Available online: June 28, 2018

Commun. Fac. Sci. Univ. Ank. Ser. A1 Math. Stat.
Volume 68, Number 1, Pages 583–594 (2019)
DOI: 10.31801/cfsuasmas.438407
ISSN 1303–5991 E-ISSN 2618-6470
http://communications.science.ankara.edu.tr/index.php?series=A1

ITERATED BOOTSTRAP PROCEDURE IN INDIVIDUAL BIOEQUIVALENCE

UFUK BEYAZTAS

Abstract. In this paper, we propose an iterating principle in the bootstrap method to assess the individual bioequivalence under 2×4 randomized crossover design. The finite sample properties of the proposed algorithm are illustrated by an extensive simulation study and a real-world example. Our findings reveal that the proposed idea have better performance than the classical percentile bootstrap confidence limits.

1. Introduction

Bioequivalence (BE) studies play an important role in the drug development process. The goal of such studies is to evaluate the therapeutic equivalence of two (or more) drugs or to study if two different galenic formulations of the same drug have a similar bioavailability and therapeutic effect. Let T be a generic drug developed as an alternative to a reference drug (R). The United States Food and Drug Administration (FDA) requires BE before marketing formulation T or new formulations of the existing drugs. A formal definition of the BE given by [4] is as follows: "Bioequivalence is defined as the absence of a significant difference in the rate and extent to which the active ingredient or active moiety in pharmaceutical equivalents or pharmaceutical alternatives becomes available at the site of drug action when administered at the same molar dose under similar conditions in an appropriately designed study." Usually, the BE studies are carried out by measuring the drug concentration in the blood by several pharmacokinetic variables. The commonly used pharmacokinetic variables are the area under the plasma concentration curve (AUC), the maximum drug concentration (C_{max}) and the time required to reach the maximum drug concentration (T_{max}). Note that the BE studies are performed with healthy volunteer subjects.

There are three types of BE: (i) The average bioequivalence (ABE), which compares the distance of average pharmacokinetic measures between the formulations

Received by the editors: March 27, 2017; Accepted: March 07, 2018.
2010 *Mathematics Subject Classification.* Primary 62F40, 62P10; Secondary 62F03.
Key words and phrases. Bioequivalence, bootstrap, hypothesis testing.

T and R; (*ii*) The population bioequivalence (PE), which compares the population means under the test and reference formulation as well as the between-subject variance in bioavailability; (*iii*) The individual bioequivalence (IBE), which additionally takes into account the subject-by-formulation interaction for the test and reference formulation. This paper aims to contribute to the IBE studies by resampling procedures. For more information about the types of BE studies please see [7].

Generally, a two sequence-four period (2×4) randomized crossover design is recommended to assess the IBE. A crossover design is a repeated measurements design so that the experimental units cross over from one treatment to another during the different time periods. Such design has formed the basis of the many clinical studies, see [15] and [8] for more information about the crossover designs in medical studies. Note that, as pointed out by referee, this design can be used properly in drug studies when the drugs under consideration are used only for symptomatic purposes not for the treatment. For analyzing the IBE, aggregate, scaled and moment based measures, which are nonlinear functions of the difference of the means (T - R) and the various variance components, are proposed by [12], [13] and [14]. [3] proposes using the upper limit of a 90% confidence interval to test the hypothesis of IBE. However, it is difficult to determine the exact distribution of the estimators calculated for the measures mentioned above since they have a nonlinear form. To overcome this problem [12] suggests to use the original nonparametric bootstrap method to build up a confidence interval for the statistics since it does not require the full knowledge of the underlying data and distributional assumptions. [16] proposes an improved procedure to assess the IBE and [10] reviews the different concepts of IBE by concentrating on the bootstrap percentile interval. See also [11] and references therein for the history of IBE and the role of the bootstrap in this context.

The theoretical properties such as consistency and accuracy of the recommended bootstrap percentile intervals to test the IBE hypothesis are discussed by [16]. The authors conclude that the FDA's bootstrap procedure using the moment-based approach yields a consistent test procedure. On the other hand, the iterated bootstrap method (see, [5]) can be useful in obtaining an arbitrarily high degree of correction and improving the efficiency of bootstrap by iterating the bootstrap argument. [5], [1], [2] and [6] provide theoretical properties of this method and prove that the iterating principle reduces the bootstrap errors in many statistical problems. This paper proposes the use of iterated bootstrap algorithm with an aim to reduce the coverage error of the percentile confidence interval in individual bioequivalence studies.

The rest of the paper is organized as follows. In Section 2 we provide a detailed information on the IBE and bootstrap methods examined in this study. An extensive Monte Carlo simulation is conducted to examine the finite sample performance of the iterated bootstrap method and the results are presented in Section 3. Section

4 presents the numerical results for the real-world example considered in this study. Finally, we conclude with some final remarks described in Section 5.

