European Journal of Pharmaceutical Sciences 44 (2011) 497–505

Contents lists available at SciVerse ScienceDirect

European Journal of Pharmaceutical Sciences

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

On the leveling-off properties of the new bioequivalence limits for highly variable drugs of the EMA guideline

Vangelis Karalis, Mira Symillides *, Panos Macheras

Laboratory of Biopharmaceutics-Pharmacokinetics, School of Pharmacy, University of Athens, Athens, Greece

ARTICLE INFO

Article history:
Received 22 July 2011
Accepted 9 September 2011
Available online 16 September 2011

Keywords:
Bioequivalence
Highly variable drugs
Leveling-off limits
Scaled limits
European Medicines Agency guideline
Simulated studies

ABSTRACT

Recently, the European Medicines Agency (EMA) issued a new guideline on the investigation of bioequivalence (BE). In case of highly variable drugs, this guideline proposes that the acceptance limits for C_{max} can gradually be expanded as a function of within-subject variability (CV_{wR}). Actually, these BE limits exhibit *leveling-off* properties since they are not allowed to scale continuously, but only up to CV_{wR} = 50%. To avoid the risk of accepting two drug products which may differ significantly, this EMA guideline also proposes the use of a secondary constraint criterion on the geometric mean ratio (GMR) of the two products under comparison. Aim of this study was to explore the *leveling-off* properties of the new EMA limits in comparison to other approaches, as well as to assess the impact of the complementary GMR criterion on the ability to declare bioequivalence. Simulated bioequivalence studies and extreme GMR plots were used to assess the performance of the EMA limits. Three sequence, three period (3×3) crossover studies with two treatments (T and R) were simulated. The R product was considered to be administered twice, while the T only once (i.e., TRR/RTR/RRT). Among others, this study revealed the *leveling-off* properties of the new EMA limits. It was also shown that the complementary GMR-constraint is only effective when a large sample size is used and at regions of CV_{wR} close to 50%. This GMR-criterion begins to be effective at sample sizes around 60 and becomes more prominent as the number of subjects participating in the BE study increases. For CV_{wR} values lower than 50%, the GMR-constraint has no role. In case of within-subject variabilities greater than 50%, the impact of the GMR-constraint diminishes due to the *leveling-off* properties of the EMA limits. Compared to the classic 0.80–1.25 or the extended 0.75–1.33 criteria, the new EMA limits are more liberal at high CV_{wR} values and allow greater differences between the two drug products to be declared bioequivalent. Finally, this study showed that the use of an *approximate* value (0.760) on the scaling factor proposed by EMA, has no impact on the performance of the new BE limits compared to other more accurate approaches.

1. Introduction

It is widely known that the efficacy of drug products depends on the *extent* (amount) and the *rate* according to which the active moiety is absorbed. These two key terms, *extent* and *rate* of absorption, constitute the basis of bioequivalence (BE) testing (FDA, 2003). Bioequivalence assessment simply means a comparison between two drug products; a product under evaluation (test, T) *versus* an innovator's product (reference, R). Accordingly, the T and R preparations, which contain the same active substance, are considered bioequivalent if their *rate* and *extent* of absorption are so similar, thus, excluding any differences in the *in vivo* performance (EMA, 2010).

However, determination of bioequivalence becomes a complicated issue in case of highly variable drugs (HVD), namely, drugs or drug products which are characterized by a within-subject variability value greater than 30% (Blume and Midha, 1993; Blume et al., 1995; Midha et al., 2007; Shah et al., 1996; Van Peer, 2010). For the purposes of this article no distinctions will be made between a highly variable drug and a highly variable drug product. In case of HVD the risk of erroneously reject bioequivalence between two drugs (producer risk) becomes relatively high. In order to face-off this problem and, therefore, increase the statistical power of the BE study, the common practice is the recruitment of a large number of subjects. However, this approach raises many ethical and financial concerns (Benet, 1995).

In order to overcome the need of an increased sample size, several methodologies have been proposed. These approaches include the widening of BE limits to pre-fixed constant values (EMA, 2010; Hauck et al., 2001; Tothfalusi et al., 2003), the use of multiple-dose studies (Blume et al., 1995), the inclusion of individual or population

* Corresponding author. Address: School of Pharmacy, University of Athens, Panepistimiopolis, Athens 15771, Greece. Tel.: +30 2107274675; fax: +30 2107 274027.
E-mail address: simillidou@pharm.uoa.gr (M. Symillides).

0928-0987/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.ejps.2011.09.008

Abbreviations

AUC	area under the curve
BE	bioequivalence
CL	classic limits of bioequivalence: 0.80–1.25
C_{max}	peak plasma concentration
CV_{wR}	coefficient of variation corresponding to s^2_{wR}
EMA_{c1}	more accurate EMA limits (of type 1)
EMA_{c2}	more accurate EMA limits (of type 2)
EMA_{nc}	modified EMA limits without the GMR-constraint
EMA_s	the composite scaled approach proposed by EMA
Ext	extended (0.75-1.33) limits of bioequivalence
Gc	the approach where the GMR constraint is used as the sole criterion of bioequivalence
GMR	geometric mean ratio
HVD	highly variable drugs
k	scaling factor of the limits proposed by EMA
LO	*leveling-off* bioequivalence limits
LO_g	modified *leveling-off* limits with a GMR-constraint
R	reference product (i.e., innovator's product)
s^2_{wR}	within-subject variability of the reference product
s_{w0}	a constant referring to the inflection point of the LO limits
s_{wR}	standard deviation corresponding to s^2_{wR}
T	test product (i.e., product under evaluation)

bioequivalence criteria (Anderson and Hauck, 1990; Endrenyi et al., 1998; FDA, 2001; Midha et al., 1997; Patnaik et al., 1997; Schall and Luus, 1993), as well as the application of scaled average BE approaches (Boddy et al., 1995; Midha et al., 1998; Tothfalusi and Endrenyi, 2003; Tothfalusi et al., 2003).

More recently, novel scaled bioequivalence approaches were proposed (Karalis et al., 2004, 2005; Kytariolos et al., 2006). According to these approaches, the BE limits scale with within-subject variability and act as an *all-in-one* criterion, i.e., limits that can be used in any case regardless of the variability of the BE study. The basic feature of these limits is their *"leveling-off"* property. This means that they scale with within-subject variability, but only between a basal value and an extreme *plateau* value. An advantage of these *leveling-off* (LO) limits is their continuous nature, since no switching criteria (e.g., a 30% value in within-subject variability) are required. The LO limits are based on appropriate functions which provide a smooth widening of the bioequivalence limits with the increase of variability.

