STATISTICS IN MEDICINE
Statist. Med. 2005; **24**:883–891
Published online 23 November 2004 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/sim.1977

Comparing the concentration curves directly in a pharmacokinetics, bioavailability/bioequivalence study

Jason J. Z. Liao*,†

Merck Research Laboratories, P.O. Box 4, West Point, PA 19486, U.S.A.

SUMMARY

In a traditional pharmacokinetics (PK), bioavailability (BA)/bioequivalence (BE) study, the same number of time points and sampling times are used for each subject. Often, an indirect inference is then made on some PK parameters such as area under the plasma concentration curve (AUC), maximum plasma concentration (C_{max}), time to maximum plasma concentration (T_{max}) or half-life. However, since these PK parameters are summarized from repeated measurements, a lot of information can be lost. The indirect inferences on some PK parameters are not always accurate. Taking the repeated measurements of the concentration curve into consideration, a functional linear model has been developed to compare concentration curves directly instead of the PK parameters. Considering the nature of repeated measurements, a multiple testing procedure is proposed to assess the equality of two concentration curves. A real data set is used to demonstrate the proposed procedure. Copyright © 2004 John Wiley & Sons, Ltd.

KEY WORDS: concentration curve; functional linear model; sampling time; repeated measurement; multiple testing

1. INTRODUCTION

In a pharmacokinetics (PK), comparative bioavailability (BA)/bioequivalence (BE) study in humans, the blood or plasma concentration curve (or profile) is often used to study the absorption and the elimination of a drug after it has been administered. The plasma or blood concentration-time curve can be characterized by taking blood samples at various time points after the drug administration. It is common to model plasma drug concentration as a function of time using compartmental model techniques based on physiological theories. The blood concentration curve is then used for pharmacokinetics modelling, and for a bioavailability and bioequivalence study. To compare the blood concentration curves between different formulations of a drug or between different administration circumstances (e.g. with/without food or simultaneous/staggered administration with other drugs), one usually estimates the PK

*Correspondence to: Jason J. Z. Liao, Merck Research Laboratories, P.O. Box 4, West Point, PA 19486, U.S.A.
†E-mail: jason_liao@merck.com

Received November 2003
Accepted June 2004

parameters such as area under the plasma concentration curve (AUC), maximum plasma concentration ($C_{\max}$), time to maximum plasma concentration ($T_{\max}$) and half-life, either by using the simple trapezoidal rule [1], or by fitting a parametric compartmental model [2]. One then carries out further statistical inferences indirectly based on the derived PK parameters. Thus, the conclusions from the parametric modelling approach are subject to the assumptions of first order or linear kinetics on the rates of transfer between compartments and the number of compartments considered in the model. These assumptions may not be attainable or appropriate in the initial tests of drug development.

It is well known that the time points sampled in the blood profile are very important. The selection of the various time points has a significant effect on the estimate of the PK parameters. Therefore, the FDA regulation requires that sampling be continued through at least three half-lives of the active drug ingredient so that majority of the elimination is completed and the remaining area beyond the last observation is negligible. In the current practice, the same number of time points and sampling times are used for each subject. Inference based on the PK parameters assumes that the parameters uniquely determine the PK profile. This assumption may not hold for many practical problems. Hence, indirect inferences on the PK derived parameters are not always accurate. In addition, the choice of the time points is very critical for the PK parameters; inaccurate time points may not truly reflect the structure of the true blood profile. There are often too few predetermined sampling time points to have a reliable estimate of the PK parameters, even though theoretically the assumed pharmacokinetic model may adequately describe the observed blood or plasma concentration-time curve [3].

Consider a bioavailability study. A marketed 100-mg tablet was found to have slower dissolution at early time points in initial batches when compared to the exact weight submultiple 50-mg tablet. Then a study was conducted to determine the relative bioavailability of a 100-mg tablet versus 2×50-mg tablets at one clinical centre. Thus, the blood profiles from a 100-mg tablet and 2×50-mg tablets need to be compared. The traditional approach for testing whether two blood profiles are the same is to use the trapezoidal rule or parametric model to show whether there is a significant difference between two profiles in terms of AUC and/or other PK parameters. A common procedure usually involves the following stages [4]:

- Running an AB/BA cross-over in which so-called test (T) and reference (R) 'formulations' are compared for each subject with the same observation time points.
- Log transforming the AUCs measured in the trial.
- Fitting a linear model to the log-AUCs in which subject and the period effects are eliminated to produce an estimate of the 'formulation effect'.
- Estimating the standard error of the estimated formulation effect.
- Comparing the results to pre-established limits of equivalence, δ_1 (a lower limit) and δ_2 (an upper limit). It is customary to use the limits $\log(0.8)$ and $\log(1.25)$ on the log-AUC scale, so that $\delta_1 = -\delta_2$ where $\delta_2 = \delta = 0.223$.

