European Journal of Pharmaceutical Sciences 38 (2009) 55–63

Contents lists available at ScienceDirect

European Journal of Pharmaceutical Sciences

journal homepage: www.elsevier.com/locate/ejps

Comparison of the reference scaled bioequivalence semi-replicate method with other approaches: Focus on human exposure to drugs

Vangelis Karalis*, Mira Symillides, Panos Macheras

Faculty of Pharmacy, University of Athens, Panepistimiopolis, Athens 157 71, Greece

ARTICLE INFO

Article history:
Received 29 July 2008
Received in revised form 6 April 2009
Accepted 30 May 2009
Available online 11 June 2009

Keywords:
Bioequivalence
Scaled average bioequivalence
Scaled bioequivalence limits
Highly variable drugs
Levelling-off limits
Replicate design

ABSTRACT

To compare the performance of the reference scaled average bioequivalence ($scABE_R$) method proposed by FDA scientists [Haidar et al., 2008. Pharm. Res. 25, 237–241] with other approaches focusing on the human exposure expressed as the product *sample size* × *periods of drug administration*. Simulated bioequivalence studies were generated assuming the partial replicate 3-way crossover design and the classic (2 × 2) crossover design. Intrasubject variability (CV_W) values ranged from 15% to 60% and sample sizes from 16 to 54. The procedures examined include: the $scABE_R$ method, the classic 0.80–1.25 approach, a levelling-off scaled BE limit (BELscW), and some other scaled bioequivalence limits. To assess the performance of the aforementioned approaches, the typical as well as novel three-dimensional modified power curves were constructed. A new index, termed %Mean Relative Difference (MRD%), was introduced in order to quantitatively compare the performance of the bioequivalence limits. The recently proposed $scABE_R$ approach showed the lowest producer risk in particular for highly variable drugs. When exposure was taken into account $scABE_R$ resulted in a desired behaviour when CV_W was low. For high CV_W values the overall performance diminished when geometric mean ratio (GMR) substantially deviated from unity. Application of the MRD% index clearly revealed that the effect of lowering the producer risk at GMR = 1 was totally counterbalanced by the rise of consumer risk at high GMR values. The classic 0.80–1.25 limits were favoured at low intrasubject variability and high exposure, whereas the levelling-off limits demonstrated a preferred overall performance when variability was high and exposure was limited.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

The term "Bioequivalence" (BE) refers to the absence of a significant difference in the *rate* and *extent* to which the active moiety of drug products becomes available at the site action when administered at the same molar dose and under similar conditions (FDA, 2003). Classically, determination of BE relies on the application of two-period, two-treatment crossover studies (EMEA, 2001; FDA, 2001). Based on the concept of average bioequivalence two drug products are considered bioequivalent if the 90% confidence interval around their mean relative bioavailability is between preset limits (usually 0.80–1.25). However, determination of bioequivalence in case of highly variable drugs (HVD) is a difficult issue which has been recognized and discussed for a long time (Blume and Midha, 1993; Midha et al., 2007). A drug is considered to be highly variable when exhibits intrasubject variability (CV_W) greater or equal to 30%. The consequences of high CV_W on BE assessment have been highlighted in the past (Blume et al., 1995a,b; Blume and Midha, 1993; Shah et al., 1996).

Over the last years, several methods have been introduced to resolve the problem of high intrasubject variability in BE studies. In practice, a large number of subjects is recruited in the BE study in order to counterbalance the large variability and achieve the required statistical power. However, many concerns are raised when a large number of subjects are exposed to drugs (Benet, 1995). In order to face-off this drawback other methods, such as the widening of the BE limits to preset constant values such as 0.75–1.33 (EMEA, 2001; Blume et al., 1995a,b; Tothfalusi et al., 2003) or

Abbreviations: BE, Bioequivalence; BEL, the classic 0.80–1.25 bioequivalence limits; BELsc, scaled bioequivalence limits; BEL_r, the classic 0.80–1.25 bioequivalence limits applied to replicate designs; BELscC, bioequivalence limits using concomitantly a point GMR constrain; BELscM, mixed scaled bioequivalence limits; BELscW, levelling-off scaled bioequivalence limits using a Weibull function; $BELscW_R$, levelling-off scaled bioequivalence limits applied to replicate designs; CV_0, switching variability value; CV_W, intrasubject variability; CV_{WR}, intrasubject variability of the reference product, GMR, geometric mean ratio of the bioavailability measures; GMR_0, true geometric mean ratio; HVD, highly variable drug; MRD%, Mean Relative Difference; *N*, sample size; R, reference drug product; $scABE_R$, reference scaled average bioequivalence limits; $scABE_{Rc}$, reference scaled average bioequivalence limits adjusted for continuity at the switching variability value; s_{W0}, regulatory standardized variation; s_{WR}, intrasubject variability of the R product (of the log-transformed data); T, test drug product.

* *Corresponding author*. Tel.: +30 210 2610175.
E-mail address: vkaralis@pharm.uoa.gr (V. Karalis).

0928-0987/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.ejps.2009.05.013

the use of steady-state studies have been proposed (Blume et al., 1995a,b).

For highly variable drugs, Boddy et al. (1995) introduced a method for expanding the BE limits in proportion to CV_W. Other scaled procedures have also been proposed which include the application of a "mixed" criterion and a "constraint" on geometric mean ratio (GMR) (Tothfalusi and Endrenyi, 2003; Tothfalusi et al., 2003). According to the mixed method (Tothfalusi and Endrenyi, 2003) the classic average BE approach should be applied up to a "switching" variability value (CV_0) value, while scaled BE limits should be used if CV_W exceeds CV_0. The "constrained" approach (Tothfalusi et al., 2003) arises from the finding that large differences between the means can be observed if the scaled BE limits are solely applied. This approach suggests the use of scaled BE limits and apply concomitantly an additional regulatory criterion which constrains the point estimate of GMR in the range 0.80–1.25.

More recently, other modified scaled BE limits approaches have been introduced. These approaches incorporate both scaling with intrasubject variability and a GMR constraint into a single criterion (Karalis et al., 2004, 2005; Kytariolos et al., 2006). The advantages of these approaches can be attributed to their continuous nature as well as the levelling-off ability, i.e., to scale until a maximum "plateau" value. Finally, an approach proposed to resolve the problem of high CV_W and concomitantly reduce the total number of subjects was the use of replicate designs namely, designs where each subject receives the same treatment more than once (Blume et al., 1995a,b; FDA, 2001; Shah et al., 1996).

