Letter to the Editor

Bioequivalence Assessment of Topical Ophthalmic Drugs Using Sparse Sampling Pharmacokinetics Data*

YU Yong Pei[1], YAN Xiao Yan[2], YAO Chen[2], and XIA Jie Lai[1,#]

In the development of eye drop medications, difficulty in sampling is a major challenge[1]. Drug concentrations in the aqueous humor (AH) can only be measured when the eye is undergoing surgery, such as cataract replacement. Sampling from tears may significantly reduce the amount of medication remaining in the eye. Owing to limitations caused by sampling difficulty, the concentration–time profile for each subject is generally unattainable when estimating the pharmacokinetics parameters for topical ophthalmic drugs. Instead, each subject can be sampled at one of several prespecified sampling times.

For the bioequivalence assessment of topical ophthalmic drugs, some of the U.S. Food and Drug Administration's product-specific guidances recommend single-dose, *in vivo*, sparse-sampling AH studies with pharmacokinetics (PK) endpoints using either a crossover or parallel study design. Ratios of the concentration-time curve up to the last measurable concentration ($AUC_{0\text{-}t}$), and C_{max} are required for the bioequivalence assessment. In sparse sampling design, a bioequivalence test for C_{max} can be conducted using Schuirmann's two one-sided tests[2] in the same manner as complete data designs. However, a bioequivalence test for the $AUC_{0\text{-}t}$, should be considered for sparse sampling situations.

In parallel study designs with sparse sampling, the AUCs of follow-on products and reference products are independent of each other. Several methods can be applied to perform bioequivalence assessment in parallel designs with sparse sampling. Meanwhile, in crossover designs, both sparse sampling AUC estimation and the correlation between treatments should be determined. Jaki et al.[3] developed three Fieller-type confidence intervals for a crossover design with a flexible sampling regime. Jaki's method is applicable to the case of sparse-sampling design and is the first technique to provide a direct estimate of confidence interval for the ratio of AUCs in a crossover design with a batch-sampling regime. Shen and Machado[4] developed a nonparametric bootstrap method for bioequivalence assessment. Bootstrap methods can be applied in both parallel and crossover designs.

The drawback of the bootstrap method is its relatively cumbersome study design, in which batches of simulations are required to estimate the sample size. However, the performance of the Fieller-type confidence interval compared with the bootstrap method is still unknown in crossover designs with sparse sampling.

In this study, we propose a Fieller-type confidence interval for the assessment of bioequivalence using sparse sampling data. The proposed method simplifies the estimation of Jaki's method. A simulation study is conducted to evaluate the customer risk (Type I error) and the empirical power of the nonparametric bootstrap method[4], Jaki's method[3], and our proposed method in sparse sampling crossover trials. Furthermore, we evaluate the performance of Jaki's method and our proposed method based on group sequential designs.

Consider a two-sequence, two-period, two-treatment crossover study. Subjects are randomized into two sequence groups: TR and RT (T = test product; R = reference product). Subjects in each sequence are then randomly assigned to Q sampling time points ($t_1, \cdots, t_Q$). The number of subjects at each time point is n_q. Each subject provides one sample in each period. Let y_{ijkq} denote the drug concentration of the ith subject at the qth time point ($q = 1, \cdots, Q$) of period j ($j = 1, 2$) in sequence k ($k = 1, 2$). Subsequently, using Bailer's algorithm, the AUC from 0 to the last time point of sequence k in jth period is approximated by:

doi: 10.3967/bes2019.099

*This work was supported by the subproject of National Major Scientific and Technological Special Project of China for 'Significant New Drugs Development' [2015ZX09501008-004].

1. College of Military Preventive Medicine, the Fourth Military Medical University, Xi'an 710032, Shaanxi, China; 2. Peking University Clinical Research Institute, Peking University Health Science Center, Beijing 100191, China

$$AUC_{jk} = \sum_{q=1}^{Q} c_q \mu_{jkq}, \quad (1)$$

where c_q is

$$\begin{aligned} c_1 &= \frac{1}{2}(t_2 - t_1) \text{ for } q = 1, \\ c_q &= \frac{1}{2}(t_{q+1} - t_{q-1}) \text{ for } q = 2, \cdots, Q-1, \\ c_Q &= \frac{1}{2}(t_Q - t_{Q-1}) \text{ for } q = Q. \end{aligned} \quad (2)$$

