Alternative Cross-over Designs for Individual Bioequivalence

Terry Hyslop [1], Boris Iglewicz [2]

[1] Biostatistics Section, Division of Clinical Pharmacology, Thomas Jefferson University, 402 Sheridan Building, 125 S. 9th Street, Philadelphia, PA 19107 Terry.Hyslop@mail.tju.edu

[2] Department of Statistics, Temple University, Speakman Hall 006-00, 1810 North 13th Street, Philadelphia, PA 19122-6083, iglewicz@sbm.temple.edu

Abstract

Three alternative cross-over study designs are considered for testing Individual Bioequivalence (IBE) according to the FDA criteria. Two types of three period designs and also a conservative test which can be accomplished in the usual two period setting are proposed. We introduce a partial-replicate design with three sequences, which includes replication of only the reference product. The estimation procedures for each of the designs and formulas to test IBE by the FDA criteria with a one-sided upper confidence bound are provided. We find that the partial-replicate design is far superior to typical three period designs for testing IBE. In many cases, it is also superior to or equivalent to the four-period design in terms of number of required product administrations, and may provide a practical alternative for IBE studies of formulations with longer half-lives. These methods will be demonstrated through a real data example.

Key Words. Individual Bioequivalence; confidence interval; cross-over design.

1. Introduction

The FDA has prepared a draft guidance (1) outlining the proposed implementation of individual bioequivalence (IBE) in pivotal bioequivalence studies. In addition, the general BA/BE preliminary guidance (2) suggests that all pivotal bioequivalence trials should be replicate design trials. Such designs could be used to assess intra-subject variability for the test (T) and reference (R) formulations in addition to the usual mean comparisons currently used in average bioequivalence studies (ABE).

The guidance provides estimation methods, sample size requirements, and confidence interval procedures for two sequence four period trials. The confidence interval based testing procedure described in the guidance utilizes the methods of Hyslop, Hsuan, and Holder (3) to conclude whether two formulations are individually bioequivalent. Four period designs present the additional challenges of patient recruitment and patient retention for the duration of the study. There is reason to believe that alternative designs will be desirable, particularly if those designs are more efficient in terms of sample size and/or the required number of administrations of drug products.

We have evaluated three alternative study designs with regard to the FDA's IBE criteria, using an extension of the methods proposed by Hyslop et al (3). Two of the designs have three periods, and the third is the usual two period crossover (TR/RT), which is typically used for average bioequivalence. We determined that IBE can be tested in each of these designs. We also found that the replication of just the R product in a three period design is sufficient to test IBE. These designs, if efficient, will offer the researcher additional study design options which may impact either the overall cost of a bioequivalence study, or at least the researcher's capability to complete the study with fewer missing data points.

Wang (4) presented a method to test IBE for a two sequence three period design (TRR/RRT). However, his paper presents inference methods for only one of the two FDA proposed criteria, he provides power and sample size for limited cases, and the reparametrization required is not intuitive. Additionally, the design he uses does not allow for administration of the test product in the second period. Kimanani and Potvin (5) have proposed an alternative criterion for IBE and present three period methods. Unfortunately, the criterion they propose is design-dependent, so that their approach is not comparable to the FDA proposal. Estimation methods for the FDA's IBE criteria, variance components of interest in IBE studies, and associated inference methods for alternative designs have not been prevalent in the literature.

We present three designs for consideration. First we will compare two types of three period designs. The first three period design has two sequences, where, one sequence contains replicates on the reference formulation (for example, RTR or TRR), while the other sequence would then have replicates on the test

formulation (TRT, RTT). We will refer to this design as a *full-replicate design*. We introduce the second three period design, which has three sequences , for example, TRR/RRT/RTR. This design has replication only on the reference formulation, and we will refer to it as a *partial-replicate design*. Wang (4) considered the first two sequences of this design, but our design is more balanced since the test formulation occurs in each period. In addition to comparing the efficiency of these two designs, we will also present some proposed estimation methods for the criteria. We also present a conservative two period test, for formulations with low intra-subject variance. This test could be utilized for long half-life drugs or for formulations with a small intra-subject variance, provided that sufficient prior information on the magnitude of the intra-subject variability is available.