From a regulatory point view, special effort has been placed on resolving the issue of high variability in BE assessment during the last years. The European medicine evaluation agency (EMEA) issued a concept paper in 2006 seeking for possible solutions (EMEA, 2006). In the United States, this issue was discussed in several workshops. Very recently, FDA scientists published an article describing their current views on this topic (Haidar et al., 2008). Aiming at reducing the *exposure* of humans to the drugs during the clinical trials, a partial replicate design was proposed which allows the determination of intrasubject variability of the reference (CV_{WR}) product. According to the proposed design the reference (R) product should be administered twice in each subject, while the test (T) product only once. Additionally, a composite scaling procedure was proposed. For CV_{WR} values greater than a preset variability cut-off point, a scaled with CV_{WR} average BE criterion ($scABE_R$) is applied together with a point-estimate constraint imposed on the GMR between the test and reference products. The recently proposed $scABE_R$ approach (Haidar et al., 2008) aims at increasing the statistical power of the BE study and concomitantly constrain the humans' exposure to drugs by reducing sample size.

However, the issue of exposure is not only related to the number of subjects enrolled in the study but also to the total human exposure to drugs which can be expressed as the product *sample size* × *periods*. Aim of this manuscript is to compare the performance of the $scABE_R$ method (Haidar et al., 2008) with other approaches placing special emphasis on human exposure. Conclusions regarding the suitability of an approach will be gathered after adjusting the same exposure among the methods under evaluation.

2. Background-theory

2.1. Traditional method: average bioequivalence—classic BE limits

Traditionally, determination of average bioequivalence of two drug products is based on the comparison of the arithmetic means of a logarithmically transformed metric such as ln(AUC) and $\ln(C_{max})$. Two drug products are considered bioequivalent if the 90% confidence interval (CI) for the difference of their logarithmic means is between preset limits (θ) imposed by the regulatory authorities (Schuirmann, 1987; FDA, 2001; EMEA, 2001). Typically, the value of θ is set equal to $\theta = \ln(1.25)$. For the classic two-period, two-treatment, crossover design the upper/lower limits of the 90% CI are given by Eq. (1):

$$\text{Upper/Lower } 90\%\,\text{CI} = \exp\left[(\mu_T - \mu_R) \pm t_{0.05} \cdot S_{\text{res}}\sqrt{\frac{2}{N}}\right] \quad (1)$$

where $t_{0.05}$ is the Student criterion for 5% significance level, N is the number of subjects participating in the BE study, while μ_T and μ_R refer to the logarithmic mean of BE measure of the test (T) and reference (R) product, respectively. The term S_{res} refers to the residual variability calculated from ANOVA and is assumed to express the intrasubject variability. Obviously, as S_{res} increases it becomes more difficult to declare bioequivalence unless a large number of subjects is used in the study.

2.2. Scaled BE limits–scaled average BE

The scaled bioequivalence limits were introduced in order to resolve the problem of high producer risk for highly variable drugs (Boddy et al., 1995). The basic feature of scaled BE limits is their gradual expansion with intrasubject variability (Boddy et al., 1995; Midha et al., 1998; Tothfalusi and Endrenyi, 2003). Scaled BE limits are calculated as a fixed multiple (k) of intrasubject variability (Boddy et al., 1995; Midha et al., 1998):

$$\text{BELsc} = \exp(k \cdot S_{\text{res}}) \quad (2)$$

where BELsc corresponds to the upper BE limit.

"Mixed" scaled method represents a variant of scaled BE limits (Tothfalusi and Endrenyi, 2003). According to the mixed method, the classic 0.80–1.25 limits should be used until a preset "switching" variability (CV_0) value. For CV_W values exceeding CV_0 scaled BE limits should be used. However, the major drawback of all these scaled BE limits arises from the fact that scaled BE limits increase continuously with intrasubject variability allowing drug products with large differences in their mean values to be declared bioequivalent. In an effort to solve this problem the incorporation of a secondary criterion on GMR values was proposed in order to confine the GMR ratio between a lower (0.80) and a maximum value (1.25) (Tothfalusi et al., 2003).

The key point of the classic definition of average BE relies on the fact that two fixed values (0.80 and 1.25) are assigned to BE limits. Based on this observation, novel scaled BE limits were proposed which increase with intrasubject variability but only until a maximum GMR-dependent plateau value (Karalis et al., 2004, 2005). In addition, more simple "levelling-off" limits were proposed which do not include a GMR-related factor (Kytariolos et al., 2006). The basic feature of these BE limits can be ascribed to their gradual expansion which combines the performance of the classic average BE at low and moderate variability with the more permissive behaviour of the expanded BE limits at high variability values. The predominant features of flexibility, continuity and levelling-off properties make these BE limits suitable for the assessment of bioequivalence irrespective of the level of variability encountered.

2.3. Reference scaled average bioequivalence approach

More recently, scientists from the FDA working group on highly variable drugs proposed the approach of reference scaled average bioequivalence ($scABE_R$) for the determination of BE in case of highly variable drugs (Haidar et al., 2008; Haidar, 2006; Davit, 2006). This method suggests that if the expected intrasubject variability of a drug is less than 30% the classic 0.80–1.25 limits should

be applied. However, for drugs with CV_{WR} exceeding 30% a reference scaled average BE approach should be used. These suggestions correspond to a switching coefficient of variation of $CV_0 = 30\%$. It was proposed that three-period BE studies should be performed in which the reference product will be administered twice and the test product once, i.e., the possible sequences are TRR, RTR, and RRT (Haidar et al., 2008). The $scABE_R$ criterion is described by Eq. (3):

$$\text{Upper/Lower BE limits} = \exp\left[\pm\ln(1.25)\cdot\frac{S_{WR}}{S_{WO}}\right] \quad (3)$$

where s_{WR} is the intrasubject variability of the R product and s_{W0} is the regulatory standardized variation. It was suggested (Haidar et al., 2008) that a value of 0.25 should be assigned to s_{W0} since the latter demonstrates a good balance between a conservative approach and a practical one. However, the value of s_{W0} is lower than the switching variation (30%), and therefore the BE limits are discontinuous at the switching variation value (Endrenyi and Tothfalusi, 2007).

In addition, Haidar (2006) proposed a constraint on the point estimate for the ratio of GMR. Since large deviations of GMR could be accepted according to the $scABE_R$ approach it was suggested that GMR should be limited to the range of 0.80–1.25. It has been demonstrated that the point GMR constraint has little impact at low CV_{WR} (e.g., 30%) values and its effect becomes more significant at high CV_{WR} values (e.g., 60%).

3. Methods

3.1. Simulated BE trials

Various conditions were simulated to compare the performance of the methods proposed for the determination of BE. Three-period studies, in which the reference product is administered twice (TRR, RTR, RRT) to each of the subjects, were considered in order to simulate the conditions of the proposed semi-replicate design. Additionally, two-period, two-treatment, crossover bioequivalence studies were simulated. In all cases, an equal number of subjects were assumed to participate in each sequence. Several sample sizes (N) ranging from 16 to 54 subjects were used to simulate the conditions of both the classic 2×2 and the semi-replicate BE studies. Bioequivalence was declared if the 90% CI around the ratio of the estimated geometric means for the two drug products was between the BE limits according to the two-one sided tests procedure (Schuirmann, 1987). The BE limits used in this analysis are listed in Table 1.