In 2010, the European Medicines Agency (EMA) issued a new guideline on the investigation of bioequivalence (EMA, 2010). In this guideline, the classic methodology for the assessment of BE using the 0.80–1.25 limits is conserved. However, this guideline brings to the light some new possibilities. Among others, the EMA 2010 guideline revives the application of replicated clinical designs, proposes a two-stage approach in BE testing, and introduces a new methodology for the assessment of BE in case of highly variable drugs or drug products (Morais and Lobato Mdo, 2010). For the area under the curve (*AUC*), the use of the classic 0.80–1.25 limits remains the only possibility; however, a scaled approach can be used for the assessment of peak plasma concentration (C_{max}).

EMA proposes the application of a replicate cross-over design which allows the estimation of within-subject variability of the R (s^2_{wR}) formulation (EMA, 2010). Obviously, knowledge of s^2_{wR} allows the estimation of the corresponding standard deviation (s_{wR}) and the coefficient of variation (CV_{wR}). Particularly, the EMA guideline suggests that for drugs with CV_{wR} values greater than 30%, the acceptance limits for C_{max} can gradually be expanded, as a function of CV_{wR}, to a maximum range of 0.6984–1.4319. Finally, EMA introduces a constraint on the point estimate of the geometric mean ratio (GMR) in the region 0.80–1.25. In other words, the proposed EMA bioequivalence limits exhibit *leveling-off* properties similar to the earlier proposed LO limits (Karalis et al., 2005; Kytariolos et al., 2006).

Aim of this study is (i) to highlight the *leveling-off* properties of the newly proposed EMA limits, (ii) to examine the performance of the new EMA limits in comparison to the approaches, such as the ones proposed in the previous guidelines (EMA, 2006, 2010), (iii) to explore the impact of the complementary GMR-constraint criterion for the assessment of BE, and (iv) to focus on the approximate values applied to the EMA bioequivalence and their possible implications.

2. Methods and materials

2.1. Background

2.1.1. Classic bioequivalence approach

Over the past years, the concept of average bioequivalence dominated in the field of bioequivalence testing (EMA, 2001, 2006, 2010; FDA, 2001). According to this approach, two drug products are considered bioequivalent if the calculated 90% confidence interval for the difference of their log-transformed mean measures of bioavailability (i.e., m_T and m_R for the T and R formulation, respectively) lies between preset limits (EMA, 2010; FDA, 2001, 2003). This definition is mathematically expressed by Eq. (1):

$$-\theta \le m_T - m_R \le \theta \tag{1}$$

where θ is usually set equal to ln(1.25).

This classic BE approach was widely used and it is still used in bioequivalence studies. However, it becomes problematic in case of highly variable drugs.

2.1.2. The new BE scaled approach of the EMA guideline

In order to resolve the problem of high variability encountered in BE studies, the newly proposed EMA guideline offers the opportunity to use a mixed scaled approach (EMA_s). It is suggested that a replicate crossover design (3- or 4-periods) could be applied where, at least, the R product is administered twice (EMA, 2010). The T preparation can either be administered once or twice. The crucial issue, however, is the replicate administration of R which allows the estimation of within-subject variability of R (i.e., (s^2_{wR}). The latter can be used to construct the scaled BE limits for C_{max}:

$$\text{Upper/Lower BE limit} = \exp(\pm k \cdot s_{wR}) \tag{2}$$

where *k* is a scaling factor set by the regulatory authorities equal to 0.760. According to the guideline, the scaled BE limits should only be used in cases where the CV_{wR} values are between 30% and 50%. The boundary value of 50% sets the extreme BE limit value equal to 1.4319 (or 0.6984), Fig. 1A. However, EMA defines the switching criterion with CV_{wR}, while scaling of the BE limits is done through s_{wR}. These two terms are related with the mathematical formula: $CV_{wR} = \sqrt{e^{s^2_{wR}} - 1}$. It should be underlined that the increasing (or decreasing) part of EMA_s limits (Fig. 1A) is not linear but follows an exponential rise (see Eq. (2)).

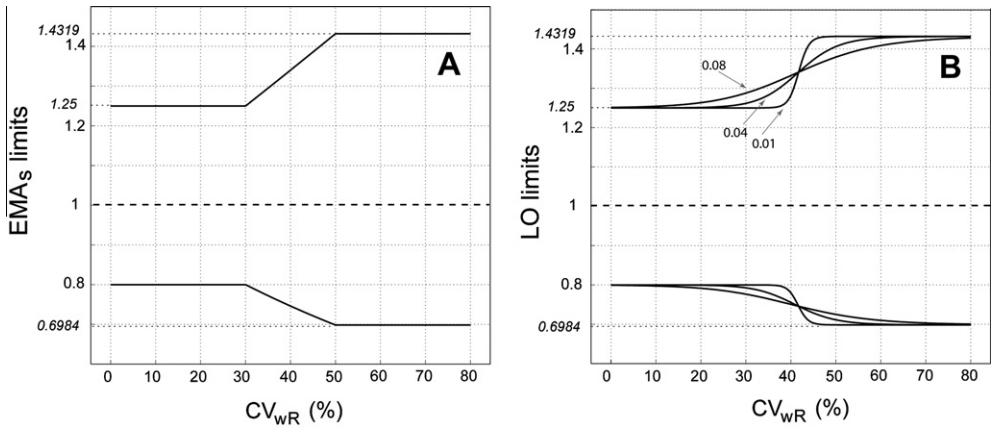

Fig. 1. EMA_s (A) and LO (B) limits as a function of within-subject variability. For the LO limits three different values for the shape parameter γ (Eq. (4)) are depicted: 0.01, 0.04, and 0.08.

In addition, the EMA guideline suggests the inclusion of a secondary criterion i.e., the point GMR estimate of the BE study should lie within the 0.80–1.25 range. This constraint is used to avoid the risk of accepting two drug products which in fact differ significantly in their GMR values. It is worthy to mention that for *AUC* no scaling is allowed and the classic 0.80–1.25 limits should always be applied regardless of the level of variability.

2.1.3. Leveling-off BE limits

The so-called "*leveling-off*" approach was recently proposed to improve the performance of simple scaled BE limits (Karalis et al., 2005; Kytariolos et al., 2006). The LO limits were defined to scale with within-subject variability, but only until an extreme value (Fig. 1B). Several functions were used to achieve the gradual widening of BE limits to the plateau value, such as the sigmoid and the Weibull functions. Adjusting appropriately all these functions, one can get similar results. Eq. (3) presents the LO limits (the upper part) in case of the sigmoid function (Kytariolos et al., 2006):

$$\text{Upper LO limit} = \alpha + \frac{\beta - \alpha}{1 + e^{-\left(\frac{CV - CV_0}{\gamma}\right)}} \quad (3)$$

where γ is a parameter controlling the rate of gradual expansion, CV_0 is the inflection point of the curve, and α, β refer to the basal and maximum value of the upper BE limit, respectively. Finally, the term *CV* corresponds to the within-subject coefficient of variation of the study.