This paper is organized as follows. We describe the statistical model in Section 2, and introduce a real data example in Section 3. Section 4 outlines the three period design alternatives with their associated estimation and inference procedures. We return to the numerical data example in Section 5. In Section 6, we describe our proposal for a conservative two period test. In Section 7, we present simulation study results which compare the efficiency of the designs as well as the validity of the level of the tests under boundary conditions. Finally, Section 8 contains our conclusions on the efficiency and applicability of these alternative crossover designs.

2. Statistical Model

We consider the following statistical model proposed by Chinchilli and Esinhart (6) with the assumption of no carry-over effects:

$$Y_{ijkl} = \mu_k + \gamma_{ikl} + \delta_{ijk} + \varepsilon_{ijkl} \qquad (2.1)$$

where $i = 1,\ldots,s$ indicates sequence, $j = 1,\ldots,n_i$ indicates subject within sequence i, $k = R,T$ indicates treatment, $l = 1,\ldots,p_{ik}$ indicates replicate on treatment k for subjects within sequence i. Y_{ijkl} is the natural logarithm of the response for replicate l on treatment k for subject j in sequence i, γ_{ikl} represents the fixed effect of replicate l on treatment k in sequence i, δ_{ijk} is the random effect for subject j in sequence i on treatment k, and ε_{ijkl} is the random error for subject j within sequence i on replicate l of treatment k. We assume that under normality of Y_{ijkl}, the ε_{ijkl}'s are mutually independent and identically distributed as

$$\varepsilon_{ijkl} \sim N\left(0, \sigma_{Wk}^{\ 2}\right), \qquad (2.2)$$

for $i = 1,\ldots,s$, $j = 1,\ldots,n_i$, $k = R,T$, and $l = 1,\ldots,p_{ik}$. Also, the random subject effects $\mathbf{m_{ij}} = \left(m_{ijR}, m_{ijT}\right)' = \mu_k + \delta_{ijk}$ are assumed to be mutually independent and distributed as

$$\mathbf{m_{ij}} \sim N_2\left[\begin{pmatrix} \mu_R \\ \mu_T \end{pmatrix}, \begin{pmatrix} \sigma_{BR}^{\ 2} & \rho\sigma_{BT}\sigma_{BR} \\ \rho\sigma_{BT}\sigma_{BR} & \sigma_{BT}^{\ 2} \end{pmatrix}\right]. \qquad (2.3)$$

The following constraint is applied to the nuisance parameters to avoid overparametrization of the model for k=R, T and

$$p = p_{iT} + p_{iR} : \sum_{i=1}^{s} \sum_{l=1}^{p_{ik}} \gamma_{ikl} = 0 \qquad (2.4)$$

These constraints yield a set of $sp-2$ unconstrained nuisance parameters, so that the number of fixed location parameters corresponds to the number of cell means in the s x p crossover design. The nuisance parameters in this model represent sequence x period interactions nested within treatment. Note that while sequence and period effects are not explicitly described as separate model parameters, their estimates can be obtained from linear combinations of the nuisance effects parameters, γ_{ikl}.