2. Methodology

A two sequence-four period randomized crossover design $TRTR$ and $RTRT$ is commonly recommended for assessing IBE. Under this design, four drug treatments in the order of $TRTR$ are administrated to the first sequence of n_1 patients whereas in the second sequence, the treatments are administrated in the order of $RTRT$ to n_2 patients. Let Y_{ijk} represents the log transformed response (i.e., $Y = \log AUC$ or $\log C_{max}$) for subject i in the jth period of sequence k, where $i = 1, \cdots, n_k$, $j = 1, 2, 3, 4$ and $k = 1, 2$. The following mixed-effect model is recommended:

$$y_{ijk} = \mu + F_l + P_j + Q_k + W_{ljk} + S_{ikl} + \epsilon_{ijk} \tag{2.1}$$

where μ is the overall mean, P_j is the fixed effect of the jth period ($j = 1, 2, 3, 4$ and $P_1 + P_2 + P_3 + P_4 = 0$), Q_k is the fixed effect of the kth sequence ($k = 1, 2$ and $Q_1 + Q_2 = 0$), F_l is the fixed effect of the lth drug formulation ($l = T$ when (k,j) = (1,1), (1,3), (2,2), (2,4) and $l = R$ otherwise, $F_t + F_r = 0$), W_{ljk} is the fixed effect of interaction (sum of W_{ljk}'s over any index is 0), S_{ikl} is the random effect of the ith subject in the kth sequence under drug formulation l and (S_{ikT}, S_{ikR}), $i = 1, \cdots, n_k$, $k = 1, 2$ are independent and identically distributed random vectors with mean 0 and unknown covariance matrix

$$\begin{pmatrix} \sigma^2_{BT} & \rho\sigma_{BT}\sigma_{BR} \\ \rho\sigma_{BT}\sigma_{BR} & \sigma^2_{BR} \end{pmatrix} \tag{2.2}$$

ϵ_{ijk}'s are independent random errors with mean 0 and variance σ^2_{Wl}, and (S_{ikT}, S_{ikR})'s and ϵ_{ijk}'s are independent (for more information please see [16]). Note that: (i) σ^2_{BT} and σ^2_{BR} are the between subject variances while σ^2_{WT} and σ^2_{WR} are the within subject variances for the test T and reference R formulations, respectively. (ii) The correlation between the test and reference formulations responses from subject i, ρ, is related to the subject-by-formulation interaction variance $\sigma^2_D = var(S_{ikT} - S_{ikR}) = \sigma^2_{BT} + \sigma^2_{BR} - 2\rho\sigma_{BT}\sigma_{BR}$. Table 1 shows the expected means and observed data of the crossover design considered in this study.

Table 1. Expected means

Sequence	Period 1	Period 2	Period 3	Period 4
1 ($TRTR$)	$\mu + F_T + P_1 + Q_1 + W_{T11} + S_{i1T} + \epsilon_{i11}$ $i = 1, \cdots, n_1$	$\mu + F_R + P_2 + Q_1 + W_{R21} + S_{i1R} + \epsilon_{i21}$ $i = 1, \cdots, n_1$	$\mu + F_T + P_3 + Q_1 + W_{T31} + S_{i1T} + \epsilon_{i31}$ $i = 1, \cdots, n_1$	$\mu + F_R + P_4 + Q_1 + W_{R41} + S_{i1R} + \epsilon_{i41}$ $i = 1, \cdots, n_1$
2 ($RTRT$)	$\mu + F_R + P_1 + Q_2 + W_{R12} + S_{i2T} + \epsilon_{i12}$ $i = 1, \cdots, n_2$	$\mu + F_T + P_2 + Q_2 + W_{T22} + S_{i2R} + \epsilon_{i22}$ $i = 1, \cdots, n_2$	$\mu + F_R + P_3 + Q_2 + W_{R32} + S_{i2T} + \epsilon_{i32}$ $i = 1, \cdots, n_2$	$\mu + F_T + P_4 + Q_2 + W_{T42} + S_{i2R} + \epsilon_{i42}$ $i = 1, \cdots, n_2$

Let Y_T, Y_R and Y'_R, respectively, represent the bioavailabilities of the administrated one test and two reference formulations. Then the two formulations are

considered individually bioequivalent if the null hypothesis given in Equation 2.3 is rejected at the α percent significance level.

$$\begin{aligned} H_0 &: \theta = \frac{E(Y_T - Y_R)^2 - E(Y_R - Y_R')^2}{\max\{\sigma_0^2, \sigma^2\}} < \theta_U \\ H_1 &: \theta < \theta_U \end{aligned} \tag{2.3}$$

where $\sigma^2 = E(Y_R - Y_R')^2$ is the variance calculated under the reference formulation, θ_U is the predetermined upper limit for IBE, and σ_0^2 is the within subject variance for the reference formulation. Note that for all the numerical analyses considered in this study the values of θ_U and σ_0^2 are chosen as $\theta_U = 2.4948$ and $\sigma_0^2 = 0.04$ as proposed by the FDA.