The AUC_{jk} can be estimated by:

$$\widehat{AUC}_{jk} = \sum_{q=1}^{Q} c_q \bar{y}_{jkq}, \quad (3)$$

with $\bar{y}_{jkq} = \frac{1}{n_q}\sum_{i=1}^{n_q} y_{ijkq}$. Subjects at different time points are independent; no covariance terms appear in the variance of $\widehat{AUC}_{jk}$, which is estimated by:

$$\hat{s}^2\left(\widehat{AUC}_{jk}\right) = \hat{\sigma}_{jk}^2 = \sum_{q=1}^{Q} \frac{c_q^2 \hat{s}_{jkq}^2}{n_q}, \quad (4)$$

with $\hat{s}_{jkq}^2 = \frac{1}{n_q - 1}\sum_{i=1}^{n_q}\left(y_{ijkq} - \bar{y}_{jkq}\right)$. For simplicity, we will use the notations as listed in Table 1 for the AUCs in each sequence and period.

To assess the bioequivalence between two products, we denote the AUCs of the test product and reference product by κ and λ, respectively, and define the ratio of the two AUCs as $\vartheta = \kappa/\lambda$. We assumed no carry-over effect, as the washout period is sufficiently long. Bioequivalence would be established if the 90% confidence interval of ϑ lies within 0.80, 1.25, which is the bioequivalence margin frequently recommended by regulatory authorities.

Based on the study of Locke[5], we derive the estimators of κ and λ, which include the individual drug effect and the mean of the fixed period effects, referred to as 'method 2,' as follows:

$$\widehat{AUC}^{T2} = \widehat{\kappa_2} = \frac{\hat{a} + \hat{d}}{2} \quad (5)$$

and

Table 1. Notations of AUCs in each sequence and period

Sequence	Period 1 (j = 1)	Period 2 (j = 2)
Sequence TR (k = 1)	$\hat{a}$	$\hat{b}$
Sequence RT (k = 2)	$\hat{c}$	$\hat{d}$

$$\widehat{AUC}^{R2} = \widehat{\lambda_2} = \frac{\hat{b} + \hat{c}}{2}. \quad (6)$$

The standard errors are then given by:

$$s^2\left(\widehat{AUC}^{T2}\right) = \hat{\xi}_{\kappa_2}^2 = \frac{1}{4}\left(\hat{\sigma}_a^2 + \hat{\sigma}_d^2\right) \quad (7)$$

and

$$s^2\left(\widehat{AUC}^{R2}\right) = \hat{\xi}_{\lambda_2}^2 = \frac{1}{4}\left(\hat{\sigma}_b^2 + \hat{\sigma}_c^2\right). \quad (8)$$

The covariance of $\widehat{\kappa_2}$ and $\widehat{\lambda_2}$ can be estimated as

$$\hat{\xi}_{\kappa_2,\lambda_2} = \frac{1}{4}\left(\hat{\sigma}_{a,b} + \hat{\sigma}_{c,d}\right). \quad (9)$$

The derivation of method 2 is given in File S1 (available in www.besjournal.com). Fieller's theorem[6] provides a general procedure for the estimation of confidence intervals for parameter ratios. Assuming that the data of each sequence and period are normally distributed, we can calculate:

$$\begin{aligned} \mathrm{A} &= \hat{\lambda}^2 - z_{\alpha/2}^2 \hat{\xi}_{\lambda}^2, \\ \mathrm{B} &= z_{\alpha/2}^2 \hat{\xi}_{\kappa,\lambda} - \hat{\lambda}\hat{\kappa}, \\ \mathrm{C} &= \hat{\kappa}^2 - z_{\alpha/2}^2 \hat{\xi}_{\kappa}^2, \end{aligned} \quad (10)$$

where $z_{\alpha/2}$ denotes the 100 (1-α) percentile of the t-distribution. Additionally, Fieller-type confidence intervals can be constructed using t-quantiles, and degrees of freedom can be calculated using the Satterthwaite approximation. Please refer to Jaki's research[3,7] for a detailed calculation process.