Visual inspection of Fig. 1A and 1B reveals the similarity of the LO approach with the new EMA limits. Actually, Eq. (3) can be rewritten in the sense of the EMA limits (EMA, 2010) as follows:

$$\text{Upper LO limit} = \alpha + \frac{\beta - \alpha}{1 + e^{-\left(\frac{S_{wR} - s_{w0}}{\gamma}\right)}} \quad (4)$$

where the terms *CV* and CV_0 are substituted by s_{wR} and the constant s_{w0}, respectively. Appropriate values should be assigned to the parameters of Eq. (4). Thus, α and β were set equal to 1.25 and 1.4319 (or 0.80 and 0.6984 for the lower limits), respectively, while the most relevant values of s_{w0} and γ were estimated after fitting Eq. (4) to the EMA_s limits.

It should be mentioned that at the time the LO limits were proposed no GMR-constraint was defined. However, this does not exclude the use of an additional GMR criterion when the LO limits are applied.

2.2. Approximate values of the EMA_s limits

A deeper inspection of the EMA_s limits reveals that an approximate value (0.760) of the scaling factor, *k*, is actually being proposed by EMA. In case of the EMA_s, the value of *k* is related to the BE limits and the variability, s_{wR}, with Eq. (2). In other words, the value of *k* is derived from the following equation:

$$k = \ln(1.25)/s_{wR} \quad (5)$$

Therefore, the underlying reason should be sought on the transformation of variability from the log-scale (s_{wR}) to the normal scale (CV_{wR}). A CV_{wR} value equal to 30% corresponds to a s_{wR} value close to, but not exactly, 0.30 (actually, $s_{wR} \approx 0.2935604...$). In turn, this discrepancy may be interpreted in two ways.

Firstly, the exact s_{wR} value can be used to estimate the correct scaling factor. This corrected value will be $k = \ln(1.25)/0.2935604 \approx 0.76012283$, namely slightly higher than the one (0.760) defined by EMA. Alternatively, the value of *k* set by EMA can be used to define another switching criterion of variability. Indeed, setting $k = 0.760$, the exact switch CV_{wR} value will be ≈30.0052858. Hereafter, the more accurate EMA criterion with $k \approx 0.76012283$ will be termed as EMA_{c1}, while the criterion with $CV_{wR} \approx 30.0052858$ as EMA_{c2}. Even though these differences are quite small, this analysis will examine the degree of discrepancies especially at the critical variability regions i.e., at $CV_{wR} = 30\%$ and 50%.

2.3. BE limits studied

Simulated bioequivalence studies were used to assess the performance of the EMA limits in comparison to other interesting cases such as the classic 0.80–1.25 limits (CL), the extended limits 0.75–1.33 (Ext) proposed in the 2001 EMA guideline, the LO limits, and the two more accurate methods, EMA_{c1} and EMA_{c2} (Table 1).

The role of GMR-constraint was also analyzed by examining the behavior of EMA limits without the GMR criterion. These limits were termed as EMA_{nc}. To further unveil the impact of the GMR criterion, a GMR-constraint was also included in the LO limits (termed as LO_g), and finally the performance of GMR was examined as a sole criterion (Gc), Table 1.

2.4. Maximum acceptable differences

The concept of maximum acceptable differences allows the investigation of the range of GMR values that become accepted by each BE approach (Schuirmann, 1987). This task can be

Table 1
Bioequivalence limits considered in the simulations.

Symbol	Description
EMA_s	The scaled BE limits proposed in the EMA guideline
EMA_{nc}	The EMA limits without the secondary GMR-constraint in the range 0.80–1.25
EMA_{c1}	EMA_s limits using a more exact value for the scaling factor: $k \approx 0.76012283$
EMA_{c2}	EMA_s limits using a more exact value for the switching coefficient of variation value: $CV_{wR} \approx 30.0052858$
LO	*Leveling-off* limits based on Eq. (4)
LO_g	*Leveling-off* limits with an additional GMR constraint: $0.80 \leqslant GMR \leqslant 1.25$
Gc	The GMR-constraint (0.80–1.25) applied as the sole criterion of bioequivalence
CL	The classic 0.80–1.25 limits
Ext	The extended (0.75–1.33) BE limits quoted in the 2001 EMA guideline (EMA, 2001).

implemented by setting the upper 90% CI equal to the upper BE limit (e.g., the one described by Eqs. (2) and (4)). To this point, it should be highlighted the distinction between two variability terms: (i) the residual variability of the study which is included in the 90% CI, and (ii) the within-subject variability of the R product which is necessary for the construction of the BE limits. Assuming that both T and R products exhibit the same within-subject variability, then these two variabilities posses the same value. Therefore, the difference between T and R formulations, which can be expressed in terms of GMR, can be represented as a function of CV_{wR}. This kind of extreme GMR plots were constructed in case of EMA_s and LO limits for several sample sizes: 24, 48, and 72 subjects.

2.5. Bioequivalence simulations – Power curves

Three sequence, three period (3×3) crossover studies with two treatments (T and R) were simulated. In these simulated studies, the R product was considered to be administered twice, while the T only once. This design results in three possible sequences: TRR/RTR/RRT. An equal number of subjects was assumed to participate in each sequence. In the current study, the simulated sample sizes were set equal to 24, 48, 72, and 108.

Bioequivalence was declared if the 90% confidence interval around the ratio of the estimated geometric means for the two drug products was between the BE limits, Table 1 (Schuirmann, 1987). Even though, this classic procedure can also be considered as an approximation method, it is preferred due to its simplicity (Endrenyi and Tothfalusi, 2007; Kytariolos et al., 2006; Tothfalusi et al., 2001). In addition, it leads to almost identical results to the other approximate method used for the evaluation of scaled BE (Midha et al., 2005; Tothfalusi and Endrenyi, 2003; Tothfalusi et al., 2003). It is also worth mentioned that when this classic method of the 90% CI is applied to a fairly large sample size (as in case of BE studies for HVD), the observed estimation accuracy is high (Endrenyi and Tothfalusi, 2007, 2008). Nevertheless, this method was tested prior to its use and was successfully applied to previous works (Karalis et al., 2004, 2005, 2009; Kytariolos et al., 2006).

The pharmacokinetic parameter under study was assumed to follow log-normal distribution. Several levels (20%, 30%, 50%, and 70%) of theoretical CV_{wR} were considered for the simulations. In all cases it was assumed that the within-subject variability of the T product was equal to that of R. It is worthy to mention that the switching variability values CV_{wR} = 30% and 50% were included in the simulation framework. This was deliberately made to examine the effect of the approximate values (used in the EMA_s) on the acceptance of BE. However, it should be mentioned that the true CV_{wR} value of a simulated study may be different from (but close to) the theoretical CV_{wR}. Therefore, in case of EMA_s, some simulated studies are following one criterion (e.g., 0.80–1.25), while some others a different criterion (e.g., scaling with k = 0.760). However, the derived overall true CV_{wR} becomes equal to the theoretical one. Hence, the results take into consideration both situations and reflect the overall performance of the BE limit.