Under model 2.1, the parameter of interest, θ is obtained as in Equation 2.4 given below

$$\theta = \frac{(F_T - F_R)^2 + \sigma_{BT}^2 + \sigma_{BR}^2 - 2\rho\sigma_{BT}\sigma_{BR} + \sigma_{WT}^2 - \sigma_{WR}^2}{\max(\sigma_0^2, \sigma_{WR}^2)} \tag{2.4}$$

where $\sigma_{WR}^2 = E(Y_R - Y_R')^2/2$. The model described in Equation 2.1 is fitted by using general linear models, maximum likelihood or restricted maximum likelihood (REML) procedures. FDA recommends the use of REML method to estimate the mean difference and variance components. REML uses an iterative process where each iteration has two steps. In the first step it uses the initial parameter estimates to estimate the fixed effects. Then, in the second step, the variance parameters are re-estimated by using the residuals obtained from the first step. These two steps are repeated when the parameter estimates do not change from one iteration to the next. It has two main properties; it is useful to estimate between and within subject variances, and it may be useful to estimate the mean differences and variance components when the data set is incomplete. Also, [17] shows that the bootstrap procedure using REML yields a consistent test procedure. On the other hand, as described by [16], it may not produce the best estimator of θ and may not be robust against the violation of the normality assumption. Also, it requires a large amount of computation time since REML estimators involves an iteration process. The moment method which is simple and robust against the normality assumption can be used to estimate the mean difference and variance components as an alternative to the method of REML. Following [16], let $\tau = \sigma_{BT}^2 + \sigma_{BR}^2 - 2\rho\sigma_{BT}\sigma_{BR} + \sigma_{WT}^2 + \sigma_{WR}^2$,

$$\begin{aligned} z_{i11} &= y_{i11} - y_{i21} \\ z_{i21} &= y_{i31} - y_{i41} \\ z_{i31} &= y_{i21} - y_{i41} \; i = 1, \cdots, n_1 \\ z_{i12} &= y_{i12} - y_{i22} \\ z_{i22} &= y_{i32} - y_{i42} \\ z_{i32} &= y_{i12} - y_{i32} \; i = 1, \cdots, n_2 \end{aligned}$$

and let $\bar{z}_{jk}$ and s^2_{jk} be the sample mean and sample variance based on $z_{1jk}, \cdots, z_{n_kjk}$, for each fixed (j,k). Then, the unbiased estimators of $F_T - F_R$, τ and σ^2_{WR} are obtained as $(\bar{z}_{11} + \bar{z}_{21} - \bar{z}_{12} - \bar{z}_{22})/4$, $(s^2_{11} + s^2_{21} + s^2_{12} + s^2_{22})/4$ and $(s^2_{31} + s^2_{32})/4$, respectively. By using these estimators, which are first and second-order moments, a moment estimator of θ is estimated as in Equation 2.5 given below

$$\begin{aligned} \hat{\theta}_{mom} &= \frac{(\hat{F}_T - \hat{F}_R)^2 + \tau - 2\hat{\sigma}^2_{WR}}{\max\{\sigma^2_0, \hat{\sigma}^2_{WR}\}} \\ &= \frac{[(\bar{z}_{11} + \bar{z}_{21} - \bar{z}_{12} - \bar{z}_{22})/4]^2 + (s^2_{11} + s^2_{21} + s^2_{12} + s^2_{22})/4 - (s^2_{31} + s^2_{32})/2}{\max\{\sigma^2_0, (s^2_{31} + s^2_{32})/4\}} \end{aligned} \tag{2.5}$$

For this study, we restricted our focus to only moment estimator for the reason of its advantageous. Following is the bootstrap algorithm proposed by FDA to set the confidence intervals when testing IBE hypothesis.

Step 1. Estimate the IBE parameter θ

$$\hat{\theta} = \frac{(\hat{F}_T - \hat{F}_R)^2 + \hat{\tau} - 2\hat{\sigma}^2_{WR}}{\max\{\sigma^2_0, \hat{\sigma}^2_{WR}\}}$$

Step 2. Let $Y_{i1} = (y_{iT11}, y_{iR11}, y_{iT12}, y_{iR12})$ and $Y_{i2} = (y_{iR21}, y_{iT21}, y_{iR22}, y_{iT22})$ denote the four observed pharmacokinetic variables in sequence 1 and 2, respectively. Let also $Y_{1,2} = \{(Y_1, \cdots, Y_{n_1}), (Y_1, \cdots, Y_{n_2})\}$ represents the n_k vectors of Y_{i1} and Y_{i2} for $k = 1, 2$. Generate n_k bootstrap samples $Y^*_{1,2} = \{(Y^*_{11}, \cdots, Y^*_{n_11}), (Y^*_{12}, \cdots, Y^*_{n_22})\}$ with replacement from $Y_{1,2}$. Note that, the bootstrap resampling is stratified by sequence.