The criteria for individual bioequivalence (1) in terms of the mixed model parameters are:

$$\frac{(\mu_T - \mu_R)^2 + \sigma_D^2 + (\sigma_{WT}^{\ 2} - \sigma_{WR}^{\ 2})}{\sigma_{WR}^{\ 2}} \le \theta_I \quad ,$$

$$\text{for } \hat{\sigma}_{WR} > \sigma_{W0} \qquad (2.5)$$

$$\frac{(\mu_T - \mu_R)^2 + \sigma_D^2 + (\sigma_{WT}^{\ 2} - \sigma_{WR}^{\ 2})}{\sigma_{W0}^{\ 2}} \le \theta_I \quad ,$$

$$\text{for } \hat{\sigma}_{WR} \le \sigma_{W0}. \qquad (2.6)$$

The methods of Hyslop et al. (3) are based on linear combinations of random variables, so we will work with the following linearized criteria to test individual bioequivalence:

$$\eta_1 = (\mu_T - \mu_R)^2 + \sigma_D^2 + (\sigma_{WT}^{\ 2} - \sigma_{WR}^{\ 2}) - \theta_I \cdot \sigma_{WR}^2 < 0 \,,$$

$$\text{for } \hat{\sigma}_{WR} > \sigma_{W0} \qquad (2.7)$$

$$\eta_2 = (\mu_T - \mu_R)^2 + \sigma_D^2 + (\sigma_{WT}^{\ 2} - \sigma_{WR}^{\ 2}) - \theta_I \cdot \sigma_{W0}^2 < 0 \,,$$

$$\text{for } \hat{\sigma}_{WR} \le \sigma_{W0} \qquad (2.8)$$

These linearized criteria provide the means for a test of IBE. In practice, a 5% test would be based on an upper 95% confidence bound of either η_1 or η_2, depending on the magnitude of

the study estimate of intra-subject standard deviation, and the regulatory constants, σ_{W0}, which the guidance specifies as 0.20, and $\theta_I = 2.495$ (1). If the upper confidence bound of the appropriate linearized criterion is less than 0, then IBE is concluded.

3. 3-period full-replicate and partial-replicate designs

The first proposed three period design has two sequences, where, without any loss of generality, we assume that the first sequence contains replicates on the reference and one period would be assigned to the test formulation. The second sequence would then consist of replicates on the test formulation with one measure of the reference formulation. An example of such a design is RTR/TRT. We will refer to this type of design as a *full-replicate design*.

We introduce the following reparametrizations, for the full-replicate design for each subject j in sequence 1, $j = 1, \ldots, n_1$:

$$I_{1j} = Y_{1jT1} - \bar{Y}_{1jR\cdot},$$

$$R_{1j} = Y_{1jR1} - Y_{1jR2},$$

and for each subject j in sequence 2, $j = 1, \ldots, n_2$:

$$I_{2j} = \bar{Y}_{2jT\cdot} - Y_{2jR1},$$

$$T_{2j} = Y_{2jT1} - Y_{2jT2}. \quad (4.1)$$

for subject j on formulations R and T. To estimate the subject by formulation interaction variance component in the full-replicate design, we begin by computing:

$$\hat{\sigma}_{Ii}^{\,2} = \frac{1}{n_i - 1}\sum_{j=1}^{n_i}\left(I_{ij} - \bar{I}_i\right)^2,\ i = 1, 2. \quad (4.2)$$

The details of the remaining variance component computations for this design are not provided. The pooled variance component estimator, $\hat{\sigma}_{I_p}^{\,2}$, estimates the subject by formulation variance component plus some additional intra-subject variance components in this design. These additional intra-subject variance components can then be combined with the unbiased intra-subject variance component estimators (Chinchilli and Esinhart, 6) in a linear combination to obtain the form of the IBE criteria. For example, based on information in Table 1a, and after some algebra,

$$E\left(\hat{\sigma}_{I_p}^{\,2} + \frac{1/2(n_2 - 1)}{(n_1 + n_2 - 2)}\bullet\hat{\sigma}_{WT}^2 + \frac{1/2(n_1 - 1)}{(n_1 + n_2 - 2)}\bullet\hat{\sigma}_{WR}^2 - (2 + \theta_I)\bullet\hat{\sigma}_{WR}^2\right)$$
$$= \sigma_D^2 + \sigma_{WT}^2 - (1 + \theta_I)\bullet\sigma_{WR}^2.$$