The theoretical GMR was gradually changed from 1.00 to 1.50 with an increasing step of 0.05. Under each condition, 40,000 bioequivalence trials were simulated and the percentage of accepted studies was recorded. Power curves were constructed by plotting the percent of acceptance in the vertical axis and the true GMR value in the horizontal axis. The entire programming work was performed in MATLAB®. (MathWorks)

3. Results

Fig. 2 presents the results of fitting the LO limits (Eq. (4)) to the data points of scaled EMA_s approach. For reasons of clarity, only the upper limits are shown. Apparently, similar fittings can be obtained for the lower limits. The estimated parameters for γ and s_{w0} were found to be equal to 0.0336 and 0.3853, respectively. Visual inspection of Fig. 2 reveals that both limits (LO and EMA_s) exhibit almost identical performance. As it is expected, the LO limits show a smoother change since they are based on a single equation which changes gradually with CV_{wR}. On the contrary, the EMA limits are piecewise continuous due to the inclusion of two switching criteria i.e., at 30% and 50%, respectively. The values for γ and s_{w0} estimated in this step were included in Eq. (4) and were further used in the simulated BE trials.

Fig. 3 depicts the extreme (maximum and minimum) GMR accepted values *versus* CV_{wR} in case of the EMA_s and the LO limits. These extreme GMR-plots are shown for several sample sizes with each one corresponding to a different curve. Two drug products are considered bioequivalent if their GMR value lies between the upper and lower boundaries. Fig. 3 reveals that in both cases the same general trend is apparent. For CV_{wR} values up to 30%, the range of maximum accepted values of both EMA_s and LO limits becomes shorter with CV_{wR}. This finding can be explained by the fact that both limits, EMA_s and LO, are constant and equal to 1.25 (or 0.80). When CV_{wR} gets greater than 30%, the extreme GMR acceptable range becomes wider, but until a maximum (or equivalently minimum) value. The widening is due to the scaling of BE limits, while the maximum (or minimum) is attributed to the restriction of scaling after CV_{wR} = 50% is reached. Obviously, the most extreme deviation between the T and R products can be observed in cases where CV_{wR} is equal to 50%. This is the CV_{wR} value after which the bioequivalence limits stay constant (1.4319 or 0.6984).

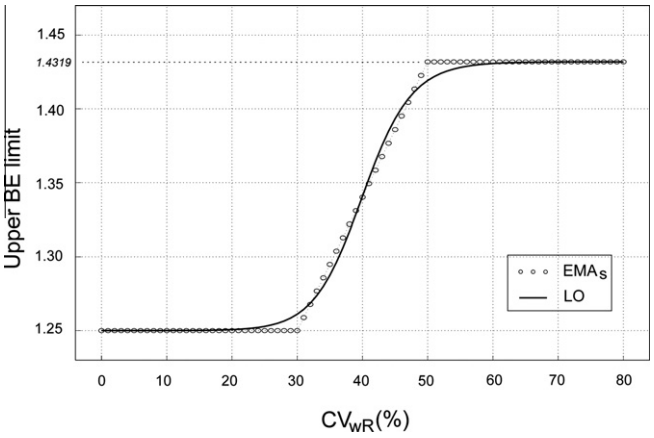

Fig. 2. Fitting of LO limits (Eq. (4)) to the EMA_s approach.

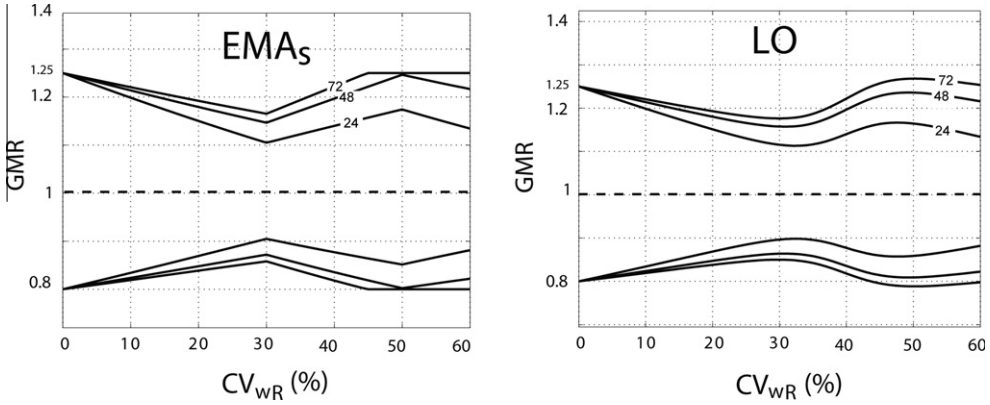

Fig. 3. Maximum accepted GMR values, for the EMA_s and LO limits, as a function of within-subject variability. Each curve corresponds to a different number (24, 48, 72) of subjects.

Therefore, for any further increase of variability, it will be more difficult to declare bioequivalence.

In addition, the role of GMR-constraint, as a cut-off limit, becomes evident in case of the EMA_s limits. The effect of GMR-constraint on bioequivalence acceptance is reflected as a sudden stop in the monotonous change of the curve. Since, no such a constraint was used in the LO, the LO plots are more smooth and their expansion is self-limited. However, the inclusion of a similar GMR-constraint in the LO limits (now termed as LOg limits), leads to plots that are similar to these obtained for EMAs (data not shown).

As the number of subjects increases, the maximum GMR accepted values become more liberal (Fig. 3). Obviously, this is anticipated because it is easier to declare bioequivalence when more subjects are recruited in the study. However, in case of EMA_s limits, the increment of sample size makes the role of GMR-constraint more prominent. If more than 72 subjects are recruited in a BE study, the flat region, due to GMR-constraint, will become wider. In other words, Fig. 3 reveals that the GMR-constraint is only effective at CV_{wR} values around 50% and when a large number of subjects is used. In the same vein, extreme GMR plots clearly show that the inclusion of less than 48 subjects makes the secondary GMR criterion ineffective.

Fig. 4 shows the percentage of studies in which BE is accepted as a function of GMR of the simulated studies. Five different bioequivalence limits were examined: EMA_s, LO, EMA_{nc} (i.e., the EMA_s without the secondary GMR-criterion), the LO limits with the additional constraint on GMR (LOg), and the use of GMR-constraint as the sole criterion (Gc) of BE (see Table 1). The upper panel of Fig. 4 represents the results in case of medium levels of CV_{wR} i.e., 20% and 30%. Besides, high CV_{wR} values (50% and 70%) are depicted in the lower panel of Fig. 4. It should be mentioned that simulation results are available for a variety of sample sizes and levels of variability as quoted in Section 2. However, due to space-limitations, only a portion of them (the most representative) is shown in this manuscript. In the same vein, depiction of LOg was omitted for reasons of clarity. Any result that will be informative for the reader is available upon request.