It has been proposed that the approaches of the classic scaled BE limits and scaled average bioequivalence lead to similar results and each method can be converted to the other (Tothfalusi et al., 2001). Besides, the use of scaled BE limits is preferable since the relevant 90% CI can be easily calculated. In case of the $scABE_R$ approach a modified Hyslop model was proposed for statistical analysis (Haidar, 2006; Hyslop et al., 2000). However, for the purposes of the current work the $scABE_R$ method was based on scaled BE limits in accord with Eq. (2). The results derived from this approximation method were compared with those from published data (Haidar, 2006) and were found almost identical. Nevertheless, the approximation method has been also successfully applied to previous studies (Karalis et al., 2004, 2005; Kytariolos et al., 2006).

Log-normal distribution was assumed for the pharmacokinetic parameter studied. The theoretical intrasubject variability values applied to the reference product (CV_{WR}) were 15% 30% 45% and 60%. The intrasubject variability for the product under evaluation (Test) was assumed to be equal to CV_{WR}. The variability of the logarithmically transformed parameters was calculated from the preset CV_{WR} according to the formula: $S_{WR} = \sqrt{\ln(1 + CV_{WR}^2)}$. The true Geometric Mean Ratio (GMR_0) was gradually changed, from $GMR_0 = 1.00$ to 1.70 with increasing step equal to 0.05.

Ten thousand BE trials were simulated under each condition. The percentage of accepted studies was recorded and power curves were constructed by plotting the percentage of acceptance versus GMR_0. For the implementation of the simulation framework, a computer program was developed in FORTRAN. This program was validated prior to use in this study by comparing some of the simulated acceptances with previous published data (Tothfalusi et al., 2001; Tothfalusi and Endrenyi, 2003; Haidar, 2006).

3.2. Exposure

This study places particular emphasis to the total human exposure to drugs, which is expressed as the product *sample size* × *periods*. For this reason three-dimensional modified power curves were constructed by plotting the % acceptance versus GMR_0 and all levels of exposure (ranging from 48 to 108).

3.3. Mean Relative Difference (MRD%)

The BE limits examined in this work exhibit different properties; some of them may achieve high percentages of acceptance at $GMR_0 = 1$ (which is desired) but concomitantly can still show increased % power values at high GMR_0 values ($GMR_0 > 1.25$) which may lead to high consumer risk (and should be avoided). Therefore, an "ideal" BE limit should exhibit power values of 100% and 5% at GMR_0 equal to 1.00 and 1.25, respectively. A new metric, termed %Mean Relative Difference (MRD%), is introduced in order to quantitatively compare the behaviour of the BE limits with the "ideal" performance. The MRD% estimate consists of three main elements: (i) the observed % acceptances, $P_{1.00}$ and $P_{1.25}$, of the BE limit at GMR values equal to 1.00 and 1.25, respectively, (ii) the % acceptances *100* and *5* corresponding to an "ideal" BE limit at GMR equal to 1.00 and 1.25, respectively, and (iii) the relative difference between the observed and the "ideal" percent of acceptances. Therefore, MRD% is calculated according to the formula:

$$\text{MRD\%} = 100\cdot\frac{[(P_{1.00} - 100)/100 + (P_{1.25} - 5)/5]}{2} \quad (4)$$

The use of the relative differences, between the observed and the "ideal" % acceptances in Eq. (4), leads to a normalization of the differences which allows their equal participation at both GMR = 1

Table 1
Methods for the determination of bioequivalence and the design upon which they applied.

Method	Criteria	Design	Reference
BEL	Classic 0.80–1.25 bioequivalence limits	2×2	FDA (2001)
BEL_r	BEL applied to the semi-replicate design	3×3	Haidar et al. (2008)
$scABE_R$	Reference scaled average bioequivalence with $s_{W0} = 0.25$	3×3	Haidar et al. (2008)
$scABE_{Rc}$	Reference scaled average bioequivalence with $s_{W0} = 0.30$	3×3	Endrenyi and Tothfalusi (2007)
BELscW	A levelling-off scaled BE limit using a Weibull function	2×2	Karalis et al. (2005)
$BELscW_R$	BELscW using intrasubject variability of the reference formulation	3×3	Current study
BELscM	Mixed model: Unscaled BE limits up to CV_W 20% and scaled BE limits with $i = 1.116$ for CV_W >20%	2×2	Tothfalusi and Endrenyi (2003)
BELscC	Scaled BE limits ($k = 1.00$) with the additional criterion: $0.80 \leq GMR \leq 1.25$	2×2	Tothfalusi et al. (2003)

and 1.25. Also, the type of the differences in Eq. (4) permits the discrimination between a possible underestimation at GMR = 1 and an overestimation at GMR = 1.25. This feature would not be feasible if the absolute values or a different arrangement of the differences had been used in Eq. (4). As long as MRD% values get closer to zero, the overall performance of the BE limits becomes better, while negative or positive values correspond to increased producer or consumer risk, respectively.

4. Results and discussion

Fig. 1 presents power curves simulated by assuming intrasubject variabilities equal to 30, 45, and 60%, while the true GMR ratio ranges from 1.00 to 1.70. The sample size was either set to $N = 24$ or 36. Ten thousand trials were simulated under each condition. The BE limits depicted on the plot can be summarized into two categories:

A Bioequivalence limits based on the semi-replicate design which include the recently proposed $scABE_R$ approach, the classic 0.80–1.25 limits applied to the semi-replicate design (BEL_r), the $BELscW_R$ limits which scale with intrasubject variability of the R product, and a modification of the $scABE_R$ method to resolve the discontinuity of BE limits at switching variability values $scABE_{Rc}$ (Endrenyi and Tothfalusi, 2007).

B BE limits relying on the conventional 2 × 2 design which include the levelling-off BELscW limit (Karalis et al., 2005) and the classic 0.80–1.25 BE limits (BEL).