Step 3. Calculate the bootstrap estimate of $\hat{\theta}$, $\hat{\theta}^*$, by using the bootstrap data set $Y^*_{1,2}$.

$$\hat{\theta}^* = \begin{cases} \frac{(\hat{F}^*_T - \hat{F}^*_R)^2 + \hat{\tau}^* - 2\hat{\sigma}^{*2}_{WR}}{\hat{\sigma}^{*2}_{WR}} & if\ \hat{\sigma}^2_{WR} \geq \sigma^2_0 \\ \frac{(\hat{F}^*_T - \hat{F}^*_R)^2 + \hat{\tau}^* - 2\hat{\sigma}^{*2}_{WR}}{\sigma^2_0} & if\ \hat{\sigma}^2_{WR} < \sigma^2_0 \end{cases} \tag{2.6}$$

Step 4. Repeat steps 2 and 3 B times.

Step 5. Determine the 95th percentile ($\hat{\theta}_{FDA}(95)$) of this generated bootstrap distribution.

Step 6. $\hat{\theta}_{FDA}(95)$ is then compared to θ_U to conclude that two formulations are bioequivalent or not.

[16] studies the properties of this test procedure and they conclude that the bootstrap procedure in the FDA's draft guidance is incorrect. The authors suggest to use the bootstrap estimate $\hat{\theta}^*$ as in Equation 2.7 given below.

$$\frac{(\hat{F}_T^* - \hat{F}_R^*)^2 + \hat{\tau}^* - 2\hat{\sigma}_{WR}^{*2}}{\max\{\sigma_0^2, \hat{\sigma}_{WR}^{*2}\}} \tag{2.7}$$

The 95th percentile calculated by using this bootstrap estimate is called the bootstrap percentile (BP) upper confidence bound for θ, $\hat{\theta}_{BP}(95)$.

The use of iterated bootstrap procedure mentioned in Section 1 may provide better upper confidence bounds since it improves the coverage accuracy of bootstrap percentile confidence intervals. Let $X_1, X_2, \cdots$ be a sequence of i.i.d. random variables from an unknown distribution $F \equiv F_\theta$, where the parameter θ is of our primary interest. Let $\chi_n = (X_1, \cdots, X_n)$ be an i.i.d. random sample from F, and let $R_n(\chi_n, \theta)$ be the pivotal quantity whose distribution is given by $G_n = G_n(\cdot, F)$. Suppose Θ is the set of all possible values of θ. Then, a level α confidence set for the parameter θ can be obtained as

$$S_n = \{t \in \Theta : R_n(\chi_n, t) \leq G_n^{-1}(\alpha)\} \tag{2.8}$$

for any given $\alpha \in (0, 1)$, where $G_n^{-1}(\alpha)$ describes the largest α-th quantile of G_n. For any sequence $\{F_n\}$ which converges to F, $G_n(\cdot, F_n)$ is supposed to converge weakly to a continuous distribution function $G = G(\cdot, F)$. Then, $G_n(R_n(\chi_n, \theta))$ is distributed as uniform $U(0, 1)$. In classical theory, G_n is approximated by its limit. However, in most cases, it is not easy to obtain its limit when the estimate of the parameter is a complicated statistic. But bootstrap method makes it possible since it does not require the full knowledge of the underlying distribution. Let $\chi_n^* = (X_1^*, \cdots, X_n^*)$ be the bootstrap sample from F_n, where F_n is the empirical distribution function which puts mass $1/n$ to each data point. Let also $\widehat{\theta}$ be the estimate of θ based on χ_n. Then, the bootstrap analogue of $R_n(\chi_n, \theta)$ with the bootstrap distribution conditional on χ_n are given as $R_n^* = R_n(\chi_n^*, \widehat{\theta})$ and $G_n^* = G_n^*(\cdot, F_n)$, respectively. Similar to the Equation 2.8, the bootstrap estimate of S_n is obtained as

$$S_n^* = \{t \in \Theta : R_n(\chi_n, t) \leq G_n^{*-1}(\alpha)\}$$

Since F_n is a consistent estimate for F, the bootstrap estimate G_n^* converges in probability to G as n increases. Moreover, $G_n^*(R_n(\chi_n^*, \widehat{\theta}))$ converges to a uniform $U(0, 1)$ distribution.