As mentioned by Westlake [5], equivalence of the PK parameters does not necessarily mean that the blood profiles are equivalent. It is not difficult to construct two different blood profiles where AUC, $C_{\max}$ and $T_{\max}$ are the same. Mauger and Chinchilli [6] also gave the same concern and proposed a new approach. However, they used a different PK derived parameter by summarizing the profile. As such, their approach is in the same category as the traditional PK parameters. By summarizing the blood concentration profile as the PK parameters, much information is lost since the concentration profile is a repeated measurement

and they are correlated [7]. Hence, a direct non-parametric approach to compare the blood profiles is needed. For this purpose, Liao [8] proposed a linear functional model to compare the concentration curve directly. In this paper, we will continue the work in this direction. In Section 2, the linear functional model is reviewed first, and then a multiple testing procedure is proposed to assess the equivalence of two concentration curves. An example is used to demonstrate the multiple testing procedure in Section 3. Discussions follow in Section 4.

2. A MULTIPLE TESTING PROCEDURE

To obtain the concentration curve, the cross-over design is commonly used with the same sampling time points for each subject. Let $X(t)$ and $Y(t)$ be the two concentration curves. The repeated measurements are $(t_{ij}, x_{ij} = X(t_{ij}), y_{ij} = Y(t_{ij}))$, $i = 1, \ldots, m$, $j = 1, \ldots, n_i$, where m is the number of subjects (patients) and n_i is the number of (X, Y) observations for the ith subject. Note that the n_i's are not necessarily the same for all subjects, and t_{ij} and $t_{i'j}$ are not necessarily the same for $i \neq i'$. Instead of making an indirect inference on the PK parameters, Liao [8] proposed a linear functional model to directly assess the blood concentration profiles on the raw scale using the nature of repeated measurements. The linear functional model takes the correlation among the concentrations into account. Unlike the current practice in which requires the same number of sampling times and time points for each subject, it has flexible sampling time points and does not require equal number of time points for each subject. Regarding this, our recommendation is as follows. First, two or three subjects from the subject population are randomly selected to form a group. Then the subjects from the same group receive the same time points while subjects from different groups receive different time points. Thus, more time points are sampled over the whole time range.

For a fixed time point t, a linear measurement error model was used for $Y(t)$ and $X(t)$. However, the coefficients are different at each time point. Please note that $Y(t)$ ($X(t)$) itself is a non-linear function of time t with a repeated measurement structure. By combining all time points, the two profiles $Y(t)$ and $X(t)$ are approximately linearly related and their relationship can be modelled as follows:

$$Y(t) = \alpha(t) + \beta(t)X^0(t) + \varepsilon(t) \tag{1}$$

$$X(t) = X^0(t) + \delta(t) \tag{2}$$

where $X^0(t)$ is the unobserved value at time point t, $\varepsilon(t)$ and $\delta(t)$ are two independent Gaussian processes with mean zero and covariance function $r(s,t)$ where s is another time point. The independence assumption of $\varepsilon(t)$ and $\delta(t)$ within a subject should hold because the two processes, $\varepsilon(t)$ and $\delta(t)$, are from the same subject with a long enough washout period separating them. The goal is to estimate $\alpha(t)$ and $\beta(t)$ and make some inferences.

Remarks

(1) The sampling time points are not necessarily the same for all subjects in the model defined by equations (1) and (2). Therefore, there is a better chance to catch the true $T_{\max}$ (therefore, $C_{\max}$) and produce more reliable estimates for the PK parameter of each subject.

(2) The correlation among the concentrations is characterized by the covariance function $r(s,t)$.