The lower limit ϑ_L and the upper limit ϑ_U of the Fieller-type confidence interval are

$$\vartheta_L = \left[-B - \left(B^2 - AC\right)^{1/2}\right]/A \quad (11)$$

and

$$\vartheta_U = \left[-B + \left(B^2 - AC\right)^{1/2}\right]/A, \quad (12)$$

respectively. As discussed by Fieller[5,6], obtaining an interpretable confidence interval requires that both $\hat{\lambda}^2/\hat{\xi}_{\lambda}^2 > t_{\alpha/2}^2$ and $\hat{\kappa}^2/\hat{\xi}_{\kappa}^2 > t_{\alpha/2}^2$ are satisfied. In other words, both κ and λ should be statistically significant

compared with 0 to construct a Fieller-type confidence interval that contains no negative values.

To evaluate the performance of the proposed approach, we simulated two-treatment, two-period, two-sequence crossover trials (TR/RT). We considered the sampling time and drug concentration based on the outcome of a pharmacokinetics study of an azithromycin eyedrop[8]. In each period, 210 subjects were randomly assigned to seven sampling time points (0.17, 0.5, 2, 4, 8, 12, and 24 h after dosing) nested in each sequence. Each subject was sampled at the same time point in each period. Thus, n_q = 30. The drug concentrations corresponding to each time point were 165, 50, 25, 10, 5, 1.5, and 0.5 μg/g of tears. The period effect was set as absent. Simulation data for each time point were obtained from multivariate log-normal distributions.

Two levels of within-subject variability (WSV) and between-subject variability (BSV) were considered in the simulation study. Intrasubject coefficient variances (CVs) were assumed to be 0.5 and 0.8 for low and high variabilities, respectively. The Intersubject CVs were assumed to be 1.2 and 1.5 for low and high variabilities, respectively. In these scenarios, the correlations between T and R in each sequence ranged from 0.64 to 0.84 for different combinations of WSV and BSV. For each scenario, 1,000 simulation trials were performed. All simulations were executed using the SAS 9.4 software. For simplicity, we designate our proposed approach 'method 1' and Jaki's approach 'method 2'.

Table 2 lists the Type I errors of the bioequivalence tests for each side of the equivalence margin as well as for different intrasubject and intersubject CV combinations. In all scenarios, the Type I errors of methods 1 and 2 were similar. The Type I errors of both methods 1 and 2 were lower than those of the bootstrap method and the nominal level of 5%. On both sides of the equivalence margin, the Type I errors of the bootstrap method were inflated.

The empirical power estimates for ϑ ranging from 0.95 to 1.05 are presented in Table 3. In scenarios 1 and 2, the empirical power values of methods 1 and 2 were similar to those of the bootstrap method. When the intrasubject CVs increased, as in scenarios 3 and 4, the empirical power values of methods 1 and 2 were slightly lower than those of the bootstrap method. It was clear that the empirical power of all methods was highly affected by the correlation between periods. As demonstrated by scenarios 2 and 3, when the overall variances were similar, with the correlation between periods ranging from 0.84 to 0.64, the empirical power decreased considerably. In group sequential designs, the empirical power values of methods 1 and 2 were similar in all scenarios.

In summary, the proposed method provides estimations for AUC ratios and the corresponding confidence intervals in accordance with the guidelines of regulatory authorities. The AUC estimators and associated standard errors were more straightforward in the proposed method. Our

Table 2. Type I errors of the bioequivalence tests on the ratio of AUCs using the (0.80, 1.25) equivalence margin

AUC_T/AUC_R	Intrasubject CV	Intersubject CV	Method 1	Method 2	Bootstrap	O'Brien-Fleming		Pocock	
						Method 1	Method 2	Method 1	Method 2
Scenario 1									
0.80	0.5	1.2	0.048	0.050	0.054	0.047	0.052	0.032	0.037
1.25	0.5	1.2	0.047	0.048	0.060	0.045	0.051	0.045	0.042
Scenario 2									
0.80	0.5	1.5	0.049	0.047	0.056	0.046	0.044	0.042	0.038
1.25	0.5	1.5	0.049	0.049	0.059	0.053	0.050	0.034	0.036
Scenario 3									
0.80	0.8	1.2	0.044	0.041	0.058	0.042	0.039	0.026	0.023
1.25	0.8	1.2	0.053	0.054	0.061	0.049	0.052	0.025	0.029
Scenario 4									
0.80	0.8	1.5	0.040	0.038	0.048	0.035	0.038	0.014	0.020
1.25	0.8	1.5	0.043	0.039	0.050	0.038	0.038	0.014	0.018

method is conservative compared with the bootstrap method, as its empirical power is slightly lower than that of the bootstrap method and no Type I error inflation is shown. Moreover, simulation analysis demonstrated that our technique is highly suitable for log-normal data. Thus, our method is applicable to either primary or sensitivity analyses in the bioequivalence assessment of topical ophthalmic drugs.