For the finite samples, the level of the confidence set given above tends to be inaccurate. One way of improving it is to use the iterating principle which is based on [1]'s prepivoting idea such that it transforms the original root $R_n(\chi_n, \theta)$ into a new root $R_{n,1}(\chi_n, \theta) = G_n^*\{R_n(\chi_n, \theta)\}$ whose distribution is less dependent to F compared to R_n. In other words, mapping R_n into $R_{n,1}$ is called prepivoting. Note that $R_{n,1}(\chi_n, \theta)$ is exactly distributed $U(0, 1)$ if $R_n(\chi_n, \theta)$ is the pivot. Let

$G_{n,1}(x) = \mathbb{P}(R_{n,1}(\chi_n, \theta) \leq x)$ be the distribution of $R_{n,1}(\chi_n, \theta)$, and let $G^*_{n,1}(x) = \mathbb{P}(R_{n,1}(\chi^*_n, \widehat{\theta}) \leq x | \chi_n)$ be its bootstrap estimate. Then

$$S^*_{n,1} \quad = \quad \{t \in \Theta : R_{n,1}(\chi_n, \theta) \leq G^{*-1}_{n,1}(\alpha)\} = \{\theta : R_n(\chi_n, \theta) \leq G^{*-1}_n[G^{*-1}_{n,1}(\alpha)]\}$$

defines α level iterated bootstrap confidence set for θ. Generally, the error in $S^*_{n,1}$ is smaller than the error in S^*_n and S_n. The iteration can be repeated continuously to reduce the coverage error of a confidence interval to a desired level. On the other hand, each iteration increases the computation burden drastically. [9] shows that each iteration reduces the coverage error by an order of $n^{-1/2}$ and n^{-1} for one-sided and two-sided intervals, respectively. By considering the computational burden of this iterative procedure, to make our proposed method more practical and widely applicable we only recommend of doing the double bootstrap where iteration is only being done once.

The iterated bootstrap algorithm used in this study works as follows: First, drawn a simple random sample of size n_k with replacement from $Y_{1,2}$. For the second level bootstrap, drawn another simple random sample of size n_k with replacement from $Y^*_{1,2}$. Let $Y^{**}_{1,2}$ be the generic bootstrap resample from $Y^*_{1,2}$. Also let $\hat{\theta}^*$ and $\hat{\theta}^{**}$ be the bootstrap estimators of $\hat{\theta}$, and $R^*_n = R(Y^*_{1,2}, \hat{\theta})$ and $R^{**}_n = R(Y^{**}_{1,2}, \hat{\theta}^*)$ be the bootstrap pivotal quantities obtained from $Y^*_{1,2}$ and $Y^{**}_{1,2}$, respectively with B_1 and B_2 being the number of first and second level bootstrap replications. Then the Monte Carlo algorithm for the iterated bootstrap for the construction of confidence sets is as follows.

(a) Resample $Y^*_{1,2}$ from $Y_{1,2}$ as explained above, and compute $R^*_{n,i}$ for $i = 1, \cdots B_1$.

(b) For each i, resample a second level bootstrap sample $Y^{**}_{1,2}$ from $Y^*_{1,2}$ and compute $R^{**}_{n,j}$ for $j = 1, \cdots, B_2$.

(c) Calculate $Z_i = \frac{1}{B_2} \sum_{j=1}^{B_2} I(R^{**}_{n,j} \leq R^*_{n,i})$ for $i = 1, \cdots, B_1$.

(d) Then, the empirical cumulative distribution of Z_i's, G^{**}_n is asymptotically $U(0,1)$, and approximate the distribution of R^*_n, $G^*_n = \mathbb{P}(R(Y^*_{1,2}, \widehat{\theta}))$.

(e) Define the level α confidence set of iterated bootstrap for R_n as $S^{**}_n = \{\theta : R_n(Y_{1,2}, \theta) \leq G^{*-1}_n[G^{**-1}_n(\alpha)]\}$.

Then the iterated versions of the bootstrap upper confidence limits, $\hat{\theta}^*_{FDA}(95)$ and $\hat{\theta}^*_{BP}(95)$, for assessing IBE hypothesis can easily be obtained by using the algorithm given above.

3. Numerical Results

To investigate the performances of the iterated bootstrap confidence limits we carried out a simulation study under different parameter settings and sample sizes, and we compared our results with classical percentile bootstrap confidence limits by means of estimated test size (α) and power of the test (β). The parameter settings (where H_0 in Equation 2.3 hold) presented in Table 2 were considered to calculate the α values, and the parameter settings (where H_1 in 2.3 is actually true) in Table 3 were considered to calculate the β values. 2×4 randomized crossover design was considered to assess the IBE. For each simulated experiment, four drug treatments were arranged in the order of $TRTR$ in sequence 1 whereas the treatment order was arranged as $RTRT$ in sequence 2. For each parameter setting, 100 experiments were simulated, and for each case, a sample of size n was generated, and $B_1 = B_2 = 2000$ bootstrap resamples were generated in each resampling operation. The boundary of the null hypothesis of individual bioequivalence (θ_U) and the within subject variances for the reference formulation (σ^2_{WR}) where chosen as $\theta_U = 2.4948$ and $\sigma^2_{WR} = 0.04$ as is proposed by the FDA. All simulations were done at the nominal $\alpha = 0.05$. The calculations were carried out using R 3.3.2. on and Intel Core i7 6700HQ 2.6 GHz PC. (The codes can be obtained from the author upon request.) The results are presented in Table 4.

