Pharm Res (2012) 29:1066–1077
DOI 10.1007/s11095-011-0651-y

RESEARCH PAPER

Bioequivalence of Highly Variable Drugs: A Comparison of the Newly Proposed Regulatory Approaches by FDA and EMA

Vangelis Karalis • Mira Symillides • Panos Macheras

Received: 1 September 2011 / Accepted: 5 December 2011 / Published online: 28 December 2011

ABSTRACT

Purpose To explore the comparative performance of the recently proposed bioequivalence (BE) approaches, FDA_s and EMA_s, by the FDA working group on highly variable drugs and the EMA, respectively; to compare the impact of the GMR-constraint on the two approaches; and to provide representative plots of % BE acceptance as a function of geometric mean ratio, sample size and variability.

Methods Simulated BE studies and extreme GMR *versus* CV plots were used. Three sequence, three period crossover studies with two treatments were simulated using four levels of within-subject variability.

Results The FDA_s and EMA_s approaches were identical when variability was <30%. In all other cases, the FDA_s method was more permissive than EMA_s. The major discrepancy was observed for variability values >50%. The GMR-constraint was necessary for FDA_s, especially for drugs with high variabilities. For EMA_s, the GMR-constraint only became effective when sample size was large and variability was close to 50%.

Conclusions A significant discrepancy in the performances of FDA_s and EMA_s was observed for high variability values. The GMR-constraint was essential for FDA_s, but it was of minor importance in case of the EMA_s.

KEY WORDS bioequivalence · European Medicines Agency · Food and Drug Administration · highly variable drugs · replicate design

ABBREVIATIONS

90% CI	90% confidence interval
AUC	area under the curve
BE	bioequivalence
C_{max}	peak plasma concentration
CV_w	coefficient of variation
CV_{wR}	coefficient of variation corresponding to s^2_{wR}
EMA_{nc}	modified EMA approach without GMR-constraint
EMA_s	scaled approach proposed by EMA
FDA_{nc}	modified FDA approach without GMR-constraint
FDA_s	scaled approach proposed by FDA scientists
GMR	geometric mean ratio
HVDs	highly variable drugs
k	scaling factor of the limits proposed by EMA
PK	pharmacokinetic(s)
R	reference product (i.e., the innovator's product)
s^2_w	within-subject variability
s^2_{wR}	within-subject variability of the reference product
s^2_{wT}	within-subject variability of the test product
s_{w0}	constant referring to regulatory standardized variation of FDA_s limits
s_{wR}	standard deviation corresponding to s^2_{wR}
T	test product (i.e., product under evaluation)

INTRODUCTION

Proving bioequivalence (BE) is essential in order to ensure therapeutic similarity between two drug products of the same active moiety (1). Usually, determination of BE relies on the comparison of the *rate* and *extent* of absorption of a product under study (Test, T) with an innovator's product (Reference, R). Therefore, two drug products of the same active substance are considered bioequivalent if they contain

V. Karalis (✉) · M. Symillides · P. Macheras
Laboratory of Biopharmaceutics-Pharmacokinetics
Faculty of Pharmacy, National and Kapodistrian University of Athens
Panepistimiopolis
Athens 15771, Greece
e-mail: vkaralis@pharm.uoa.gr

the same active moiety, are at the same molar dose, and their *extent* and *rate* of absorption values are so similar to minimize the risk of any difference in their *in vivo* performances (1–3). Area under the curve (*AUC*) and peak plasma concentration (C_{max}) are routinely used to describe these two properties namely, *extent* and *rate* of absorption or *total* and *peak* exposure, respectively.

However, BE assessment becomes a difficult issue in case of highly variable drugs or drug products (HVDs). For reasons of simplicity in this study, no distinction will be made between highly variable drugs or drug products. Thus, HVDs are considered the drugs which exhibit high within-subject variability (s^2_w) values. In particular, it has widely been accepted that the within-subject coefficient of variation (CV_w) of HVDs is greater than or equal to 30% (4–8). In case of HVDs, the risk of erroneously rejecting BE between two drugs (producer risk) arises, unless a large number of subjects is recruited in the study. This issue, however, raises many ethical and financial concerns about the participation of large number of healthy subjects in clinical trials (9,10). This opinion is further encouraged by the inherent property of HVD to be safe drugs with wide therapeutic indices (11).