At low CV_{wR} values (20% and 30%), EMA_s and LO limits exert almost identical performance. When CV_{wR} is 20% (Fig. 4A), no scaling is effective and both limits are actually equal to 1.25 or 0.80. When the theoretical CV_{wR} is 30% (Fig. 4B), some studies exhibit true CV_{wR} values greater than 30%, while some others lower than 30%. Thus, in case of EMA_s approach, some simulated studies will follow the 0.80–1.25 limits, while some others the scaling criterion with $k = 0.760$. The results shown in Fig. 4B reflect the overall performance, namely, they take into consideration both situations. Similar remarks can also be made for the 50% switching variability value.

When CV_{wR} gets equal to 50% (Fig. 4C), the performance of EMA_s becomes slightly different from that of the LO limits. Nevertheless, in the critical region of 80% power, the discrepancy between the% acceptance values of EMA_s and LO does not exceed 5%. In any way, this small discrepancy can be attributed to the fact that the GMR-constraint of the EMA_s limits becomes effective. This finding is in accordance with the theoretical expectations shown above in extreme GMR plots (Fig. 3); it was shown that as CV_{wR} gets close to 50% and the sample size is large enough, then the GMR-constraint has a role for the determination of bioequivalence. As CV_{wR} deviates from the switching value of 50%, either to greater (as in Fig. 4D) or to lower values, there is no role for the GMR-constraint. Hence, the discrepancy between LO and EMA_s (or LO_g) limits gradually diminishes and it is finally vanished at higher CV_{wR} values.

In order to further validate this hypothesis, one may compare the results for LO and LO_g i.e., also include a GMR-constraint in the LO limits. Indeed, the performance of LO_g limits is identical to that of EMA_s (data not shown). In the same vein, the results derived for EMA_{nc} (i.e., the EMA_s approach without the GMR criterion) always coincide with the LO limits (Fig. 4). Overall, EMA_s and LO_g show almost identical performance, while the same exists for the other pair: EMA_{nc} and LO limits.

Furthermore, the performance of the GMR-constraint as a sole criterion (Gc) is presented in Fig. 4. At low CV_{wR} values (Fig. 4A and 4B), the use of Gc results in a very permissive behavior. In other words, the percentage of acceptance for the Gc criterion is by far higher than any other bioequivalence limit. However, as CV_{wR} values tend towards 50% this discrepancy gets lower. When $CV_{wR} = 50\%$ and an increased number of subjects is used (e.g., 72 as in Fig. 4C), the Gc performance becomes identical to that of EMA limits. This is expected since at this level of variability the GMR-constraint is more strict than the 0.6984–1.4319 limits of the EMA approach and determines the outcome of bioequivalence. For CV_{wR} values greater than 50%, application of the scaled EMA limits leads to stricter criteria than Gc. Thus, the Gc approach becomes again more permissive and the discrepancy, between EMA_s and Gc, is enhanced.

Fig. 5 presents the percentage of accepted BE studies using EMA_s in comparison to the typical 0.80–1.25 limits, the extended 0.75–1.33 limits, and the two more accurate approaches EMA_{c1} and EMA_{c2} (see Table 1). In all cases, the performance of EMA_{c1} was identical to that of EMA_{c2}. Thus, for reasons of clarity only results for EMA_{c1} are shown in Fig. 5. As it is expected, at low CV_{wR} values (Fig. 5A and 5B), both EMA_s and CL exhibit an almost similar behavior. However, as variability increases, the new EMA_s limits become much more permissive than the 0.80–1.25 limits (Fig. 5C and 5D). This attribute is evident at all GMR values, namely, either

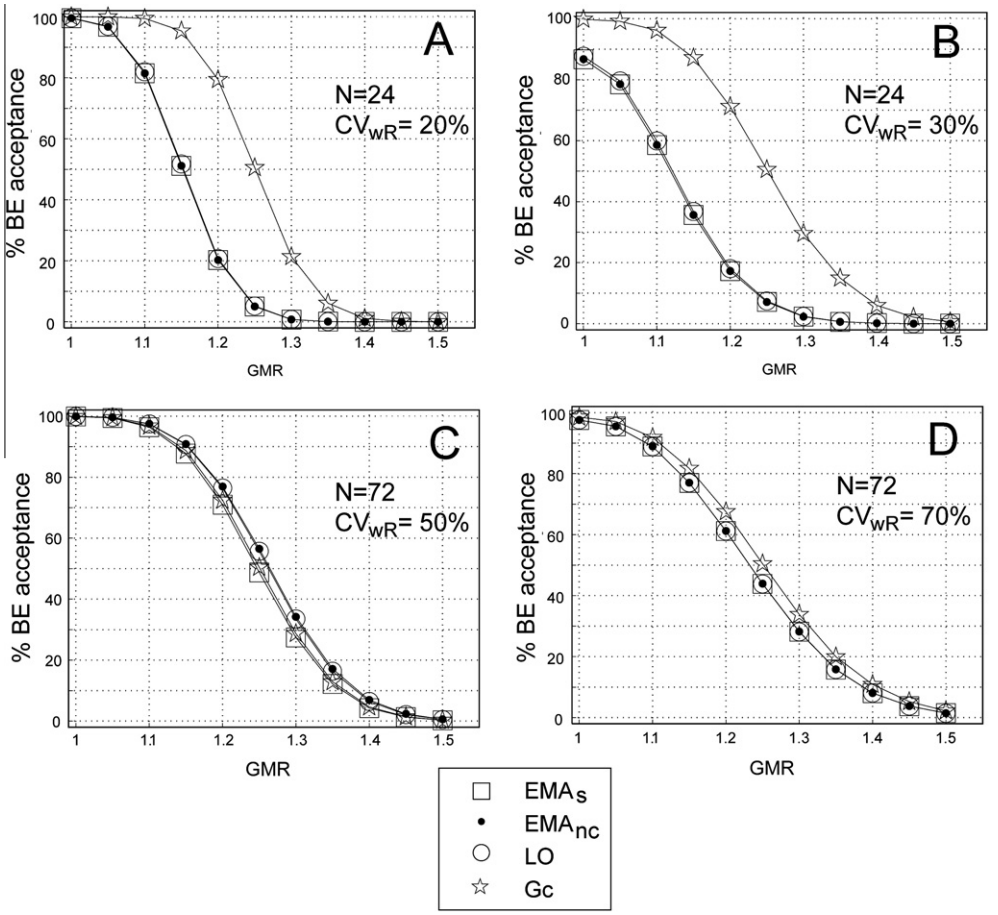

Fig. 4. Percentage of 3x3 BE studies accepted, by four procedures: EMA_s, EMA_{nc}, LO, and Gc (see Table 1) at various ratios of GMR. Upper panel (A and B): 24 subjects, CV_{wR} = 20% and 30%. Lower panel (C and D): 72 subjects, CV_{wR} = 50% and 70%.

the two drug products under comparison are identical (GMR = 1) or differ significantly. Another characteristic, is the fact that as the number of subjects in the study is increased, the difference between EMA and CL limits diminishes (data not shown).