Our findings show that, for all sample sizes and parameter settings under considered, iterated bootstrap methods outperform their conventional counterparts in terms of power values. The estimated test sizes of the conventional bootstrap methods ($\hat{\theta}_{FDA}(95)$ and $\hat{\theta}_{BP}(95)$) are smaller than the nominal size $\alpha = 0.05$, in general. For the iterated bootstrap methods ($\hat{\theta}^*_{FDA}(95)$ and $\hat{\theta}^*_{BP}(95)$), the test sizes are larger than the nominal size under first parameter setting. On the other hand, for the other parameter settings, they tend to have a reasonable test sizes, which are close to the nominal size $\alpha = 0.05$, but still larger than the ones obtained by the conventional methods. It is not a surprising result since a large test size corresponds to a large power of the test.

Table 2. Parameter settings under the null hypothesis for type I error simulation

Parameter setting	$F_T - F_R$	σ^2_{WR}	σ^2_{WT}	σ^2_{BR}	σ^2_{BT}	ρ	θ
1	0.3	0.01	0.04	0.01	0.04	0.9	3.350
2	0.3	0.01	0.06	0.01	0.04	0.9	3.850
3	0.4	0.03	0.04	0.01	0.07	0.9	5.059
4	0.4	0.01	0.02	0.02	0.03	0.9	4.397

TABLE 3. Parameter settings under the alternative hypothesis for power simulation

Parameter setting	$F_T - F_R$	σ^2_{WR}	σ^2_{WT}	σ^2_{BR}	σ^2_{BT}	ρ	θ
1	0.1	0.02	0.06	0.02	0.03	0.9	1.397
2	0.3	0.02	0.06	0.02	0.05	0.9	1.576
3	0.1	0.01	0.03	0.01	0.03	0.9	1.720
4	0.2	0.02	0.05	0.02	0.03	0.9	1.897

TABLE 4. Simulation results: Estimated test sizes and power of the tests

Parameter setting	Method	Sample size							
		16		24		32		48	
		α	β	α	β	α	β	α	β
1	$\hat{\theta}_{FDA}(95)$	0.06	0.64	0.01	0.75	0.02	0.85	0.02	0.99
	$\hat{\theta}_{BP}(95)$	0.06	0.82	0.01	0.92	0.02	0.92	0.02	1.00
	$\hat{\theta}^*_{FDA}(95)$	0.20	0.76	0.20	0.81	0.14	0.91	0.22	0.98
	$\hat{\theta}^*_{BP}(95)$	0.21	0.99	0.20	1.00	0.14	1.00	0.22	1.00
2	$\hat{\theta}_{FDA}(95)$	0.04	0.61	0.00	0.62	0.00	0.80	0.00	0.95
	$\hat{\theta}_{BP}(95)$	0.04	0.77	0.00	0.77	0.00	0.91	0.00	0.98
	$\hat{\theta}^*_{FDA}(95)$	0.11	0.74	0.08	0.74	0.06	0.90	0.06	0.96
	$\hat{\theta}^*_{BP}(95)$	0.13	0.94	0.08	1.00	0.06	1.00	0.06	1.00
3	$\hat{\theta}_{FDA}(95)$	0.00	0.64	0.00	0.83	0.00	0.85	0.00	0.95
	$\hat{\theta}_{BP}(95)$	0.00	0.64	0.00	0.83	0.00	0.85	0.00	0.95
	$\hat{\theta}^*_{FDA}(95)$	0.04	0.93	0.03	1.00	0.01	0.99	0.00	1.00
	$\hat{\theta}^*_{BP}(95)$	0.07	0.93	0.06	1.00	0.06	1.00	0.02	1.00
4	$\hat{\theta}_{FDA}(95)$	0.00	0.46	0.00	0.53	0.00	0.60	0.00	0.78
	$\hat{\theta}_{BP}(95)$	0.00	0.63	0.00	0.63	0.00	0.74	0.00	0.90
	$\hat{\theta}^*_{FDA}(95)$	0.04	0.64	0.04	0.71	0.01	0.82	0.02	0.89
	$\hat{\theta}^*_{BP}(95)$	0.10	0.92	0.07	0.93	0.08	0.99	0.05	1.00

4. A REAL-WORLD EXAMPLE

In this section, we studied the performances of the iterated and traditional bootstrap methods for assessing IBE by a real-world example. For this purpose, we used a dataset given by the FDA: The antihypertensive patch dataset (see Table 5) which is consisted of a total of 37 subjects. For this dataset, an antihypertensive patch were administered to the first sequence of 18 patients in the order of $TRRT$, and in the second sequence, the treatments were administrated in the order of

$RTTR$ to 19 patients. This dataset have a large subject-by-formulation interaction ($\sigma_D > 0.15$), and hence it is necessary to test the IBE. We used logarithmically transformed AUC and C_{max} datasets to test the IBE. Both classical and iterated bootstrap methods were used to constructed upper limit of 95% confidence interval for IBE and the results are presented in Table 6.