Both our method and Jaki's method first estimate AUCs and the corresponding standard errors for each sequence and period, and then construct the confidence intervals using Fieller's theorem. The difference between these two methods is attributed to the estimators of κ and λ, that is, the AUC estimators of the follow-on product and reference product, respectively. Theoretically, Jaki's estimators include the individual drug effect and fixed effect of the first period as a nuisance effect[3,9]. Our estimators include the individual drug effect and the mean of the fixed effects for all periods. If we assumed no period effects, the expectations of the two estimators would be identical. Our method is significantly simpler than Jaki's method, without a decrease in empirical power or an increase in Type I error. Neither of these two methods provides direct estimators of the individual drug effect, i.e., the 'pure' treatment effect, owing to the nuisance effect. This is because in a crossover design, individual drug effects cannot be estimated if the period effects are considered[10]. Fortunately, the 'pure' treatment effect does not need to be separated in bioequivalence assessment. For instance, in a parallel design, bioequivalence is assessed using the AUCs of each group, without considering the separation of the direct drug effect from the peculiarities of the period. Similarly, the estimators κ and λ are the parameters of interest when estimating the effects of a product.

In vivo bioequivalence studies for topical ophthalmic drugs are often conducted using a large sample size to achieve sufficient empirical power. Our investigation demonstrated that the empirical power levels of our method and Jaki's method in a group sequential design using O'Brien–Fleming boundaries are both similar to that of the fixed sample size design. Thus, group sequential designs can be utilized to reduce sample size by permitting early stopping.

Well-developed study designs will considerably accelerate the development of drugs. Fieller-type confidence interval approaches (our proposed method and Jaki's method) provide greater flexibility

Table 3. Empirical power values of the bioequivalence tests on the ratio of AUCs using the (0.8, 1.25) equivalence margin

AUC_T/AUC_R	Intrasubject CV	Intersubject CV	Method 1	Method 2	Bootstrap	O'Brien-Fleming		Pocock	
						Method 1	Method 2	Method 1	Method 2
Scenario 1									
0.95	0.5	1.2	0.888	0.890	0.890	0.889	0.890	0.861	0.851
1.00	0.5	1.2	0.946	0.942	0.945	0.946	0.942	0.917	0.904
1.05	0.5	1.2	0.869	0.862	0.871	0.873	0.876	0.843	0.832
Scenario 2									
0.95	0.5	1.5	0.830	0.819	0.838	0.832	0.823	0.784	0.775
1.00	0.5	1.5	0.892	0.891	0.892	0.894	0.890	0.857	0.862
1.05	0.5	1.5	0.820	0.809	0.819	0.820	0.812	0.770	0.754
Scenario 3									
0.95	0.8	1.2	0.491	0.491	0.513	0.473	0.473	0.341	0.352
1.00	0.8	1.2	0.564	0.552	0.585	0.548	0.540	0.406	0.402
1.05	0.8	1.2	0.523	0.505	0.544	0.497	0.500	0.377	0.370
Scenario 4									
0.95	0.8	1.5	0.391	0.380	0.418	0.365	0.369	0.256	0.249
1.00	0.8	1.5	0.459	0.439	0.485	0.433	0.421	0.291	0.273
1.05	0.8	1.5	0.415	0.402	0.449	0.383	0.384	0.258	0.260

in the study designs of sparse sampling crossovers. Based on parametric methods such as the Fieller-type confidence interval, other methodologies, including scaled average bioequivalence and sample size re-estimation, can be considered. Furthermore, Fieller-type confidence interval approaches can be applied to bioequivalence studies with pharmacodynamic endpoints.

Acknowledgments We thank LetPub (www.letpub.com) for the linguistic assistance during the preparation of this manuscript.