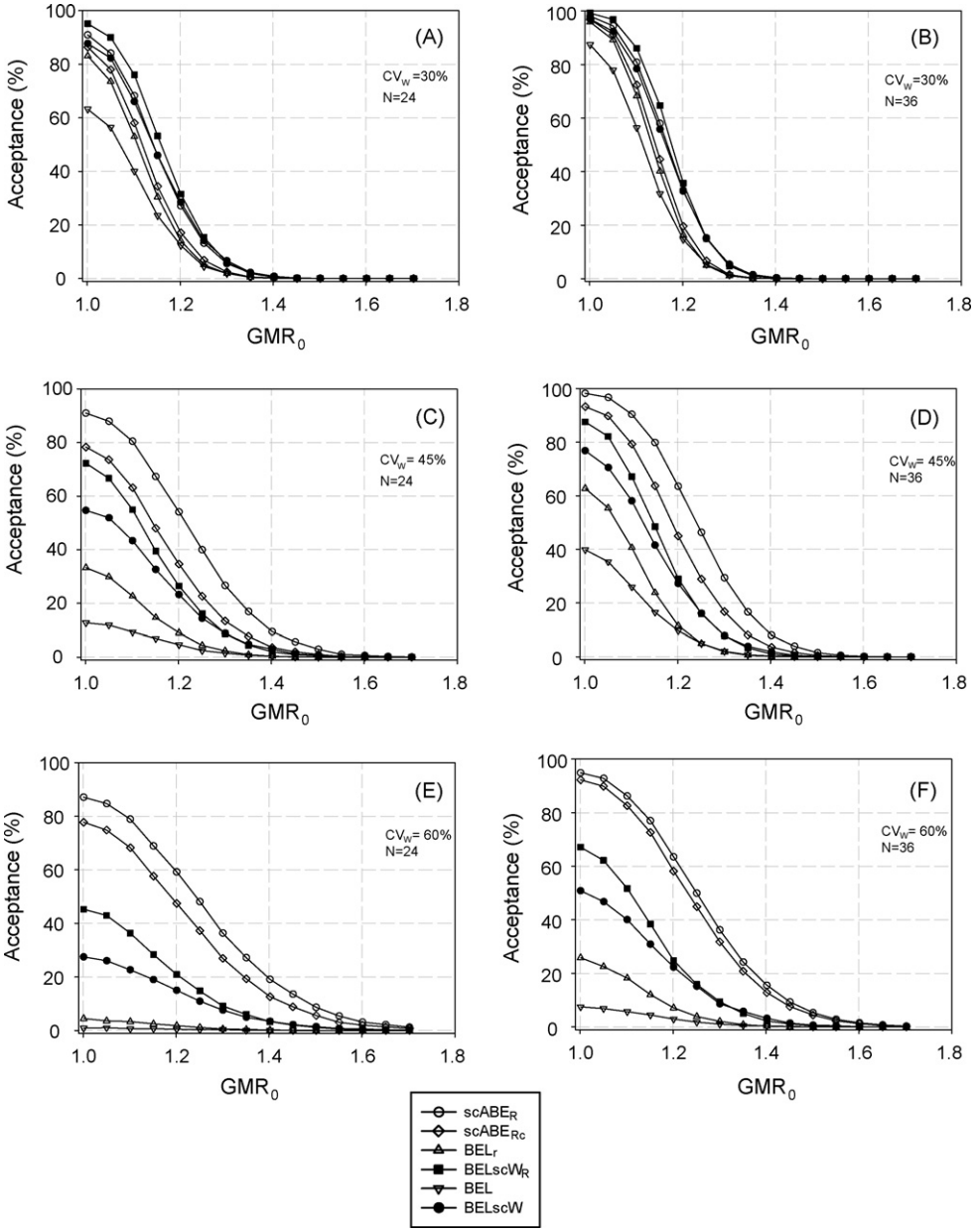

Fig. 1. Acceptance (%) of bioequivalence studies by six procedures at various ratios of true geometric mean ratio (GMR_0) assuming $N = 24$ and 36. Under each condition, a number of 10,000 studies were simulated at two levels of intrasubject variability ($CV_W = 30\%$ and $CV_W = 60\%$). Two different types of study designs were simulated: (i) Three-period design in which the reference product is administered twice (TRR, RTR, RRT) to each of the subjects, and (ii) the classic two-period, two-treatment 2 × 2 crossover design.

Assuming a sample size of 24 subjects and setting $CV_W = 30\%$ (Fig. 1A) $BELscW_R$ exhibits the highest statistical power when $GMR_0 = 1$ which is followed by $scABE_R$, BELscW, scABERc, and BEL_r. As expected the classic BEL shows the lowest percentage of acceptance. As GMR_0 values increase these relative differences among the limits are retained. When sample size is increased to $N = 36$ (Fig. 1B) the percent acceptance of BE studies rises for all BE limits, but the relative performance of the BE limits remains the same.

For higher CV_W values, 45% and 60% (Fig. 1C–F), the $scABE_R$ and $scABE_{Rc}$ limits show a much greater probability of declaring bioequivalence when GMR = 1 in comparison to the two levelling-off scaled BE limits (BELscW, $BELscW_R$) and the two classic limits (BEL and BEL_r). Overall, $scABE_R$ followed by $scABE_{Rc}$ are the most permissive BE limits and they seem to be extremely permissive even at high GMR_0 values (greater than 1.25). The s_{W0} value used for $scABE_R$ is 0.25, while the s_{W0} value for $scABE_{Rc}$ is 0.30; this makes $scABE_R$ more permissive than $scABE_{Rc}$ (Fig. 1). Regarding the two levelling-off limits, $BELscW_R$ appears more permissive than BELscW, while both show an intermediate behaviour comparing to the two $scABE_R$ and the two classic BE limits.

It should be highlighted that even though $scABE_R$ and $scABE_{Rc}$ exhibit desirable behaviour when the two drug products are truly bioequivalent (i.e., at $GMR_0 = 1$), they are extremely permissive at high GMR values, i.e., in cases where the product under evaluation differs by more than 25% from the brand-name product. For example, assume the case of a highly variable drug with an intrasubject variability of 60% and 36 subjects recruited in the BE study (Fig. 1F). If the two drug products were truly bioequivalent ($GMR_0 = 1$), the probability of declaring bioequivalence would be equal to 95%, 92%, 26%, 67%, 8% and 51% for $scABE_R$, $scABE_{Rc}$, BEL_r, $BELscW_R$, BEL, and BELscW, respectively. Under these conditions, the use of the classic BE limits (0.80–1.25) yields very low acceptances when applied either to the 2×2 or the semi-replicate design. BELscW and $BELscW_R$ exhibit much better performances, while both $scABE_R$ and $scABE_{Rc}$ achieve in showing an ideal behaviour with very high power values. However, when the two drug products differ by 25% the percent of accepted studies is still high for $scABE_R$ and $scABE_{Rc}$ (power values range from 45% to 50%). On the contrary, the power estimates for BELscW and $BELscW_R$ are much lower and close to 15%.

Overall, the use of the semi-replicate design with the BE limits $scABE_R$ and scABERc leads to low producer risk estimates which are much better than the values derived from the classic 2×2 design. However, a non-desired feature of the $scABE_R$ and $scABE_{Rc}$ approaches is the fact that they are still very permissive even at very high GMR values.