Let $1_{n_i}=(1,\ldots,1)^{\mathrm{T}}$, $X_i=(x_{i1},\ldots,x_{in_i})^{\mathrm{T}}$, $Y_i=(y_{i1},\ldots,y_{in_i})^{\mathrm{T}}$ and $K_i(t)$ be a diagonal matrix with elements $\{K((t-t_{i1})/h),\ldots,K((t-t_{in_i})/h)\}$, where $K(w)$ is a kernel function and h is the bandwidth. Then,

$$\hat{\alpha}(t)=\bar{Y}_{\cdot}-\hat{\beta}(t)\bar{X}_{\cdot} \tag{3}$$

where $\bar{X}_{\cdot}=(1/a_1)(1/m)\sum_{i=1}^{m}1_{n_i}^{\mathrm{T}}K_i(t)X_i$, $\bar{Y}_{\cdot}=(1/a_1)(1/m)\sum_{i=1}^{m}1_{n_i}^{\mathrm{T}}K_i(t)Y_i$ and $a_1=(1/m)\sum_{i=1}^{m}1_{n_i}^{\mathrm{T}}K_i(t)1_{n_i\cdot}$ and

$$\hat{\beta}(t)=\begin{cases}\dfrac{m_{yy}-m_{xx}+\sqrt{(m_{yy}-m_{xx})^2+4m_{xy}^2}}{2m_{xy}} & \text{if } m_{xy}\neq 0\\ 0 & \text{if } m_{xy}=0\end{cases} \tag{4}$$

where $m_{xx}=(1/m)\sum_{i=1}^{m}(X_i-1_{n_i}\bar{X}_{\cdot})^{\mathrm{T}}K_i(t)(X_i-1_{n_i}\bar{X}_{\cdot})$, $m_{yy}=(1/m)\sum_{i=1}^{m}(Y_i-1_{n_i}\bar{Y}_{\cdot})^{\mathrm{T}}K_i(t)(Y_i-1_{n_i}\bar{Y}_{\cdot})$ and $m_{xy}=(1/m)\sum_{i=1}^{m}(X_i-1_{n_i}\bar{X}_{\cdot})^{\mathrm{T}}K_i(t)(Y_i-1_{n_i}\bar{Y}_{\cdot})$ [8].

If one knows the true value $X_i^0(t)$, then one can estimate the covariance function $r(s,t)$ by $(1/m)\sum_{i=1}^{m}(X_i-X_i^0)^{\mathrm{T}}(s)(X_i-X_i^0)(t)$ or the median for the variance at a time point. Similar to Fuller's [9] linear measurement error model, one can estimate $X_i^0(t)$ by $(X_i(t)+\hat{\beta}(t)(Y_i(t)-1_{n_i}\hat{\alpha}(t)))/1+\hat{\beta}^2(t)$.

In a regression model setting, an important inference is to test the equivalence of the parameters to some specified values. Hence, the general hypothesis test is $H_0:\alpha(t)=a_0(t)$, $\beta(t)=b_0(t)$, for all $t\in[t_L,t_U]$, where $[t_L,t_U]$ is the range of interest. In most applications, special cases of the general hypothesis test, such as $a_0(t)$ and $b_0(t)$ held constant against t, are of interest. One particular special case is to test if the two profiles are equivalent, that is, $H_0:\alpha(t)=0$, $\beta(t)=1$, for all $t\in[t_L,t_U]$, where $[t_L,t_U]$ is the range of interest. Please note that in current bioequivalence practice, the null and alternative hypotheses H_0 and H_{a} are often switched, which correspondingly implies the type II error being controlled instead of the type I being controlled. In order to mimic the current practice of bioequivalence study, the confidence interval for $\log(Y)(t)-\log(X)(t)$ at all time points can be constructed and then the current common boundaries $\log(0.8)$ and $\log(1.25)$ or modified boundaries can be used to assess the equivalence.

To make a decision, a confidence interval can be constructed for both $\alpha(t)$ and $\beta(t)$ to determine whether the hypothesized $a_0(t)$ and $b_0(t)$ are within the corresponding confidence intervals. However, a formal testing procedure can be developed as follows.