Author contributions YY performed the derivation, analyzed the data, and drafted the manuscript. JX conceived the research questions and revised the manuscript. YY, XY, and CY designed and performed the simulation study. XY and CY critically commented and revised the manuscript. All authors read and approved the final version of the manuscript.

Competing Interests The authors declare no competing interests.

[#]Correspondence should be addressed to XIA Jie Lai, Tel: 86-13571999716, E-mail: xiajielai@fmmu.edu.cn

Biographical note of the first author: YU Yong Pei, male, born in 1984, MS, majoring in biostatistics.

Received: April 5, 2019;

Accepted: September 16, 2019

REFERENCES

1. Deng F, Ranta VP, Kidron H, et al. General Pharmacokinetic Model for Topically Administered Ocular Drug Dosage Forms. Pharmaceutical Research, 2016; 33, 2680−90.
2. Chow SC, Liu Jp. Design and analysis of bioavailability and bioequivalence studies. Boca Raton: CRC Press, 2009.
3. Jaki T, Pallmann P, Wolfsegger MJ. Estimation in AB/BA crossover trials with application to bioequivalence studies with incomplete and complete data designs. Stat Med, 2013; 32, 5469−83.
4. Shen M, Machado SG. Bioequivalence evaluation of sparse sampling pharmacokinetics data using bootstrap resampling method. J Biopharm Stat, 2017; 27, 257−64.
5. Locke CS. An exact confidence interval from untransformed data for the ratio of two formulation means. J Pharmacokinet Biopharm, 1984; 12, 649−55.
6. Fieller EC. Some Problems in Interval Estimation. Journal of the Royal Statistical Society. Series B (Methodological), 1954; 14, 175−85.
7. Jaki T, Wolfsegger MJ, Lawo J-P. Establishing Bioequivalence in Complete and Incomplete Data Designs Using AUCs. J Biopharmaceutical Stat, 2010; 20, 803−20.
8. Chiambaretta F, Garraffo R, Elena PP, et al. Tear concentrations of azithromycin following topical administration of a single dose of azithromycin 0.5%, 1.0%, and 1.5% eyedrops (T1225) in healthy volunteers. Eur J Ophthalmol, 2008; 18, 13−20.
9. Yan Z. Comments on 'Estimation in AB/BA crossover trials with application to bioequivalence studies with incomplete and complete data designs'. Stat Med, 2013; 32, 5484−6.
10. Jones B, Kenward MG. Design and analysis of cross-over trials. Boca Raton: CRC Press/Taylor & Francis, 2014.

File S1. The derivation of the proposed approach

The general model of crossover trials with sparse sampling data is given as:

$$AUC_{jk}=\mu+P_j + F_{(j,k)} + e_{jk} \tag{1}$$

where

μ is the overall mean;

P_j is the fixed effect of jth period, where j = 1, 2;

$F_{(j,k)}$ is the fixed effect of the formulation in the kth sequence at jth period, where k = 1, 2. Thus,

$$F_{(j,k)} = \begin{cases} F_T & \text{if } k = j \\ F_R & \text{if } k \neq j \end{cases} \quad k = 1,2;\ j = 1,2\ . \tag{2}$$

e_{jk} is the random error of AUC_{jk}.

Note that the AUCs of sequence TR at period 1 and period 2, $(AUC_{11}, AUC_{21})'$, follow a bivariate normal distribution with mean vector ϕ_1 and covariance Σ_1, where

$$\phi_1 = \begin{bmatrix} \mu + F_T + P_1 \\ \mu + F_R + P_2 \end{bmatrix} \text{ and } \Sigma_1 = \begin{bmatrix} \sigma_T^2 + \sigma_S^2 & \sigma_S^2 \\ \sigma_S^2 & \sigma_R^2 + \sigma_S^2 \end{bmatrix}. \tag{3}$$

Where σ_T^2 and σ_R^2 intra-subject variance of F_T and F_R, respectively, and σ_S^2 is the inter-subject variance.