The aforementioned comparison does not take into account the exposure of volunteers to drugs. Although the number of subjects participating in each study is the same between the semi-replicate and the 2×2 design, the total drug exposure is different. Assuming N subjects in each period, the partial replicate 3-way crossover design leads to 3^*N human exposure, while the 2×2 design to a 2^*N exposure. In other words, a more fair comparison of the performance of the BE limits should rely on the same extent of human exposure to drugs.

Fig. 2 illustrates the percentage of accepted BE studies as a function of GMR ratio for various levels of exposure. Exposure is expressed as the product of sample size by the number of periods of drug administration. Therefore, these three-dimensional plots allow a direct comparison of the performance of the BE limits for several exposure values and GMR ratios. Two-period crossover simulated studies were performed assuming sample sizes from 24 to 54 subjects and four levels of intrasubject variability: 15%, 30%, 45%, and 60%. For the semi-replicate design sample sizes ranging from 16 to 36 were considered. In both cases, the derived exposure estimates were from 48 to 108 *human* × *drug administrations*. Since the performance of $scABE_R$ was very similar to the behaviour of $scABE_{Rc}$ and both of them are relying on the semi-replicate design, the latter was omitted from subsequent analysis for reasons of simplicity.

A general worth noting and merely expected feature is the fact that as exposure increases the % acceptance observed for all BE limits increase but in a different manner (Fig. 2). The % acceptance for some BE limits stays almost stable, other show a smooth increase, while in other cases a more steep rise becomes evident. This finding is attributed to the different properties of each BE limit. Scaled BE limits tend to increase continuously with intrasubject variability, while levelling-off limits are extended until a maximum plateau value. These inherent properties of the BE limits is the underlying reason of the different response of the BE limits to the exposure increase.

At low CV_W values ($CV_W = 15\%$) all methods exhibit almost identical performance irrespective of the exposure and the GMR value (Fig. 2). Both at $GMR_0 = 1$ and 1.25 all methods show a desired behaviour. For the borderline CV_W value of 30%, $BELscW_R$ and BELscW show the highest percent acceptances when the two drug products are truly bioequivalent. This feature is more evident at low exposure values. As exposure increases the power values of $scABE_R$, BEL_r, and BEL gradually increase and tend to be similar to those of $BELscW_R$ and BELscW. It is noteworthy to mention that the two levelling-off BE limits ($BELscW_R$ and BELscW) exhibit almost identical performance even though the first is based on a partial replicate 3-way crossover design and the second on the classic 2×2 crossover design.

When intrasubject variability was set equal to 45% three main groups of performance were distinguished (Fig. 2). Method $scABE_R$ exhibits the highest percentage of acceptance both at $GMR_0 = 1$ and 1.25. While the first property is desired, the fact that $scABE_R$ results in high power values when two drugs are different represents a major concern regarding the consumer risk. The classic BE limits (BEL and BEL_r) show an almost identical behaviour; low % acceptances (equal to 5%) when $GMR_0 = 1.25$, and low probability of declaring bioequivalence at $GMR_0 = 1$ which tends to increase steeply as exposure values rise. Besides, the two levelling-off BE limits exhibit an intermediate behaviour. Both of them are adequately permissive (power values are close to 90%) when the two drug products are truly bioequivalent, while at high GMR values both $BELscW_R$ and BELscW appear significantly less permissive than $scABE_R$.

As intrasubject variability becomes higher ($CV_W = 60\%$) the aforementioned properties of the BE limits are still present but the differences between them become more potent (Fig. 2). Three groups of performance can also be distinguished. Method $scABE_R$ is the most permissive BE limit at all GMR_0 values. Even though this is a desired feature for truly bioequivalent drug products, the high statistical power at GMR_0 values greater than 1.25 leads to concerns for the consumer risk. On the contrary, BEL and BEL_r illustrate low percentages of acceptance for all GMR_0 values. Besides, $BELscW_R$ and BELscW show an intermediate behaviour; at $GMR_0 = 1$ these limits are adequately high, while at $GMR_0 = 1.25$ both $BELscW_R$ and BELscW are significantly less permissive than $scABE_R$.

Fig. 3 shows the percentage of studies in which BE is declared as a function of exposure and GMR_0 by applying two different scaled BE limits: the mixed method (BELscM) and the approach with a "constraint" on GMR, BELscC (Tothfalusi and Endrenyi, 2003; Tothfalusi et al., 2003). For reasons of comparison the classic 0.80–1.25 limit and BELscW were also included. In all cases, the classic 2×2 crossover design was applied. Al low CV_W values ($CV_W = 15\%$) BEL, BELscW, and BELscM show similar performance for all GMR_0 and exposure levels. BELscC is also highly permissive when the two drug