The dataset given in Table 5 is analyzed using the REML method by [7] and based on their results they state that: (i) ABE can be concluded based on the values obtained for AUC, but the C_{max} data fail, (ii) PBE can be concluded in either case, i.e., based on the AUC and on the C_{max} data, and (iii) IBE cannot be concluded in either case. This may be due to a high subject-by-formulation interaction that seems to be present in this dataset, as pointed out by the FDA. According to our results (see Table 6) IBE can be concluded for C_{max} dataset but the IBE test is rejected for AUC dataset (since the upper limits of 95% CI calculated for this dataset are greater than the boundary of the null hypothesis specified by the FDA) when $\hat{\theta}_{FDA}(95)$ and $\hat{\theta}_{BP}(95)$ are used. On the other hand, IBE can be concluded in either case when the iterated bootstrap upper confidence limits ($\hat{\theta}^*_{FDA}(95)$ and $\hat{\theta}^*_{BP}(95)$) are used.

5. Conclusion

In this study, we propose to use iterated bootstrap algorithm to test the individual bioequivalence hypothesis under 2×4 randomized crossover design, and we compare their performances with the conventional bootstrap methods both by simulations and a case study. The important result produced by iterated bootstrap algorithm is that the power of the test obtained by this method are significantly better than those obtained by classical bootstrap methods. Hence, this paper shows that more reliable results can be obtained by using iterated bootstrap method to assess the individual bioequivalence. On the other hand, the proposed iteration inflates the type I error rate to approximately 20% (but have more power than the traditional bootstrap) when the parameter θ is close to the boundary of the null hypothesis of individual bioequivalence θ_U (as in the first parameter setting (see 2)) and the sample size is small. Such high type I error rate may not be acceptable in drug studies since high type I error increases the chance of rejecting the null hypothesis when it is actually true. In this case, the number of iterations and/or bootstrap replicates B can be increased to decrease the type I error rate. As a final comment, it should be noted that the iterated bootstrap requires much more computational cost than the traditional bootstrap. For example, while the bootstrap runs only for B resamples in a simulation iterated bootstrap requires $B^2 + B$ resamples in the first iteration (double bootstrap), $B^3 + B^2 + B$ resamples in the second iteration, and so on. However, it is worth trying iteration considering its performance and the increasing technology.

Acknowledgement. The author thanks the anonymous referees for their valuable suggestions and comments, which helped to improve the manuscript significantly.

TABLE 5. The antihypertensive patch dataset

Subject	Sequence	Period							
		Period (AUC)				Period (C_{max})			
		1	2	3	4	1	2	3	4
1	RTTR	1020.65	1321.31	900.42	1173.61	109	145	106	146
2	TRTR	950.59	1637.71	2076.75	1485.93	96.3	194	341	316
3	RTTR	1188.82	1440.99	1501.20	1481.27	128	155	138	192
4	TRTR	774.44	585.89	801.26	773.51	87.6	56.2	89.1	84
5	TRTR	1563.08	1571.75	1917.37	1886.05	161	145	194	178
6	RTTR	1119.22	781.20	800.85	942.50	119	66.9	82	117
7	RTTR	1876.81	1726.01	1653.70	1111.10	232	170	194	135
8	TRTR	2549.54	3738.21	3800.33	5408.38	229	393	395	677
9	TRTR	2291.93	1223.74	1949.10	3184.15	204	126	202	365
10	RTTR	1392.92	826.36	1220	1607.52	222	68.6	112	200
11	RTTR	5239.22	8894.11	7726.47	7451.66	871	1710	1090	1450
12	TRTR	1044.18	1023	1178.20	1155.25	91.1	111	196	104
13	TRTR	744.57	985.58	1721.01	4217.64	80.2	127	215	413
14	RTTR	1629.67	2081.88	1302.65	2805.07	168	263	134	355
15	RTTR	3054.97	3370.78	2644.44	5941.36	323	502	401	630
16	TRTR	3469	1712.59	1680.07	3285.23	449	284	141	405
17	TRTR	3006.95	3063.28	1764.34	2055.51	289	277	162	203
18	RTTR	2323.41	1063.45	960.10	2629.35	344	131	101	718
19	TRTR	4989.43	6439.82	4945.42	2321.03	744	1150	769	263
20	RTTR	2673.38	1686.63	2260.34	4632.96	361	226	538	691
21	TRTR	2081.19	1028.75	758.83	1168.12	295	108	73.2	140
22	RTTR	10843.61	13162.65	13505.79	13575.90	1530	1330	1520	1650
23	TRTR	736.50	947.58	1426.96	681.66	87.4	124	151	75.5
24	RTTR	2747.09	3651.63	2543.63	1056.48	353	480	300	110
25	TRTR	2064.25	2251.24	2228.06	2633.27	253	414	314	470
26	TRTR	1092.48	1141.68	1550.98	996.55	138	118	163	95.5
27	RTTR	2011.28	2109.67	2902.35	2283.60	467	444	512	495
28	RTTR	3793.47	4165.73	4666.95	3274.41	727	454	471	473
29	RTTR	1427.53	1591.38	1909.97	1911.43	139	183	167	164
30	TRTR	2333.74	2878.94	1698.30	1142.33	308	355	156	98.9
31	RTTR	1932.80	1620.69	2279.44	3251.14	334	228	289	528
32	TRTR	1835.61	2760.92	3188.04	2480.39	167	232	321	236
33	TRTR	8330.61	6064.54	8737.60	8353.62	954	873	857	930
34	RTTR	3612.64	2494.45	3153.79	6386.19	491	417	527	1010
35	RTTR	1061.92	987.86	1422.71	1220.58	97.4	94.1	186	103
36	TRTR	2212.39	1438.48	1984.76	2640.43	226	137	237	237
37	RTTR	2252.76	2262.88	1957.66	3084.05	304	255	301	685