In the past, several approaches were proposed to overcome the problem of high variability in BE testing. Among others, these methods include the application of multiple-dose studies (5), the conduction of replicate designs (1,2,5,7), the introduction of the concept of individual BE (12–17), the widening of BE limits to other pre-fixed values (2,18–20), and the use of scaled BE approaches (20–26).

Recently, new approaches aiming to resolve the problem of high variability, have been proposed by the FDA working group on HVDs (27,28) and by the latest guideline of the European Medicines Agency (EMA) (3). For reasons of simplicity and uniformity in this analysis, these two newly proposed methods for bioequivalence assessment will be termed as FDA_s and EMA_s. Both approaches propose the application of either a full replicate or a semi-replicate design (3,27–30). The key point in both approaches is the demand that at least the reference product should be administered twice in each individual. The latter allows the estimation of within-subject variability of the R product (s^2_{wR}) or, equivalently, the corresponding standard deviation (s_{wR}). In turn, both approaches suggest the use of scaling with s_{wR} bioequivalence limits when the within-subject coefficient of the R product (CV_{wR}) exceeds a pre-set switching value of 30%. Finally, FDA_s and EMA_s introduce a secondary constraint criterion on the ratio of geometric means (GMR) of T and R in the region 0.80–1.25. However, the EMA approach further imposes an extreme limit of CV_{wR} upon which the scaled limits should be used. Thus, the EMA_s limits can be expanded only up to a maximum range 0.6984–1.4319. No such criterion is present for the FDA_s method. In addition, different scaling factors are included in the FDA_s and EMA_s approaches. Therefore, despite the fact that these two approaches rely on a similar ground, it is anticipated that they will exhibit a different performance in BE studies.

Aim of this study was to explore the comparative performance of the recently proposed approaches FDA_s and EMA_s (3,27,28). This task was achieved by using simulation studies, assuming different levels of variability values, and considering several sample sizes. Furthermore, this analysis focused on the role and the relative impact of the GMR-constraint on the FDA_s and EMA_s approaches. Finally, this study provides the reader with representative plots regarding the probability of declaring BE, with the FDA_s and EMA_s methods, for many GMR estimates, four different levels of variability, and a wide range of sample sizes.

MATERIALS AND METHODS

Bioequivalence Assessment

Classic BE

Classically, determination of bioequivalence relies on the concept of average BE (1). Two drug products are declared bioequivalent if the calculated 90% confidence (CI) interval around the difference of their mean measures of bioavailability (in the log-transformed scale) lies between some pre-set limits defined by the regulatory authorities (1,3,14,31).

Obviously, the use of a confidence interval implies that the decision of bioequivalence depends on both the difference between the T and R measures of bioavailability (e.g., *AUC*, C_{max}) and the residual variability of these pharmacokinetic (PK) measures as well. For BE purposes, the residual variability is considered to reflect the within-subject variability of the drug under study.

This average BE testing approach is widely used worldwide; however, it becomes problematic in case of highly variable drugs. Obviously, in such cases, it turns out to be difficult to declare bioequivalence, unless a large number of subjects is recruited in the study. In order to overcome this problem, the FDA_s and EMA_s approaches have recently recommended the application of reference-scaled approaches.

The FDA Approach (FDA_s)

Scientists from the FDA working group on highly variable drugs proposed that, in case of HVDs, a reference scaled average bioequivalence approach could be used (27,28). Actually, this opinion of the FDA scientists had been

introduced few years earlier (32–35). This R-scaled approach suggests the use of either a full-replicate or a semi-replicate design (27–30). In particular for the semi-replicate design, the R product is administered twice, while the T product only once. In other words, three sequence (RTR, TRR, and RRT), three period studies can be conducted. A basic feature of this design is the fact that it allows the estimation of the within-subject variability (i.e., s^2_{wR}) of the R product and, thus, its corresponding CV_{wR} value.