In regard to the 0.75–1.33 limits, the currently proposed EMA approach presents again an expected behavior. At CV_{wR} values such as 20% and 30% (Fig. 5A and 5B), both EMA_s and Ext exhibit an almost similar behavior. This is due to the fact that in these low variabilities, the effective criterion for EMA_s and Ext is the 0.80–1.25 acceptance range. However, as CV_{wR} (and residual variability of the BE study) increases, these two approaches behave differently. In such cases (e.g., when CV_{wR} = 50% or 70%), the EMA_s become much more permissive than Ext limits. This finding is associated with the wider interval (namely, 0.6984–1.4319) of EMA_s limits compared to the 0.75–1.33 range. As sample size increases, the difference in the performance of EMA_s and Ext become less pronounced.

It should be mentioned that the 0.80–1.25 and the 0.75–1.33 limits had originally been proposed for the simple 2 × 2 crossover design (EMA, 2001). However, in order to have fairly comparable data for the purposes of the current analysis, the 0.80–1.25 and 0.75–1.33 limits were appropriately applied to the semi-replicate 3 × 3 design.

Fig. 5 also depicts the comparative performance of EMA_s *versus* the two more exact methods, EMA_{c1} and EMA_{c2} (Table 1). A complete description of the EMA_{c1} and EMA_{c2} limits is given in the "Methods" section. Visual inspection of Fig. 5 reveals that no discrepancy can be observed in their ability to declare bioequivalence. None of EMA_{c1} or EMA_{c2} limits exhibits a significantly different behavior that the one seen by EMA_s. In other words, the use of the approximated values does not seem to affect their performance in power curves.

For all bioequivalence limits depicted in Figs. 4 and 5, the increase of sample size results in an enhanced ability to declare bioequivalence. However, one advantage of EMA_s (or the LO) limits is the ability to declare bioequivalence with fewer subjects. In other words, the *leveling-off* limits (either EMA_s or LO) can achieve the same statistical power with the typical 0.80–1.25 limits without the need of recruiting many subjects. This feature is depicted in Fig. 6 where the statistical power of EMA_s and CL limits is plotted against the GMR of the study in case of two levels of variability (30% and 50%). Each curve in Fig. 6 corresponds to a different number of subjects. Visual inspection of Fig. 6 reveals that, for highly variable drugs, the EMA_s approach can achieve higher statistical power. For example, in order to achieve 80% power given that CV_{wR} = 50% and GMR = 1.05, then 72 and not more than 30 subjects should be recruited according to the CL and the EMA_s approach, respectively. Obviously for variability values higher than 50%, these differences are magnified. It is noteworthy that in case of CV_{wR} equal to the switching criterion 30%, both approaches lead to almost similar results; the% Power achieved with EMA_s is only slightly higher than the% power of CL.

4. Discussion

This study examined the application of the newly proposed EMA limits to the determination of bioequivalence. It was shown that these new BE limits are actually *leveling-off* limits, since they

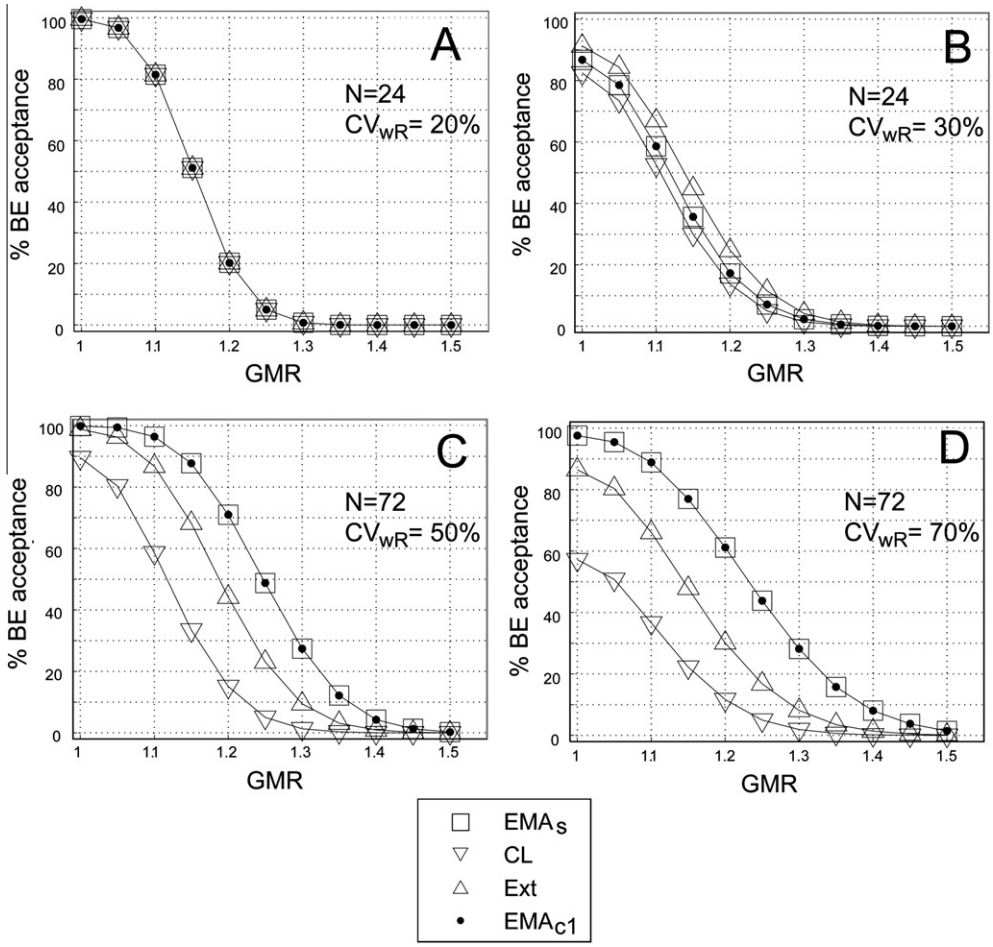

Fig. 5. Percentage of 3x3 BE studies accepted, by four procedures: EMA_s, CL, Ext, EMA_{c1} (see Table 1) at various ratios of GMR. Upper panel (A and B): 24 subjects, CV_{wR} = 20% and 30%. Lower panel (C and D): 72 subjects, CV_{wR} = 50% and 70%.

are composed of three parts: (i) a basal value which is 1.25 or 0.80 depending on whether the upper or lower limit is considered, (ii) an intermediate variable with CV_{wR} segment, and (iii) a maximum (or minimum) constant value when CV_{wR} exceeds a specific value (Fig. 1A). The switch from one type to another is defined in the EMA guideline on the basis of the CV_{wR} value (EMA, 2010).