Similarly, the mean vector ϕ_2 and covariance vector Σ_2 of the AUCs of sequence RT, $(AUC_{12}, AUC_{22})'$, are given by

$$\phi_2 = \begin{bmatrix} \mu + F_R + P_1 \\ \mu + F_T + P_2 \end{bmatrix} \text{ and } \Sigma_2 = \begin{bmatrix} \sigma_R^2 + \sigma_S^2 & \sigma_S^2 \\ \sigma_S^2 & \sigma_T^2 + \sigma_S^2 \end{bmatrix}. \tag{4}$$

Recall the notations of AUCs in each sequence and period in Table I, let the ratio of AUCs $\vartheta = \kappa/\lambda = (\mu + F_T) / (\mu + F_R)$, define

$$U_k = \begin{cases} (a - \vartheta b)/2, & k = 1 \\ (d - \vartheta c)/2, & k = 2 \end{cases}\ . \tag{5}$$

Then U_1 follows a normal distribution with mean $(P_1 - \vartheta P_2)/2$ and variance $\sigma_\vartheta^2/4$, where and U_2 follows a normal distribution with mean $(P_2 - \vartheta P_1)/2$ and variance $\sigma_\vartheta^2/4$. Since samples from sequence TR(k = 1) and RT(k = 2) are independent, parameters U_1 and U_2 are normally distributed with equal variance. Then confidence interval for ϑ can be obtained based on an unpaired two-sample statistic, that is:

$$\sigma_\vartheta^2 = \left(\sigma_T^2 + \sigma_S^2\right) - 2\vartheta\sigma_S^2 + \vartheta^2\left(\sigma_R^2 + \sigma_S^2\right). \tag{6}$$

$$T = \frac{\hat{U}_1 + \hat{U}_2}{S_U} \tag{7}$$

where

$$\hat{U}_1 + \hat{U}_2 = \frac{1}{2}\left(\hat{a} - \vartheta\hat{b}\right) + \frac{1}{2}\left(\hat{d} - \vartheta\hat{c}\right) = \frac{1}{2}\left(\hat{a} + \hat{d}\right) - \frac{\vartheta}{2}\left(\hat{b} + \hat{c}\right) \tag{8}$$

and

$$S_U^2 = \frac{1}{4}\left(\hat{\sigma}_a^2 + \vartheta^2\hat{\sigma}_b^2 - 2\vartheta\hat{\sigma}_{a,b}\right) + \frac{1}{4}\left(\hat{\sigma}_d^2 + \vartheta^2\hat{\sigma}_c^2 - 2\vartheta\hat{\sigma}_{c,d}\right) = \frac{1}{4}\left(\hat{\sigma}_a^2 + \hat{\sigma}_d^2\right) + \frac{\vartheta^2}{4}\left(\hat{\sigma}_b^2 + \hat{\sigma}_c^2\right) - \frac{\vartheta}{2}\left(\hat{\sigma}_{a,b} + \hat{\sigma}_{c,d}\right). \tag{9}$$

Then, the T statistics can be reformulated as

$$T = \frac{\hat{\kappa} - \vartheta\hat{\lambda}}{\sqrt{\hat{\xi}_\kappa^2 + \vartheta^2\hat{\xi}_\lambda^2 - 2\vartheta\hat{\xi}_{\kappa,\lambda}}} \tag{10}$$

where $\hat{\kappa}$ and $\hat{\lambda}$ are the point estimate of AUC_T and AUC_R, respectively.

$$\widehat{AUC}_T = \hat{\kappa} = \frac{\hat{a} + \hat{d}}{2}$$
$$\widehat{AUC}_R = \hat{\lambda} = \frac{\hat{b} + \hat{c}}{2} \tag{11}$$

The variances of $\hat{\kappa}$ and $\hat{\lambda}$ are given by:

$$\begin{aligned} s^2\left(\widehat{AUC}_T\right) &= \hat{\xi}_\kappa^2 = \frac{1}{4}\left(\hat{\sigma}_a^2 + \hat{\sigma}_d^2\right) \\ s^2\left(\widehat{AUC}_R\right) &= \hat{\xi}_\lambda^2 = \frac{1}{4}\left(\hat{\sigma}_b^2 + \hat{\sigma}_c^2\right) \end{aligned} \tag{12}$$

The covariance of $\hat{\kappa}$ and $\hat{\lambda}$ can be estimated as

$$\hat{\xi}_{\kappa_2,\lambda_2} = \frac{1}{4}\left(\hat{\sigma}_{a,b} + \hat{\sigma}_{c,d}\right). \tag{13}$$