In the full-replicate design, the above linear combination of estimators can be used to estimate the reference-scaled IBE criterion in equation (2.7). Also, with some minor adjustments, a similar combination can be used to estimate the variance components in equation (2.8):

$$\hat{\sigma}_{I_p}^{\,2} + \frac{1/2(n_2 - 1)}{(n_1 + n_2 - 2)}\bullet\hat{\sigma}_{WT}^2 + \frac{1/2(n_1 - 1)}{(n_1 + n_2 - 2)}\bullet\hat{\sigma}_{WR}^2 - (2)\bullet\hat{\sigma}_{WR}^2 - (\theta_I)\bullet\hat{\sigma}_{W0}^2$$

The second proposed three period design has three sequences, for example RTR/TRR/RRT. This design has replication only on the reference formulation in each of the sequences. For this reason, we will refer to it as a *partial-replicate design*. The following alternate reparametrizations are required for the partial-replicate design:

$$I_{ij} = Y_{ijT1} - \bar{Y}_{1jR\cdot} \qquad R_{ij} = Y_{ijR1} - Y_{ijR2} \quad (4.3)$$

for subject j in sequence i, $j = 1, \ldots, n_i$, $i = 1, \ldots, s$.

Also compute the subject by formulation component of variance in this design by computing:

$$\hat{\sigma}_{Ii}^{\,2} = \frac{1}{n_i - 1}\sum_{j=1}^{n_i}\left(I_{ij} - \bar{I}_i\right)^2,\ i = 1, 2, 3. \quad (4.4)$$

The additional variance component computations for this design are not provided. As with the full-replicate design, linear combinations of the estimators provide unbiased estimates of the variance component parameters in the FDA criteria in equations (2.7) and (2.8). The variance component estimators and their distributions under the assumption of normality can be used to provide the necessary elements for point and confidence interval estimation to test IBE.

Using the described reparametrizations, we have modified the confidence interval procedures described by Hyslop et al. (3) for these designs. The basic concept is to find a linear combination of independent random variables of known distribution which can be used to estimate the statistic of interest, in this case, the IBE criteria. The methods for the full-replicate design are summarized in Table 2a, while the methods for the partial-replicate design are summarized in Table 2b.

---Insert Table 2a and Table 2b here --

4. A conservative test in two period two sequence design

We next considered the two period, two sequence design which is the standard design used for average bioequivalence. In this design, TR/RT, the difference of the means can be

assessed, and the mean square error from this design estimates:

$$E(MSE) = \tfrac{1}{2}\left(\sigma_D^2 + \sigma_{WT}^2 + \sigma_{WR}^2\right) \qquad (6.1)$$

Then,

$$E\left(\hat{\delta}^2 + 2 \cdot MSE\right) \approx \delta^2 + \sigma_D^2 + \sigma_{WT}^2 + \sigma_{WR}^2 . \qquad (6.2)$$

The right side in (4.2) will always be greater than both the reference-scaled and constant-scaled linearized criteria. Hence, an individual bioequivalence test based on (4.2) would be conservative. For large values of σ_{WR}^2, it will be too conservative to be useful, as we will demonstrate. However, for small values of σ_{WR}^2, the use of this estimation method provides a useful alternative to higher-order replicate designs.

Balaam's design (TR/RT/TT/RR) could also be utilized as a 2 period design. Separate estimation of σ_{WR}^2 and σ_{WT}^2 could be computed and inference could again follow the techniques outlined in Hyslop et al (3) for the complete IBE criterion (reference and constant scaled) instead of a partial estimator. This design would be very inefficient as only approximately 25% of the subjects would be utilized to estimate the intra-subject variance components, resulting in a relatively small number of degrees of freedom for inference.