Actually, BE limits with such properties were proposed few years earlier and termed as "*leveling-off*" limits (Karalis et al., 2005; Kytariolos et al., 2006). However, the so-called "*leveling-off*" limits were based on a single function instead of using *switching variability* values (Fig. 1B). It was quoted that the values of the functions' parameters (i.e., the basal and maximum value) could be set by the regulatory authorities. Different values of these parameters would lead to different strictness of the LO function to declare bioequivalence. In this study, the assignment of parameters' values was achieved by fitting the LO function to the BE limits proposed by EMA (EMA, 2010). Besides, an advantage of the LO limits is their continuous nature, which allows them to be applied as an "all-in-one" criterion. It should be stated that at the time the LO limits were proposed, the replicate design was not encouraged by the regulatory authorities (EMA, 2001, 2006). So, the LO limits were defined on the basis of the residual variability of the study. However, the LO limits can still work with CV_{wR} by substituting residual variability by within-subject coefficient of variation of the R product (s_{wR}).

The extreme GMR plots, depicted in Fig. 3, further reveal the full analogy of the LO with the EMA limits. In both cases (LO and EMA_s), the extreme GMR acceptance range presents the following general trend as a function of CV_{wR}: (i) initially a shrinkage, (ii) then an extension up to a maximum, and finally (iii) a monotonous convergence towards unity. The only difference between LO and EMA_s can be focused on the smoothness of the curves, with the LO limits to be more smooth. On the contrary, the EMA_s procedure utilizes switching criteria (at CV_{wR} = 30% and 50%) and the corresponding curves are characterized by more steep changes. In addition, the approach proposed by EMA also include a GMR-constraint which is reflected on the extreme GMR plots as the flat part of the curves (Fig. 3). The latter is shown in the regions close to the switching value 50% and become more evident as the number of subjects participating in the study is getting larger.

The ability of the EMA_s to declare bioequivalence was found to be identical to that of LO limits in case of low to medium CV_{wR} values (Fig. 4). However, as variability increases, a small discrepancy is observed between the percentage of studies in which BE is accepted by the two approaches (EMA_s and LO). This discrepancy becomes more evident when CV_{wR} is close to the switching value CV_{wR} = 50% and a large number of subjects is assumed. Plausibly, the sample size plays an important role. For a low number of subjects, i.e., up to 48, the GMR-constraint has no role (see Fig. 3). The effect of GMR-constraint begins at sample sizes around 60 and becomes more prominent as the number of subjects enrolled in the BE study gets higher. For example, when 72 subjects are included (Fig. 4C), the discrepancy between EMA_s and LO (where no GMR-constraint is used) leads to a 5% difference in the percentages of acceptance when the power of the study is 80%. In case of higher power values, such as 90% (at GMR = 1.15), this difference practically vanishes. The underlying reason, for this discrepancy, is the

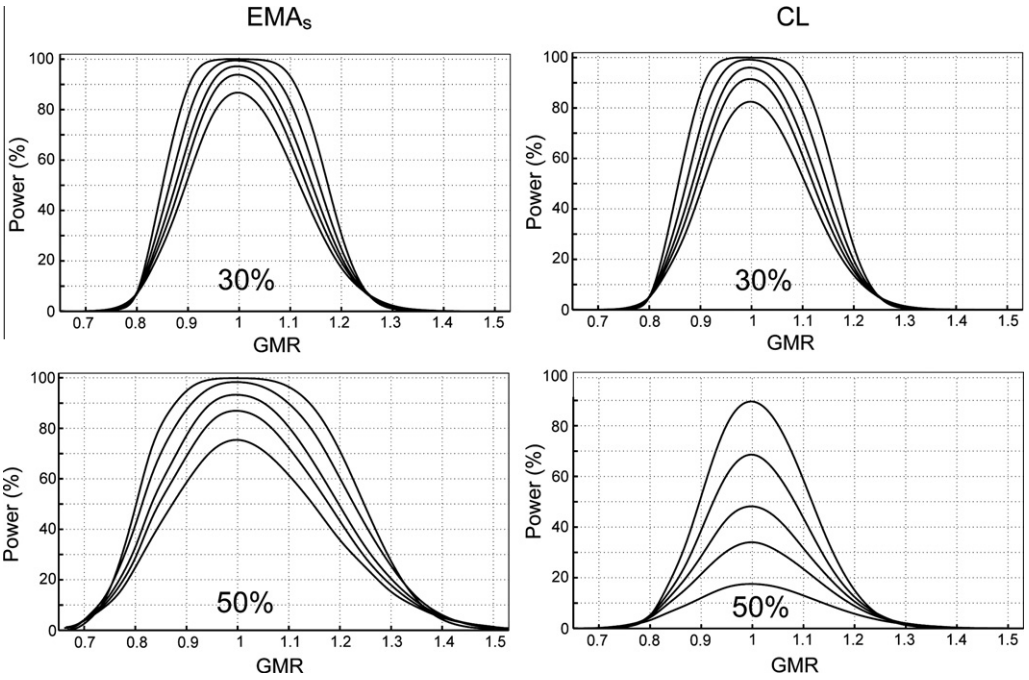

Fig. 6. Percent probability of correctly concluding BE (Power%), for EMA_s and CL limits, at several ratios of GMR assuming 3 × 3 design. Two levels of within-subject variability (30% and 50%) are depicted. Power curves refer to sample size of (from bottom to top): 24, 30, 36, 48, and 72.

GMR-constraint used in the EMA_s. Inclusion of a similar criterion into the LO limits led to a performance identical to that of EMA_s. Besides, the behavior of the GMR-criterion as a sole method for bioequivalence assessment was also studied. It was observed that at low variabilities, the Gc method was more permissive than any other approach (Fig. 4A and 4B). For CV_{wR} values close to 50%, the performance of Gc was similar to the EMA limits. Apparently, at this level of variation, the GMR-constraint of the EMA approach is stricter than the scaled limits (see also Fig. 2); hence, determination of BE depends only on this criterion.

Compared to the typical 0.80–1.25 and the extended 0.75–1.33 limits (proposed in the previous 2001 EMA guideline), the EMA_s exhibit the following behavior (Fig. 5): (i) at low variabilities all three approaches (EMA_s, CL, and Ext) coincide, (ii) for CV_{wR} greater than 30% the CL limits exhibit the most strict performance, (iii) when CV_{wR} is greater than (but close to) 30%, the Ext limits are more permissive than EMA_s, (iv) as CV_{wR} increases the EMA_s limits become more liberal than Ext and a higher probability to declare BE is observed. In other words, the newly proposed EMA_s allow greater differences between the two drug products to be declared bioequivalent, a finding that is attributed to the wider acceptance interval (0.6984 – 1.4319) of the EMA_s.

Another issue examined in this study, was the impact of the approximate values (e.g., $k = 0.760$) used in the EMA_s approach. In order to evaluate any possible problems, power curves were constructed placing special emphasis at the CV_{wR} values at which the probability to observe any differences is maximized (i.e., at 30% and 50%). However, extensive simulations, using several samples and variabilities, did not reveal any differences in power curves.