REFERENCES

[1] Beran, R., Prepivoting to reduce level error of confidence sets, *Biometrika* (1987), 74, 151-173.

[2] DiCiccio, T. J. and Romano, J. P., A Review of Bootstrap Confidence Intervals, *Journal of The Royal Statistical Society Series B-statistical Methodology* (1998), 50, 338-354.

TABLE 6. Results of IBE for the antihypertensive patch dataset

Pharmacokinetic parameters	Method	Upper limit of 95% CI	θ_U
AUC	$\hat{\theta}_{FDA}(95)$	2.8028	2.4948
	$\hat{\theta}_{BP}(95)$	2.5410	2.4948
	$\hat{\theta}^*_{FDA}(95)$	1.1510	2.4948
	$\hat{\theta}^*_{BP}(95)$	0.7868	2.4948
$C_{\max}$	$\hat{\theta}_{FDA}(95)$	1.9648	2.4948
	$\hat{\theta}_{BP}(95)$	1.9648	2.4948
	$\hat{\theta}^*_{FDA}(95)$	0.4863	2.4948
	$\hat{\theta}^*_{BP}(95)$	0.4863	2.4948

[3] FDA, In vivo bioequivalence studies based on population and individual bioequivalence approaches, Guidance for Industry, Rockville, MD: Center for Drug Evaluation and Research, Food and Drug Administration, 1997.
[4] FDA, Guidance for industry: Bioavaibility and bioequivalence studies for orally administered drug products: General Consideration, Rockville, MD: Center for Drug Evaluation and Research, Food and Drug Administration, 2003.
[5] Hall, P. On the bootstrap and confidence intervals, *Annals of Statistics* (1986), 54, 1431-1452.
[6] Hall, P. and Martin, M. A., On bootstrap resampling and iteration, *Biometrika* (1988), 75, 661-671.
[7] Hauschke, D., Steinijans, V., Pigeot, I., Bioequivalence studies in drug development, John Willey & Sons, 2007.
[8] Jones, B. and Kenward, M. G., Design and analysis of cross-over trials, Chapman & Hall/CRC, 2003.
[9] Martin, M. A., On bootstrap iteration for coverage correction in confidence intervals, *Journal of the American Statistical Association* (1990), 85, 1105–1118.
[10] Pigeot, I., The bootstrap percentile in Food and Drug Administration Regulations for bioequivalence assessment, *Drug Information Journal* (2001), 35, 1445-1453.
[11] Pigeot, I., Hauschke, D. and Shao, J., The bootstrap in bioequivalence studies, *Journal of Biopharmaceutical Statistics* (2011), 21, 1126-1139.
[12] Schall, R. and Luus, H. G., On population and individual bioequivalence, *Statistics in Medicine* (1993), 12, 1109-1124.
[13] Schall, R., Assessment of individual and population bioequivalence using the probability that bioavailabilities are similar, *Biometrics* (1995), 54, 615-626.
[14] Schall, R. and Williams, R. L., Towards a practical strategy for assessing individual bioequivalence, *Journal of Pharmacokinetics and Biopharmaceutics* (1996), 24, 133-149.
[15] Senn, S. J., Cross-over trials in clinical research, John Willey, 1993.
[16] Shao, J., Chow, S. C. and Wang, B., The bootstrap procedure in individual bioequivalence, *Statistics in Medicine* (2000), 19, 2741-2754.
[17] Shao, J., Kubler, J. and Pigeot, I., Consistency of the Bootstrap Procedure in Individual Bioequivalence, *Biometrika* (2000), 87, 573-585.

Current address: Ufuk Beyaztas: Bartin University, Department of Statistics, Bartin/TURKEY
E-mail address: ufukbeyaztas@gmail.com
ORCID Address: https://orcid.org/0000-0002-5208-4950