Let $t_1,\ldots,t_k$ be the distinct time points pooled over all subjects. Suppose $\hat{\theta}_j=(\hat{\alpha}(t_j),\hat{\beta}(t_j))^{\mathrm{T}}$ follows a normal distribution with mean $\theta(t_j)$ and covariance matrix Σ_j, $j=1,\ldots,k$ [9]. For testing $H_0:\theta(t)=\theta_0$, for all $t\in[t_L,t_U]$ where $\theta_0=(a_0,b_0)^{\mathrm{T}}$, H_0 is rejected if

$$(\hat{\theta}_j-\theta_0)^{\mathrm{T}}\Sigma_j^{-1}(\hat{\theta}_j-\theta_0)>\chi_2^2,\quad j=1,\ldots,k$$

holds at an adjusted significant level.

For this purpose, a sharper Bonferroni procedure for multiple tests [10] is adopted as follows. Let $p_1,\ldots,p_k$ be the corresponding p-values for the test at the k distinct time points. Then rank these p-values from the largest to the smallest such as $p_{(k)}\geqslant p_{(k-1)}\geqslant\cdots\geqslant p_{(2)}\geqslant p_{(1)}$ and compare the ranked p-values against the following α-levels: $\alpha,\alpha/2,\ldots,\alpha/(k-1),\alpha/k$,

respectively. If all p-values are greater than the corresponding adjusted α-levels, then the two profiles are declared not significantly different from each other.

Remarks

(1) The above testing procedure starts with the largest p-value $p_{(k)}$. If $p_{(k)} \leqslant \alpha$, then the procedure ends and one can declare that the two profiles are significantly different from each other over the whole time range. Otherwise, $p_{(k-1)}$ is compared to $\alpha/2$. And so on.

(2) In practice, people may be interested in an overall assessment using only one p-value instead of the assessment at each distinct time point. Then the p-values from different sampling time points can be combined to achieve this goal [11].

3. AN EXAMPLE

Consider the following study for an angiotensin II receptor drug for the treatment of hypertension with the marketed 100-mg tablets. Initial batches were found to have slower dissolution at early time points when compared to the exact weight submultiple 50-mg tablet. However, both 50- and 100-mg tablets were manufactured at the same site, maintained the same ratio of active to inactive ingredients and met the same release specification. Therefore, a study was conducted to assess the impact of slower dissolution at early dissolution time points *in vitro* to the *in vivo* performance by determining the relative bioavailability of 100-mg tablets versus two 50-mg tablets. The study was an open-label, randomized, two-period cross-over to determine the relative bioavailability of a 100-mg tablet verse 2×50-mg tablets to be conducted at one clinical centre. Twenty healthy male and female volunteers were enrolled. Females were of non-childbearing potential. Each subject received a single oral dose of each of the two treatments (Treatment A: one 100-mg tablet; Treatment B: two 50-mg tablets) in a two-period cross-over fashion. Plasma for treatments A and B were collected over the 36 hours post-dose. There was a minimum 7-day washout between periods. Plasma samples were collected in heparinized containers at pre-dose and at 0.25, 0.5, 0.75, 1, 1.25, 1.5, 1.75, 2, 3, 4, 5, 6, 7, 8, 10, 12, 18, 24, 30, and 36 hours post-dose. The samples were kept at $-20°$C. A comparative analysis of PK profiles of Treatment A and Treatment B was monitored.

Let $X(t)$ be the PK profile (drug concentration curve) of Treatment A and $Y(t)$ be the corresponding PK profile of Treatment B. The objective is to determine whether there is any significant difference between the two PK profiles of $X(t)$ and $Y(t)$. Using the traditional trapezoidal rule method, there is no significant difference between $X(t)$ and $Y(t)$ in terms of log-AUC parameters using the approach outlined in Section 1. Now the functional linear model defined by equations (1) and (2) is applied to this data set. This model addresses the comparison of two PK profiles directly and provides an alternative as a robust method against any model mis-specification and any assumption of the concentration curve.

First, a bandwidth $h = \delta \times T$ is selected. In this example, $T = 36$ hours. For this purpose, $\sum_{ij} (\mathrm{Var}(\hat{\beta})(t_{ij}) + \mathrm{BIAS}^2(\hat{\beta})(t_{ij}))$ is minimized with the choice of a Gaussian kernel following Liao [8]. Listed in Table I are the sums of the MSEs at bandwidth $h = \delta \times T$ for 10 different δ values. Table I shows that the minimum MSE is achieved with $\delta = 0.2$. Note that δ is a percentage of the whole time interval; therefore, $\delta = 0.2$ indicates that 20 per cent of the data should be used to estimate the parameters at each time point. The estimated slope and its 95 per cent confidence intervals are shown in Figure 1.