5. Simulation studies

We completed simulation studies to show that the approximate confidence intervals maintain nominal level at the null hypothesis boundary and to compare the efficiency of the designs we considered over a range of parameter values. For the full-replicate design, simulations were based on the two sequence three period crossover replicate design for two formulations, T and R. For the partial-replicate design, simulations were based on three sequence three period crossover designs with replication on only one formulation (R). For the two period two sequence design, simulations were based on the sequences TR/RT. Simulations were performed in SAS v 6.12 on a Pentium PC. For each simulation, random deviates from normal and chi-square distributions (with appropriate degrees of freedom) were selected to correspond to the parameters for the selected study design with sample size, means, and variances as listed in the tables and figures. Upper confidence limits were then computed for each simulation according to the procedures outlined in Sections 4 and 6. Power and level estimates were determined by computing the proportion of simulations with $H_{\eta_1} < 0$ and $H_{\eta_2} < 0$ for the reference-scaled and constant-scaled criterion, respectively. Level estimates were computed for various combinations of parameters on the IBE boundary. Simulation scenarios included studies where δ was on the IBE boundary, σ_D was assumed to be 0, and $\sigma_{WT}^2 = \sigma_{WR}^2$. We also considered studies where $\sigma_{WT}^2 / \sigma_{WR}^2 = \{0.90, 0.95, 1.05, 1.10\}$ and $\sigma_D = \{0.05, 0.10, 0.20\}$ and δ was then computed so that the resulting criterion was on the IBE reference-scaled or constant-scaled boundary. Sample sizes for boundary simulations included {10, 20, 30, 40} subjects. Level estimates are based on 15,550 simulations per scenario assuring that 95% confidence intervals for the nominal values at 5% are within ± .34%. Power estimates are based on 1,540 simulations per scenario assuring that 95% confidence intervals of estimated power are no wider than ± 2.5%. Sample size estimates to obtain 90% power were computed by simulating a larger number of sample sizes for each of the designs. Sample size estimates were then obtained by examining the resulting empirical power estimates until the first sample size at which 90% power was exceeded. In all simulations, the results reported for power and level estimates are based on the FDA mixed criterion. This criterion consists of using the reference-scaled criterion (equation (2.7)) for $\hat{\sigma}_{WR} > 0.20$, and the constant-scaled criterion(equation (2.8)) for $\hat{\sigma}_{WR} \le 0.20$.

Both the full-replicate and partial-replicate designs maintained the approximate nominal level. Empirical alpha levels ranged from 4.2% to 5.8% over the range of parameters examined for these designs. As expected, the two period, two sequence design is quite conservative, with empirical levels ranging from <0.1 % to 1.1% over the range of parameters investigated. The two-period test procedure becomes most conservative for values of σ_{WR}^2 which approach σ_{W0}^2.

Power for the three-period designs are compared. The power increases when $\sigma_{WT} < \sigma_{WR}$ for both designs. Since a larger variance value is subtracted, the overall estimate is decreased, and the probability of declaring equivalence is increased. These figures also demonstrate the superiority of the partial-replicate design over the full-replicate design for IBE studies. For example, the partial-replicate design has

approximately 80% power when $n=30$, $\sigma_{WT}=\sigma_{WR}=0.30$, $\mu_T-\mu_R=0.05$, $\sigma_D=0$; while the full-replicate design has only about 70% power for this same scenario. Similar results were obtained for other levels of reference product intra-subject standard deviation (data not shown).

In Table 3 and Table 4 we present a comparison of three designs in terms of power to conclude individual bioequivalence. We compare the two period two sequence design (2P2S), the three period partial-replicate design (3P3S), and the four period 2 sequence design (4P2S). We have excluded the three period full-replicate design from this table since we have shown that it is inefficient for IBE in comparison to the partial-replicate design. Table 3 presents simulation results in terms of number of subjects required for 90% power. Table 4 presents simulation results in terms of number of administrations (# of subjects * # of periods). The simulations were carried out as described above for all three of the designs. In addition, the minimum sample size considered was 12, since studies of smaller sample sizes would not meet the FDA's current considerations (1).