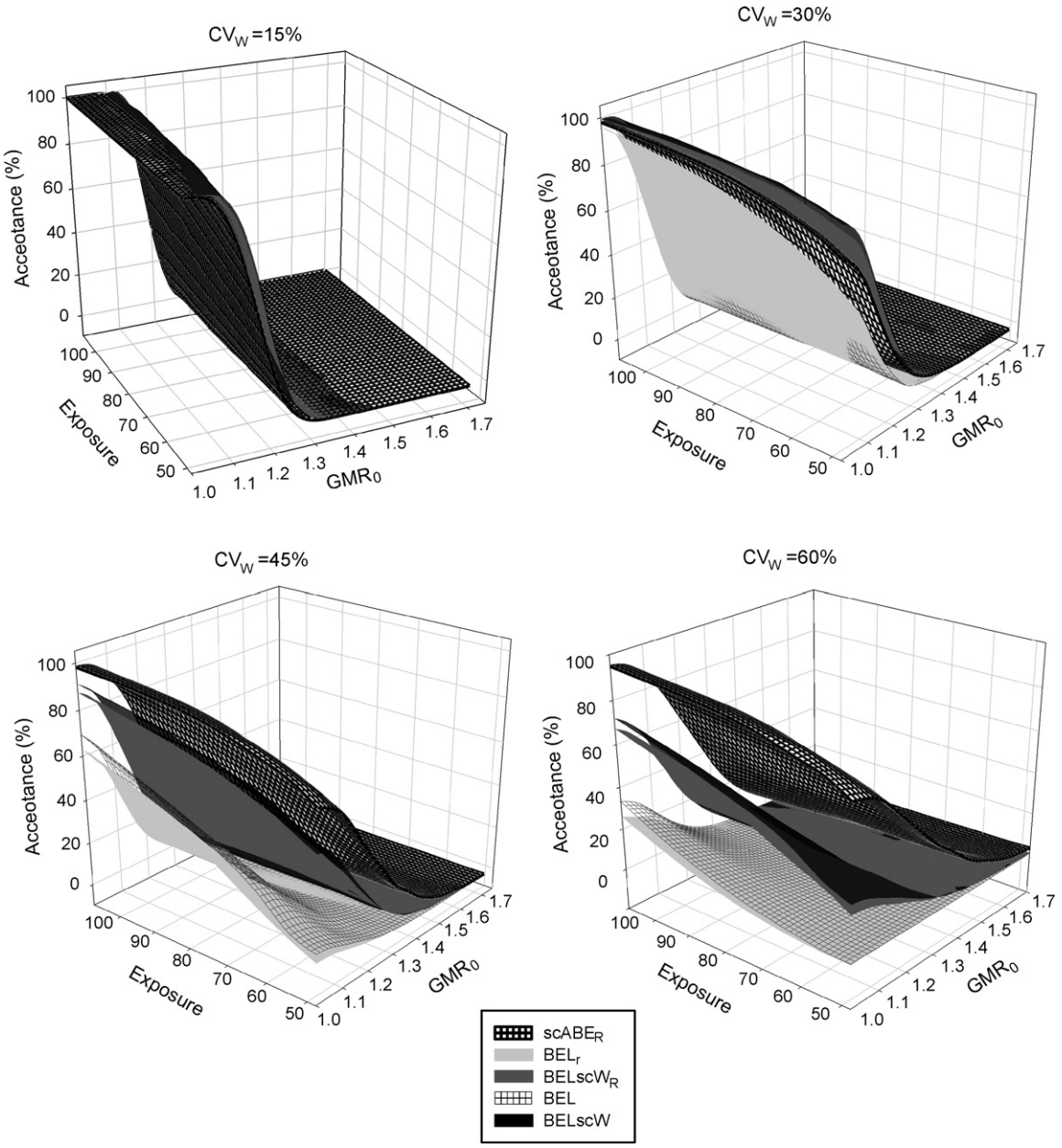

Fig. 2. Three-dimensional power curves illustrating the acceptance (%) of bioequivalence studies by five procedures at various ratios of true geometric mean ratio (GMR_0) and *Exposure* values ranging from 48 to 108 human × periods. Under each condition, a number of 10,000 studies were simulated at four levels of intrasubject variability (CV_W equal to 15%, 30%, 45%, and 60%). Two different types of study designs were simulated: (i) Three-period design in which the reference product is administered twice (TRR, RTR, RRT) to each of the subjects, and (ii) the classic two-period, two-treatment 2 × 2 crossover design. See Table 1 for the criteria and the design of each one of the methods used.

products are truly bioequivalent. However, a much steeper decline in percent acceptance becomes apparent for BELscC when GMR_0 deviates from unity.

When intrasubject variability was set equal to 30% BELscM was the more permissive followed by BELscC, BELscW, and finally by BEL (Fig. 3). This relative ranking was obvious for all GMR_0 and exposure values. At a higher CV_W level (CV_W = 45%), the pre-referred ranking was preserved but the differences in performance among the BE limits became more evident. Similar findings were also observed when intrasubject variability was set equal to 60%. However, it should be highlighted that firstly BELscM and secondly BELscC tend to decline very smoothly as GMR_0 increases. This leads to very high percent of acceptances at high GMR values especially for BELscM and in a lower degree for BELscC.

Overall, methods like $scABE_R$, BELscM, and BELscC appear very permissive at all GMR_0 values (Fig. 3). Other approaches, such as the classic 0.80–1.25 limits applied to the 2 × 2 and partial replicate 3-way crossover designs tend to be too strict for declaring bioequivalence. Finally, the levelling-off limits exhibit a more conservative behaviour which is characterized by an adequately high percent of acceptance at GMR_0 = 1 and power values much lower than $scABE_R$ and the other scaled BE limits when GMR_0 is greater than 1.25.

In order to quantify the deviations from the ideal behaviour of the various approaches the concept of percent Mean Relative Difference (MRD%, Eq. (4)) in BE studies is introduced. MRD% is a simple index which indicates how close to the ideal behaviour an approach lies. The ideal value for MRD% is zero and corresponds

