1989). Restricted unbiased iterative generalized least-squares estimation. Biometrika 76:622-623.
Goldstein H (2010). Multilevel statistical models: Wiley
Haidar S, Davit B, Chen ML, Conner D, Lee L, Li Q, Lionberger R, Makhlouf F, Patel D, Schuirmann D (2008). Bioequivalence Approaches for Highly Variable Drugs and Drug Products. Pharm. Res. 25(1):237-241.
Li X, Xu Y (2011). Guest Editors' Note: Bayesian Approaches for Medical Product Evaluation—Today and Tomorrow. J. Biopharmaceut. Stat. 21:869-870.
McNally RJ, Iyer H, Mathew T (2003). Tests for individual and population bioequivalence based on generalized p-values. Stat. Med. 22:31-53.
Patterson SD, Jones B (2002). Replicate Designs and Average, Individual, and Population Bioequivalence. Simulation Assessment of Performance of Novel Procedures and the Proposed FDA Methods for Bioequivalence Assessment. GlaxoSmithKline Report.
Rashbash J, Steele F, Browne W, Prosser B (2008). A User's Guide to MLwiN. Version 2.10. London: Centre for Multilevel Modelling. In Institute of Education, University of London. Available at www.cmm.bristol.ac.uk
Riviere JE, Papich MG (2009). Veterinary pharmacology and therapeutics. Blackwell Publication. pp. 835-864
Shao J, Chow SC, Wang B (2000). The bootstrap procedure in individual bioequivalence. Stat. Med. 19(20):2741-2754.
Wang W, Gene Hwang J, Dasgupta A (1999). Statistical tests for multivariate bioequivalence. Biometrika 86:395-402.
Willavize SA, Morgenthien EA (2006). Comparison of models for average bioequivalence in replicated crossover designs. Pharmaceut. Stat. 5:201-211.