---Insert Table 3 here --

In terms of number of subjects, the 2P2S design is as efficient as the 3P3S design and the 4P2S design only for those cases where σ_{WR}=0.10 and $\sigma_{WT}<\sigma_{WR}$ (Table 3). However, the 2P2S is more efficient in terms of number of administrations for several more cases where σ_{WR}=0.10 and $\sigma_{WT}\geq\sigma_{WR}$ (Table 4). The two-period design should not be used for $\sigma_{WR}\geq 0.20$, as the test is extremely conservative, and simulations of up to n = 62 still do not achieve even 50% power for $\sigma_{WT}<\sigma_{WR}$. For all remaining variability levels, the 4P2S design is always the most efficient in terms of number of subjects required for IBE studies with 90% power (Table 3).

In terms of number of administrations, the 2P2S design is again preferred for σ_{WR}=0.10 except for the case when σ_{WT}/σ_{WR}=1.20 and $\Delta=0.10$. In this case, the 3P3S design has a slight advantage. For values of $\sigma_{WR}\geq 0.20$, the 2P2S design should not be used, and the advantage in terms of design fluctuates between the 3P3S and 4P2S designs. The 4P2S design is more efficient for low to moderate variability (σ_{WR}=0.20-0.30), while the 3P3S design is usually more efficient for higher variability ($\sigma_{WR}\geq 0.30$).

6. Discussion and Conclusions

The design of crossover studies for experiments requires careful planning. Putt and Chinchilli (7), Jones and Kenward (8), and Senn (9) describe some of the considerations which should occur in the design phase. Noteably, investigators need to evaluate whether there is the possibility of sequence effects, period effects, and simple carryover effects. Design selection should be based on careful consideration of these possibilities. For individual bioequivalence (IBE), we have investigated some of the additional factors which may influence study design, namely the related costs of subject recruitment and subject retention, and of study conducted lab work.

It is well known that the addition of sequences (Balaam's design) and/or periods (three period and four period designs) not only improves the efficiency of the treatment effect estimates, but can also allow for estimation of intra-subject variances and orthogonal estimation of treatment and carryover effects (6,7,8,9,10). This literature provides guidance to the researcher on optimal designs for estimation of fixed effects. Additional work has been presented concerning subject retention in higher order crossover designs (11, 12). With the implementation of IBE and PBE, future research is needed to assess the optimality of higher order designs which include both fixed and random effect estimation.

For IBE studies, the four-period two-sequence design offer minimum subject recruitment for the situations we considered, and also provides a strongly balanced design (6-7) which allows for estimation of simple carryover, as well as sequence and period effects. However, if subject retention is an important factor, the three-period partial-replicate design, introduced here, offers the minimum number of administrations (number of subjects* number of periods) in many of the simulated cases. The partial-replicate design would be desirable if there is concern that subjects may not complete a higher-order crossover study, or if costs of administration and evaluation are prohibitive for a four period design. We have shown that the three period partial-replicate design is superior to the three period full-replicate design for testing IBE with the FDA criteria.

The usual two-period setting may also be used for testing IBE when the intra-subject variability is sufficiently small. Additionally, in many

situations, the partial-replicate design is equivalent to or superior to the four period design in terms of number of required product administrations, providing an efficient alternative for longer half-life formulations.

Table 2a. Confidence interval construction for full-replicate design (RTR / TRT).