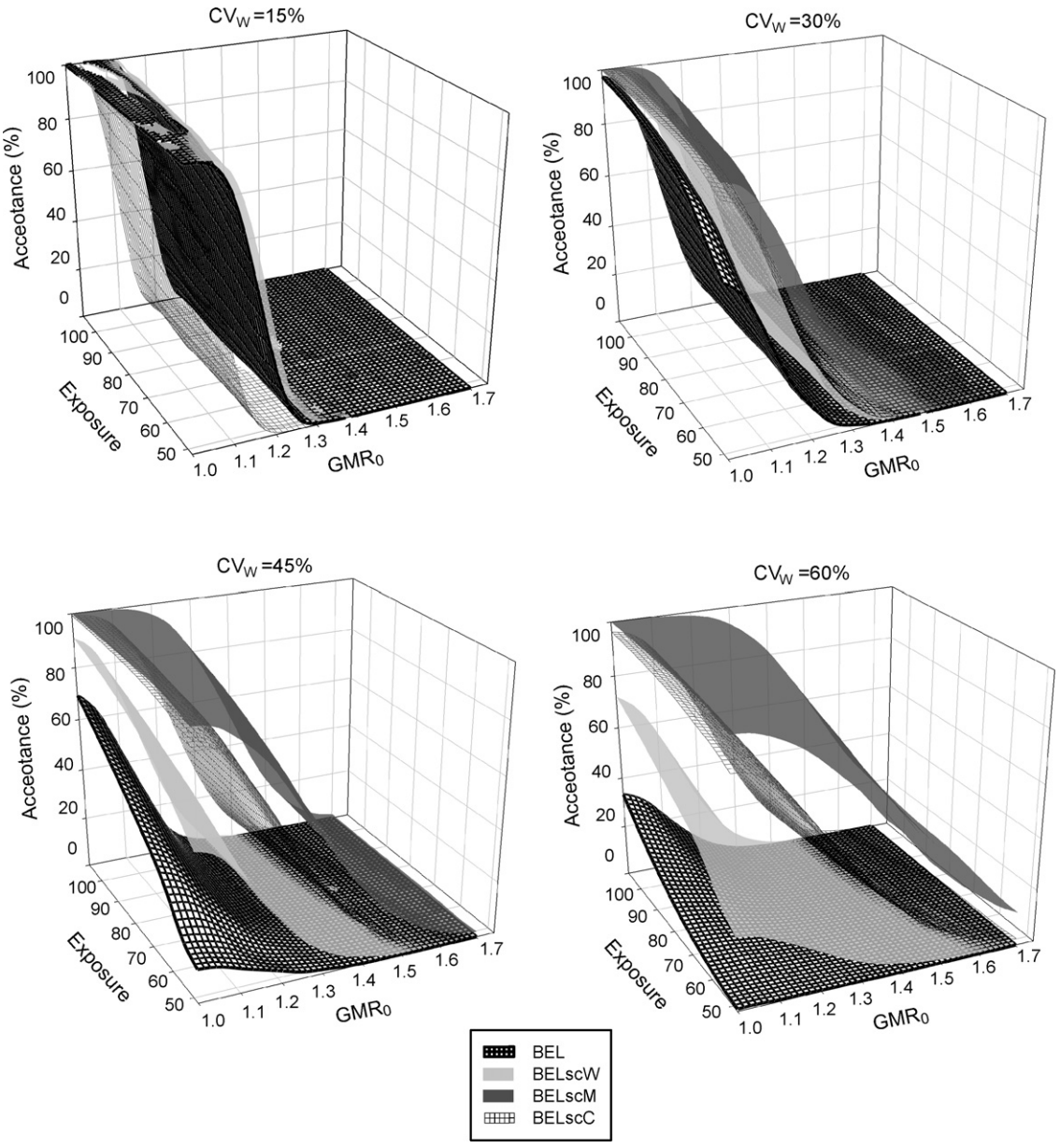

Fig. 3. Three-dimensional power curves illustrating the acceptance (%) of bioequivalence studies at various ratios of true geometric mean ratio (GMR_0) and *Exposure* values ranging from 48 to 108 *human* × *periods*. Under each condition, a number of 10,000 two-period, two-treatment crossover design BE studies were simulated at four levels of intrasubject variability (CV_W equal to 15%, 30%, 45%, and 60%). Four different BE limits are presented: the classic 0.80–1.25 BEL, the BELscW, the "mixed" criterion (BELscM), and the "constrained" approach (BELscC).

to an approach which Exhibits 100% acceptance when the two drug products are truly bioequivalent ($GMR_0 = 1$) and 5% acceptance when GMR_0 is equal to 1.25. As MRD% values are deviating from zero, to negative or positive values, the performance of the BE limit is becoming poorer.

For the BE limits shown in Figs. 2 and 3, MRD% values were calculated from Eq. (4). Fig. 4 is a three dimensional plot displaying the MRD% values as a function of exposure and intrasubject variability. Visual inspection of Fig. 4 reveals that the recently proposed $scABE_R$ limit exhibits a low MRD% value when both exposure and CV_W are low. However, there is a rapid increase with exposure and CV_W which leads to very high MRD% values. This can be explained by the fact that $scABE_R$ shows high % acceptances when $GMR_0 = 1$ which leads to desired MRD% values, but $scABE_R$ retains its permissiveness at high GMR_0 levels, instead of being close to 5%, resulting in a MRD% increase.

The classic 0.80–1.25 BE limits (BEL and BEL_r) exhibit an ideal performance when intrasubject variability is low (particularly when CV_W is lower than 30%) and exposure is high. However, as intrasubject variability rises and exposure diminishes both BEL and BEL_r tend to reach a MRD% value of −100%. This finding originates from the inability of BE limits to prove bioequivalence when two drugs are truly bioequivalent in case of highly variable drugs.

The opposite behaviour was observed for the two levelling-off BE limits (BELscW, $BELscW_R$). For high intrasubject variability and low exposure values $BELscW_R$ and BELscW resulted in MRD% values very close to zero. As exposure increases and variability declines the calculated MRD% value of the levelling-off limits reaches a value of 100%, i.e., exactly the opposite value of that observed for BEL_r and BEL for low exposure and high variability.

It should be mentioned (data not shown) that the performance of the other scaled BE limits is similar to the $scABE_R$, i.e., they exhibit

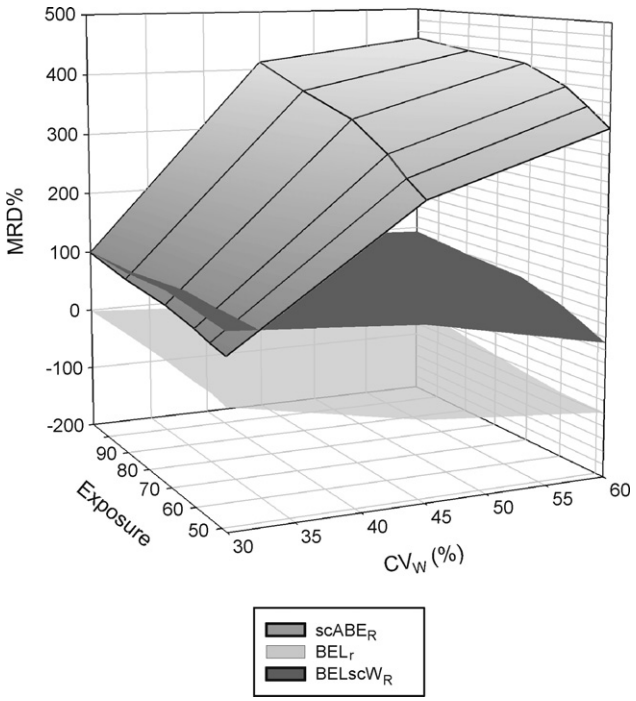

Fig. 4. Three-dimensional plots of the %Mean Relative Difference (MRD%) versus *Exposure* and intrasubject variability (CV_W) for three different approaches: $scABE_R$, BEL_r, and $BELscW_R$. Plots for BEL and BELscW were almost identical to BEL_r and $BELscW_R$, respectively, and were omitted for simplicity reasons. See Table 1 for the criteria and the design of each one of the methods used.

an ideal MRD% value at GMR = 1 but as GMR_0 increases the decline of the power is rather smooth which leads to high percentage of acceptances and large MRD% values.

A requirement of bioequivalence studies is to use an adequate number of subjects to achieve the necessary statistical power and concomitantly to not unnecessarily expose humans to drugs. In case of highly variable drugs statistical power diminishes due to the increased intrasubject variability. A possible solution to this problem is the use of replicate designs. Recently, scientists from FDA proposed the use of a semi-replicate design where the reference product is administered twice and the product under evaluation once. This design allows the estimation of intrasubject variability (CV_{WR}) of the reference product which is further used to estimate scaled with CV_{WR} average BE limits. According to the proposed method high statistical power values are derived for all geometric mean ratios of the two drug products. For high intrasubject variability values the % acceptance, after applying the proposed $scABE_R$ method, is always greater than any other method based on the classic 2×2 design with the same number of subjects as in the semi-replicate design.

However, attention should be paid not only to the number of subjects but also to the total human exposure to drugs. The proposed semi-replicate design includes more periods of drug administration which result in an increased exposure of humans to drugs. However, comparisons between different methods should take into account the "exposure" factor and any conclusions regarding the suitability of a method should be made after adjusting the same exposure among the methods.