Table I. Bandwidth selection for the example.

δ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
$100 \times (\sum_{ij}(\mathrm{Var}(\hat{\beta})(t_{ij}) + \mathrm{BIAS}^2(\hat{\beta})(t_{ij}))$	11.094	10.126	10.392	10.477	10.505	10.512	10.512	10.509	10.506	10.504

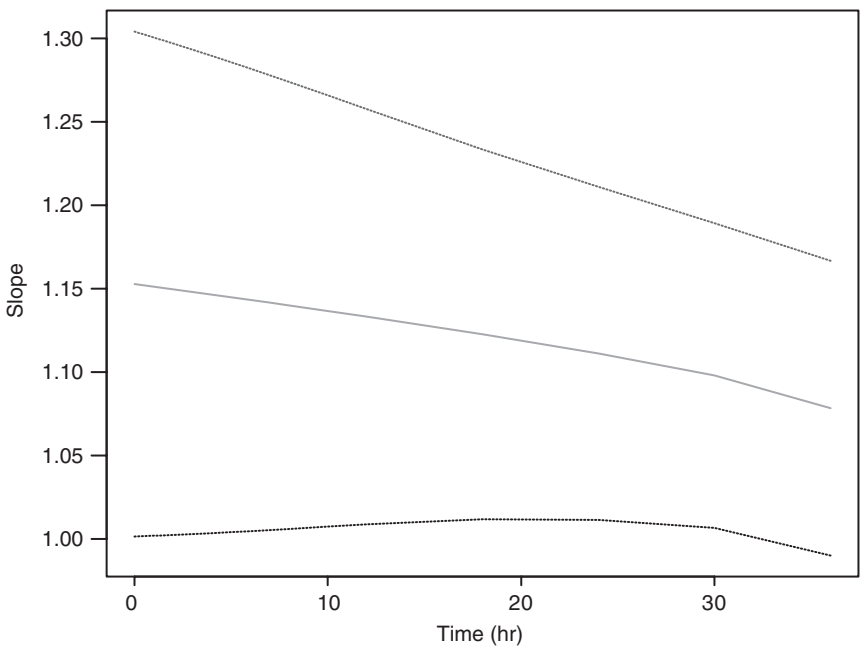

Figure 1. Estimated slope with 95 per cent confidence bands at bandwidth $\delta = 0.2$.

Figure 1 indicates that the confidence interval of $\beta(t)$ does not contain 1 for the entire time range. Therefore, the two PK profiles $X(t)$ and $Y(t)$ are statistically significantly different from each other. The confidence interval evaluation gives a non-equality conclusion instead of the equivalence conclusion by the PK parameter (AUC) approach. Figure 2 shows the estimated intercept and slope curves. It indicates that the slope is a decreasing function of time but that the intercept is an increasing function of time.

Now the multiple testing procedure is applied to this data set to assess the equivalence. A χ^2-test is conducted at each of the 21 sampling time points (0, 0.25, 0.5, 0.75, 1, 1.25, 1.5, 1.75, 2, 3, 4, 5, 6, 7, 8, 10, 12, 18, 24, 30, and 36 hours) and the corresponding p-values are 0.1074, 0.1064, 0.1054, 0.1044, 0.1034, 0.1023, 0.1012, 0.1001, 0.0990, 0.0945, 0.0898, 0.0849, 0.0801, 0.0752, 0.0704, 0.0615, 0.0543, 0.0530, 0.0870, 0.1110, and 0.1412, respectively. The comparison of the p-values to the adjusted α-levels is listed in Table II. Table II indicates that all p-values are greater than the corresponding adjusted α-values. Thus, the two profiles have no statistically significant difference. It also shows that the two profiles