$H_{U_0}=(1-\alpha)$ level upper confidence limit	E_{U_0}=point estimate	$U_{U_0}=(H_{U_0}-E_{U_0})^2$
$H_D=\left(\lvert\hat{\delta}\rvert+t_{1-\alpha,n_1+n_2-2}\left(\frac{1}{4}\sum_{i=1}^{2}n_i^{-1}\hat{\sigma}_{WT}^2\right)^{1/2}\right)^2$	$E_D=\hat{\delta}^2$	U_D
$H_I=\frac{((n_1+n_2-2)\cdot E_I)}{\chi^2_{\alpha,n_1+n_2-2}}$	$E_I=\hat{\sigma}_{I_P}^2$	U_I
$H_T=\frac{((n_2-1)\cdot E_T)}{\chi^2_{1-\alpha,n_2-1}}$	$E_T=\frac{0.5\cdot(n_2-1)}{(n_1+n_2-1)}\hat{\sigma}_{WT}^2$	U_T
$H_R=\frac{((n_1-1)\cdot E_R)}{\chi^2_{1-\alpha,n_1-1}}$	$E_R=\left\{\frac{0.5(n_1-1)}{(n_1+n_2-2)}-(2+\theta_I)\right\}\cdot\hat{\sigma}_{WR}^2$	U_R
$H_R'=\frac{\left((n_1-1)\cdot E_R'\right)}{\chi^2_{1-\alpha,n_1-1}}$	$E_R'=\left\{\frac{0.5(n_1-1)}{(n_1+n_2-2)}-2\right\}\cdot\hat{\sigma}_{WR}^2$	U_R'

95% upper confidence bounds of linearized criteria:

$$H_{U_0}=(E_D+E_I+E_T+E_R)+(U_D+U_I+U_T+U_R)^{1/2}$$

$$H_{U_0}=\left(E_D+E_I+E_T+E_R'-\theta_I\cdot\sigma_{W0}^2\right)+\left(U_D+U_I+U_T+U_R'\right)^{1/2}$$

Table 2b. Confidence interval construction for partial-replicate design (TRR / RTT / RTR).

$H_{U_0}=(1-\alpha)$ level upper confidence limit	E_{U_0}=point estimate	$U_{U_0}=(H_{U_0}-E_{U_0})^2$
$H_D=\left(\lvert\hat{\delta}\rvert+t_{1-\alpha,n^2+n^2+n^2-3}\left(\frac{1}{9}\sum_{i=1}^{3}n_i^{-1}\hat{\sigma}_{IVC}^2\right)^{1/2}\right)^2$	$E_D=\hat{\delta}^2$	U_D
$H_I=\frac{((n_1^2+n_2^2+n_3^2-3)\cdot E_I)}{\chi^2_{\alpha,n_1+n_2+n_3-3}}$	$E_I=\hat{\sigma}_{IVP}^2$	U_I
$H_R=\frac{((n_1^2+n_2^2+n_3^2-3)\cdot E_R)}{\chi^2_{1-\alpha,n_1+n_2+n_3-3}}$	$E_R=-(1.5+\theta_I)\cdot\hat{\sigma}_{WR}^2$	U_R
$H_R'=\frac{((n_1+n_2+n_3-3)\cdot E_R')}{\chi^2_{1-\alpha,n_1+n_2+n_3-3}}$	$E_R'=-(1.5)\cdot\hat{\sigma}_{WR}^2$	U_R'

95% upper confidence bounds of linearized criteria:

$$H_{U_0}=(E_D+E_I+E_R)+(U_D+U_I+U_R)^{1/2}$$

$$H_{U_0}=\left(E_D+E_I+E_R'-\theta_I\cdot\sigma_{WR}^2\right)+\left(U_D+U_I+U_R'\right)^{1/2}$$

Table 3. Comparison of IBE Designs for 90% power*, based on number of subjects.