5. Conclusions

In 2010, the European Medicines Agency issued a new guideline on bioequivalence assessment. Even though, not all issues are discussed, this guideline opened new ways in the methodology of planning and assessing BE studies (Marzo and Fontana, 2011). Among others, the new EMA 2010 guideline, proposed a novel procedure for the determination of bioequivalence in case of highly variable drugs or drug products. The suggested approach includes the application of replicate designs where the R product is administered twice allowing, thus, the estimation of the within-subject coefficient variation of the reference formulation. EMA suggests the use of the estimated CV_{wR} to estimate scaled BE limits when CV_{wR} lies between 30% and 50%. Outside these variability values, an upper and a lower bound is set for the BE limits. In addition, the EMA guideline proposes the inclusion of a secondary constraint on the point GMR of the study.

Aim of this study was to analyze the properties of the new BE limits proposed by EMA (EMA, 2010). Among others it was shown that:

(i) the complementary GMR-constraint is effective only at regions of within-subject variability close to the switching value of 50% and when a large sample size is used. No role of the GMR-constraint was found when a moderate number of subjects (e.g., up to 48) is enrolled. The GMR-criterion begins to be effective at sample sizes around 60 and becomes more prominent as the number of subjects is getting higher. Nevertheless, due to the *leveling-off* properties of the EMAs limits, the impact of the GMR-constraint diminishes with a further increase of variability.

(ii) at low variabilities, the new EMA limits exhibit performance identical to the classic 0.80–1.25 or even to the previously reported extended 0.75–1.33 limits. However, as variability rises, the EMAs limits become more liberal and allow greater differences between the two drug products to be declared bioequivalent. This property allows the recruitment of a fewer number of subjects in the bioequivalence study.

(iii) these new BE limits are in essence *leveling-off* limits, since they are composed of a basal value (1.25 or 0.80), an intermediate variable with variability segment, and an extreme (1.4319 or 0.6984) value. BE limits with such properties

were proposed few years earlier and were termed as "*leveling-off*" (LO) limits (Karalis et al., 2005; Kytariolos et al., 2006). After adjusting the parameters of the LO limits to be in agreement with the EMAs limits, the performances of both limits (EMAs and LO) were found to be almost identical.

(iv) the approximate value of $k = 0.760$, applied to the EMA limits, has no impact on the performance of the new BE limits.

References

Anderson, S., Hauck, W., 1990. Consideration of individual bioequivalence. J. Pharmacokinet. Biopharm. 18, 259–273.

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

Blume, H., Elze, M., Potthast, H., et al., 1995. Practical strategies and design advantages in highly variable drug studies: multiple dose and replicate administration design. In: Blume, H., Midha, K. (Eds.), Bio-international 2: Bioavailability. Bioequivalence and Pharmacokinetic studies. Medpharm Scientific Publishers, Stuttgart, pp. 117–122.

Blume, H., Midha, K., 1993. Bio-international '92, conference on bioavailability, bioequivalence and pharmacokinetic studies. J. Pharm. Sci. 82, 1186–1189.

Boddy, A., Snikeris, F., Kringle, R., et al., 1995. An approach for widening the bioequivalence acceptance limits in the case of highly variable drugs. Pharm. Res. 12, 1865–1868.

EMA (European Medicines Agency), 2001. Evaluation of Medicines for Human Use, CPMP. Note for Guidance on the Investigation of Bioavailability and Bioequivalence, London.

EMA (European Medicines Agency). 2006. Evaluation of Medicines for Human Use, CHMP efficacy working party] therapeutic subgroup on pharmacokinetics: Questions & Answers on the Bioavailability and Bioequivalence Guideline, London.

EMA (European Medicines Agency), 2010. Committee for Medicinal Products for Human Use, CHMP. Guideline on the Investigation of Bioequivalence, London.

Endrenyi, L., Amidon, G., Midha, K., et al., 1998. Individual bioequivalence. attractive in principle, difficult in practice. Pharm Res 15, 1321–1325.

Endrenyi, L., Tothfalusi, L., 2007. Determination of bioequivalence for highly-variable drugs. AAPS Annual Meeting: Current issues and advances in the determination of bioequivalence. November 11th-15th. San Diego, USA.

Endrenyi, L., Tothfalusi, L., 2008. Evaluation of bioequivalence of highly variable drugs. Clin. Res. Regulat. Affairs. 25, 93–117.

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

FDA (Food and Drug Administration). 2003. Center for Drug Evaluation and Research (CDER), Bioavailability and Bioequivalence Studies for Orally Administered Drug Products. General Considerations, Rockville, MD.

Hauck, L., Parekh, A., Lesko, L., et al., 2001. Limits of 80%-125% for *AUC* and 70%-143% for C_{max}. What is the impact on the bioequivalence studies? Int. J. Clin. Pharmacol. Ther. 39, 350–355.

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

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

Karalis, V., Symillides, M., Macheras, P., 2009. Comparison of the reference scaled bioequivalence semi-replicate method with other approaches: focus on human exposure to drugs. Eur. J. Pharm. Sci. 38, 55–63.

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

Marzo, A., Fontana, E., 2011. Critical considerations into the new EMA guideline on bioequivalence. Arzneimittelforschung. 61, 207–220.

Midha, K., Rawson, M., Hubbard, J., 1997. Individual and average bioequivalence of highly variable drugs and drug products. J. Pharm. Sci. 86, 1193–1197.

Midha, K., Rawson, M., Hubbard, J., 1998. Bioequivalence. Switchability and scaling. Eur. J. Pharm. Sci. 6, 87–91.

Midha, K., Rawson, M., Hubbard, J., 2005. The bioequivalence of highly-variable drugs and drug products. Int. J. Clin. Pharmacol. Ther. 43, 485–498.

Midha, K., Shah, V., Singh, G., Patnaik, R., 2007. Conference report: bio-international. J. Pharm. Sci. 96, 747–754.

Morais, J.A., Lobato Mdo, R., 2010. The new European Medicines Agency guideline on the investigation of bioequivalence. Basic Clin. Pharmacol. Toxicol. 106, 221–225.

Patnaik, R., Lesko, L., Chen, M.L., Williams, R., 1997. Individual bioequivalence. New concepts in the statistical assessment of bioequivalence metrics. Clin. Pharmacokin. 33, 1–6.

Schall, R., Luus, H., 1993. On population and individual bioequivalence. Stat. Med. 12, 1109–1124.

Schuirmann, D.J., 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., et al., 1996. Evaluation of orally administered highly variable drugs and drug formulations. Pharm. Res. 13, 1590–1594.

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

Van Peer, A., 2010. Basic variability and impact on design of bioequivalence studies. Clin. Pharmacol. Toxicol. 106, 146–153.