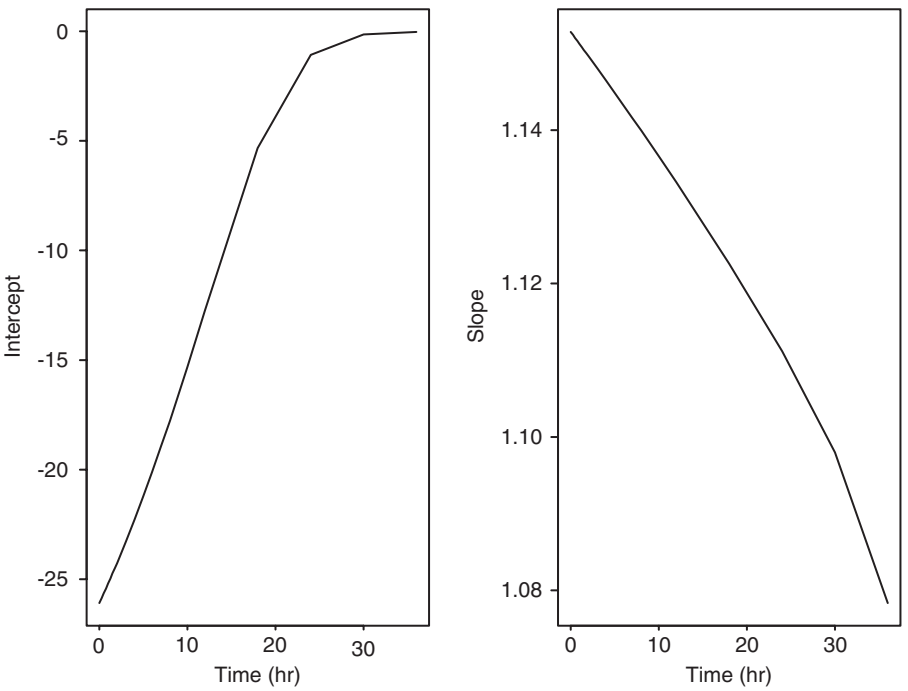

Figure 2. Estimated slope and intercept curves at bandwidth $\delta = 0.2$.

at two ends have a better agreement because of the higher p-values. Please note that multiple testing gives a different conclusion from that of the confidence interval evaluation. It is well known that confidence interval evaluation and p-value evaluation sometimes give different conclusions. The fact that the χ^2-test uses an ellipse acceptance region but the confidence interval assessment uses a rectangle acceptance region may explain the conclusion difference.

4. SUMMARY AND DISCUSSIONS

In this paper, a multiple testing procedure was proposed to compare two concentration curves directly instead of the confidence interval approach. The testing procedure used an elliptical acceptance region as opposed to a rectangular acceptance region used by a confidence interval assessment. The multiple testing procedure gives a p-value for each time points. Therefore, the strength of the agreement between two profiles at each time point can be evaluated. If one wants only one p-value to compare the two concentration curves in general, then the p-values from all time points can be combined from all time points to make an overall conclusion.

It is well known that the sampling time points have a significant effect on the value for the PK derived parameters such as AUC, $C_{\max}$ and $T_{\max}$ calculated through the trapezoidal rule. Slightly different sampling time points around the true $T_{\max}$ can change the results significantly. Ideally, it is desirable to have as many time points as possible. Using the linear functional

Table II. Comparison of p-values to the adjusted α-levels.

Ranked p-values	Adjusted α-levels	$p \leqslant$ adjusted α-level	Time point (hr)
$p_{(21)} = 0.1412$	0.05	No	$t = 36$
$p_{(20)} = 0.1110$	0.05/2	No	$t = 30$
$p_{(19)} = 0.1074$	0.05/3	No	$t = 0$
$p_{(18)} = 0.1064$	0.05/4	No	$t = 0.25$
$p_{(17)} = 0.1054$	0.05/5	No	$t = 0.5$
$p_{(16)} = 0.1044$	0.05/6	No	$t = 0.75$
$p_{(15)} = 0.1034$	0.05/7	No	$t = 1$
$p_{(14)} = 0.1023$	0.05/8	No	$t = 1.25$
$p_{(13)} = 0.1012$	0.05/9	No	$t = 1.5$
$p_{(12)} = 0.1001$	0.05/10	No	$t = 1.75$
$p_{(11)} = 0.0990$	0.05/11	No	$t = 2$
$p_{(10)} = 0.0945$	0.05/12	No	$t = 3$
$p_{(9)} = 0.0898$	0.05/13	No	$t = 4$
$p_{(8)} = 0.0870$	0.05/14	No	$t = 24$
$p_{(7)} = 0.0839$	0.05/15	No	$t = 5$
$p_{(6)} = 0.0801$	0.05/16	No	$t = 6$
$p_{(5)} = 0.0752$	0.05/17	No	$t = 7$
$p_{(4)} = 0.0704$	0.05/18	No	$t = 8$
$p_{(3)} = 0.0615$	0.05/19	No	$t = 10$
$p_{(2)} = 0.0543$	0.05/20	No	$t = 12$
$p_{(1)} = 0.0530$	0.05/21	No	$t = 18$