	$\Delta=0.00$ $\sigma_D=0.0$			$\Delta=0.05$ $\sigma_D=0.0$			$\Delta=0.10$ $\sigma_D=0.0$		
σ_{WR}/σ_{WT}	$\sigma_{WR}=0.10$			$\sigma_{WR}=0.10$			$\sigma_{WR}=0.10$		
	0.80	1.00	1.20	0.80	1.00	1.20	0.80	1.00	1.20
2P2S	**12**	14	16	**12**	14	18	**12**	16	20
3P3S	**12**	**12**	**12**	**12**	**12**	**12**	**12**	**12**	**12**
4P2S	**12**	**12**	**12**	**12**	**12**	**12**	**12**	**12**	**12**
σ_{WR}/σ_{WT}	$\sigma_{WR}=0.20$			$\sigma_{WR}=0.20$			$\sigma_{WR}=0.20$		
	0.80	1.00	1.20	0.80	1.00	1.20	0.80	1.00	1.20
2P2S	--	--	--	--	--	--	--	--	--
3P3S	18	27	45	21	27	48	36	36	54
4P2S	**14**	**20**	**30**	**14**	**20**	**34**	**16**	**24**	**40**
σ_{WR}/σ_{WT}	$\sigma_{WR}=0.30$			$\sigma_{WR}=0.30$			$\sigma_{WR}=0.30$		
	0.80	1.00	1.20	0.80	1.00	1.20	0.80	1.00	1.20
2P2S	--	--	--	--	--	--	--	--	--
3P3S	27	45	75	30	48	75	33	48	75
4P2S	**24**	**34**	**54**	**24**	**36**	**60**	**24**	**40**	**64**
σ_{WR}/σ_{WT}	$\sigma_{WR}=0.40$			$\sigma_{WR}=0.40$			$\sigma_{WR}=0.40$		
	0.80	1.00	1.20	0.80	1.00	1.20	0.80	1.00	1.20
2P2S	--	--	--	--	--	--	--	--	--
3P3S	27	45	75	30	48	75	33	48	78
4P2S	**24**	**34**	**58**	**24**	**36**	**58**	**24**	**36**	**64**
σ_{WR}/σ_{WT}	$\sigma_{WR}=0.50$			$\sigma_{WR}=0.50$			$\sigma_{WR}=0.50$		
	0.80	1.00	1.20	0.80	1.00	1.20	0.80	1.00	1.20
2P2S	--	--	--	--	--	--	--	--	--
3P3S	30	45	75	30	48	75	33	48	78
4P2S	**22**	**36**	**58**	**22**	**36**	**60**	**24**	**36**	**64**

* Power and sample size estimates based on 1,540 simulations per combination of parameters for each study design. Simulation methods are described in Section 5.

References

1. FDA (1999). Guidance for Industry: Average, Population, and Individual approaches to establishing bioequivalence. US Department of Health and Human Services, Food and drug administration, Center for Drug Evaluation and Research.
2. FDA (1999). General BA and BE guidance. US Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research.
3. Hyslop, T., Hsuan, F., Holder, D.J. A small sample confidence interval approach to individual bioequivalence, *Statistics in Medicine*, 19: 2885-2897, 2000.
4. Wang, W. On testing of individual bioequivalence, *Journal of the American Statistical Association*, 94, 880-887, 1999.
5. Kimanani, E.K., Potvin, D. A parametric confidence interval for a moment-based scaled criterion for individual bioequivalence, *Journal of Pharmacokinetics and Biopharmaceutics*, 25, 595-614, 1998.
6. Chinchilli V., Esinhart J. Design and analysis of intra-subject variability in cross-over experiments, *Statistics in Medicine*, 15, 1619-1634, 1996.
7. Putt M., Chinchilli V. A mixed effects model for the analysis of repeated measures cross-over studies, *Statistics in Medicine*, 18, 3037-3058, 1999.
8. Jones B., Kenward M.G. *Design and analysis of cross-over trials*, Chapman and Hall,London, U.K., 1989.
9. Senn, S. *Cross-over trials in clinical research*, Wiley, New York, 1993.
10. Chow S. -C., Liu J-P. *Design and analysis of bioavailability and bioequivalence studies.* Marcel Dekker, Inc. New York, p 33, 1992.
11. Low J.L., Lewis S.M., Prescott P. Assessing robustness of crossover designs to subjects dropping out, *Statistics and Computing*, 9, 219-227, 1999.
12. Richardson BA. Flack VF. The analysis of incomplete data in the three-period two-treatment cross-over design for clinical trials. *Statistics in Medicine*. 15(2), 127-143, 1996.