In this vein, several BE limits were compared in the current analysis. Overall, it was concluded that the use of the classic 0.80–1.25 BE limits is favoured when intrasubject variability is low and exposure is high, i.e., many subjects are recruited. On the other hand, levelling-off limits succeed in the opposite conditions; they demonstrate the best overall performance when variability is high and a limited number of subjects or drug administrations are used. The other scaled BE limits, including $scABE_R$, BELscM, BELscC, exhibit almost ideal properties with high percentages of acceptance when the two drug products under comparison are similar. However, as intrasubject variability increases and the two drug products under comparison do differ, very high MRD% values are obtained and their performances become poor. This finding arises from the inherent properties of scaled BE limits to exhibit high percentages of acceptance when the two drug products differ significantly. The levelling-off limits may not show the best performance at all GMR values, however, they exhibit the most promising overall behaviour in case of highly variable drugs.

Finally, it is important to mention that the application of the reference scaled average bioequivalence approach apart from the application of the 3-period design requires also to: (i) pre-define the switching variability value (CV_0), (ii) set the regulatory standardized variation ($s_{W0} = 0.25$), and (iii) apply the 0.80–1.25 point GMR constraint. Levelling-off scaled BE limits (such as BELscW) encompass all (i)–(iii) properties into a single criterion. In addition, the performance of the levelling-off limits is almost identical when exposure is the same, irrespectively if they are applied to a replicate or a simple 2×2 design. The latter allows the application of simple crossover designs which do not demonstrate the drawbacks of the replicate designs such as the recruitment of each subject for a long-time period, the high drop-out rate, and the increased risk of adverse events.

References

Benet, L., 1995. Bioavailability and bioequivalence: definitions and difficulties in acceptance criteria. In: Midha, K.K., Blume, H.H. (Eds.), Bio-International: Bioavailability, Bioequivalence and Pharmacokinetics. Medpharm, Stuttgart, pp. 27–35.

Blume, H., Elze, M., Potthast, H., Schug, B., 1995a. Practical strategies and design advantages in highly variable drug studies: multiple dose and replicate administration design. In: Blume, H.H., Midha, K. (Eds.), Bio-international' 92: Bioavailability, Bioequivalence, and Pharmacokinetic Studies. Medpharm, Stuttgart, pp. 117–122.

Blume, H.H., Midha, K.K., 1993. Bio- International 92, Conference on Bioavailability Bioequivalence and Pharmacokinetic Studies. J. Pharm. Sci. 11, 1186–1189.

Blume, H.H., McGilveray, I.J., Midha, K.K., 1995b. Bio-International 94, conference on bioavailability bioequivalence and pharmacokinetic studies. Eur. J. Pharm. Sci. 3, 113–124.

Boddy, A., Snikeris, F., Kringle, R., Wei, G., Oppermann, J., Midha, K., 1995. An approach for widening the bioequivalence acceptance limits in the case of highly variable drugs. Pharm. Res. 12, 1865–1868.

Davit, B.M., 2006. Highly variable drugs—bioequivalence issues: FDA proposal under consideration. Meeting of FDA Committee for Pharmaceutical Science, Rockville, MD.

Endrenyi, L., Tothfalusi, L., 2007. Determination of Bioequivalence for Highly-variable Drugs. AAPS Annual Meeting Current Issues and Advances in the Determination of Bioequivalence, San Diego, CA.

European Medicines Agency (EMEA), 2001. Note for Guidance on the Investigation of Bioavailability and Bioequivalence. Committee for Proprietary Medicinal Products (CPMP), London.

European Medicines Agency (EMEA), 2006. Committee for Medicinal Products for Human Use (CHMP). Concept Paper for an Addendum to the Note for Guidance on the Investigation of Bioavailability and Bioequivalence: Evaluation of Bioequivalence of Highly Variable Drugs and Drug Products. London.

Food and Drug Administration, 2003. Guidance for industry. Bioavailability and Bioequivalence Studies for Orally Administered Drug Products—General Considerations.

Food and Drug Administration. 2001. Statistical Approaches to Establishing Bioequivalence. Center for Drug Evaluation and Research (CDER), Rockville, MD.

Haidar, S.H., 2006. Evaluation of the scaling approach for highly variable drugs. In: Meeting of FDA Committee for Pharmaceutical Science, Rockville, MD.

Haidar, S.H., Davit, B., Chen, M.L., Conner, D., Lee, L., Li, Q.H., Lionberger, R., Makhlouf, F., Patel, D., Schuirmann, D.J., Yu, L.X., 2008. Bioequivalence approaches for highly variable drugs and drug products. Pharm. Res. 25, 237–241.

Hyslop, T., Hsuan, F., Holder, D., 2000. A small sample confidence interval approach to assess individual bioequivalence. Stat. Med. 19, 2885–2897.

Karalis, V., Macheras, P., Symillides, M., 2005. Geometric mean ratio dependent scaled bioequivalence limits with levelling-off properties. Eur. J. Pharm. Sci. 26, 54–61.

Karalis, V., Symillides, M., Macheras, P., 2004. Novel scaled average bioequivalence limits based on GMR and variability considerations. Pharm. Res. 21, 1933–1942.

Kytariolos, J., Karalis, V., Macheras, P., Symillides, M., 2006. Novel scaled bioequivalence limits with levelling-off properties. Pharm. Res. 23, 2657–2664.

Midha, K., Rawson, M., Hubbard, J., 1998. Bioequivalence: switchability and scaling. Eur. J. Pharm. Sci. 6 (87), Y91.

Midha, K.K., Shah, V.P., Singh, G.J., Patnaik, R., 2007. Conference report: Bio-International 2005. J. Pharm. Sci. 96, 747–754.

Schuirmann, D., 1987. A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. J. Pharmacokin. Biopharm. 15, 657–680.

Shah, V., Yacobi, A., Barr, W., Benet, L., Breimer, D., Dobrinska, M., Endrenyi, L., Fairweather, W., Gillespie, W., Gonzales, M., Hooper, J., Jackson, A., Lesko, L., Midha, K., Noonan, P., Patnaik, R., Williams, R., 1996. Absorption of orally administered highly variable drugs and drug formulations. Pharm. Res. 13, 1590–1594.

Tothfalusi, L., Endrenyi, L., 2003. Limits for the scaled average bioequivalence of highly variable drugs and drug products. Pharm. Res. 20, 382–389.

Tothfalusi, L., Endrenyi, L., Midha, K., 2003. Scaling or wider bioequivalence limits for highly variable drugs and for the special case of C_{max}. Int. J. Clin. Pharmacol. Ther. 41, 217–225.

Tothfalusi, L., Endrenyi, L., Midha, K., Rawson, M., Hubbard, J., 2001. Evaluation of the bioequivalence of highly-variable drugs and drug products. Pharm. Res. 18, 728–733.