model, different sampling time points as well as different numbers of sampling points can be used for different subjects and therefore, more time points can be used in the entire time interval. However, that is not practical for the current approach, since too much blood will be drawn for each subject. Thus, our recommendation is as follows. First, several groups are formed by picking two or three subjects randomly to form each group. The subjects from the same group receive the same time points while subjects from different groups receive different time points. Thus, more time points are sampled over the whole time range. Under the linear functional model, the PK parameters can be estimated with more accuracy and the effects of the sampling time points may not be so serious when using the PK parameters estimated from the predicted curves under the linear functional model.

It is well known that the outliers can dramatically affect the conclusion from a statistical inference and that the outliers usually occur in the early and later sampling time points [3]. The linear functional model approach is a regression-type approach, therefore, it can detect the outliers in the raw concentration level or subject level in traditional ways such as through residual analysis (it is currently under further evaluation). However, the current approaches using summarized PK parameters cannot achieve this directly.

When a cross-over design is used in a study, we follow the current practice to assume that there is no sequence effect, which is a reasonable assumption [12]. However, we might easily adjust these effects by including them as covariates in the linear functional model defined by equations (1) and (2). With the link to the current practice in a bioequivalence study, the null and alternative hypotheses H_0 and H_a are often switched, which correspondingly implies the type II error being controlled instead of the type I being controlled. In order to mimic the current practice of bioequivalence study, the confidence interval for $\log(Y)(t) - \log(X)(t)$ at all

time points can be constructed and then the current common boundaries log(0.8) and log(1.25) or modified boundaries can be used to assess the equivalence. This procedure is similar to the current common practice since C_{max} is the concentration at the time T_{max} and since it is one of the values in the concentration curve. Normal and log-normal distributions are two most commonly used distributions in a PK study. Depending on whether the concentration has a normal distribution or a log-normal distribution, direct comparison of concentration curves can be done to raw concentration or log-concentration accordingly. Other methods of directly comparing the concentration curves can also be explored in future research.

ACKNOWLEDGEMENTS

We thank the editor and one referee for their valuable comments which improved the presentation. The thanks also go to Dr Anup Majumdar and Dr Saurabh Mukhopadhyay for their comments.

REFERENCES

1. Rowland M, Tozer TN. *Clinical Pharmacokinetics*, *Concepts and Applications*. Williams & Wilkins: Baltimore, 1995.
2. Davidian M, Giltinan DM. *Nonlinear Models for Repeated Measurement Data*. Chapman & Hall: New York, 1995.
3. Chow SC, Liu JP. *Design and Analysis of Bioavailability and Bioequivalence Studies*. Marcel Dekker: New York, 1992.
4. Senn S. Statistical issues in bioequivalence. *Statistics in Medicine* 2001; **20**:2785–2799.
5. Westlake WJ. Bioavailability and bioequivalence of pharmaceutical formulations. In *Biopharmaceutical Statistics For Drug Development*, Peace KE (ed.). Marcel Dekker: New York, 1988.
6. Mauger DT, Chinchilli VM. An alternative index for assessing profile similarity in bioequivalence trials. *Statistics in Medicine* 2000; **19**:2855–2866.
7. Lindsey JK, Jones B, Jarvis P. Some statistical issues in modeling pharmacokinetic data. *Statistics in Medicine* 2001; **20**:2775–2783.
8. Liao JJZ. A functional linear model for comparing two pharmacokinetics profiles. *Journal of Statistical Planning and Inference* 2003; **116**:521–535.
9. Fuller WA. *Measurement Error Models*. Wiley: New York, 1987.
10. Hochberg Y. A sharper Bonferroni procedure for multiple tests of significance. *Biometrika* 1988; **75**:800–802.
11. Kost JT, McDermott MP. Combining dependent *P*-values. *Statistics and Probability Letters* 2002; **60**:183–190.
12. Carrasco JL, Jover L. Assessing individual bioequivalence using the structural equation model. *Statistics in Medicine* 2003; **22**:901–912.