According to the FDA$_s$ method, a reference-scaled (with s_{wR}) approach should be used for drugs which have within-subject variability greater than or equal to 30%, Fig. 1a. Actually, the FDA$_s$ definition (27) implies the use of CV_{wR} as a switching criterion rather than the variability s^2_{wR} term itself. The scaled BE limits imposed by the FDA$_s$ approach are mathematically expressed by Eq. (1):

$$Upper/Lower\ limits = \exp\left(\pm\ln(1.25)\cdot\frac{s_{wR}}{s_{w0}}\right) \quad (1)$$

where s_{w0} is a regulatory standardized variation. It has been suggested (27) that a value of 0.25 can be assigned to s_{w0} in order to exhibit a good balance between a conservative and a practical behavior. To this point, it should be clarified that for simplicity reasons, in this paper, no distinction will be made between the *population* (e.g., μ_i, σ_i) and the relevant *sample* terms (e.g., m_i, s_i).

In addition, the FDA approach includes a secondary constraint on the point GMR estimate of the BE study. According to this criterion, the point GMR of the study should lie within the range 0.80–1.25. This constraint is used to avoid the risk of accepting two drug products which in fact may differ significantly in their GMR values. An interesting property of the GMR-constraint is reflected on the fact that if the estimated GMR is exactly at the boundary, then there is less than a 50% probability that the true GMR is in the range 0.80–1.25.

Finally, it is worth mentioned that the FDA$_s$ procedure can be applied to both *AUC* and C_{max} estimates. The basic characteristics of the FDA$_s$ approach are listed in Table I.

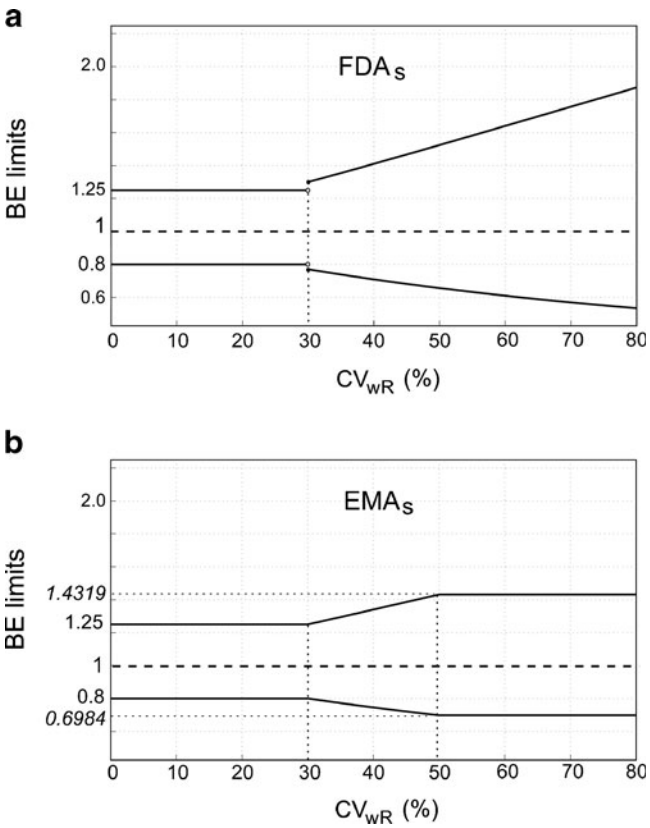

Fig. 1 FDA$_s$ (**a**) and EMA$_s$ (**b**) bioequivalence limits as a function of within-subject variability of the reference product (CV_{wR}). In case of FDA$_s$ (**a**) a "discontinuity" of the limits is apparent at 30%.

The EMA Approach (EMA$_s$)

In August 2010, the European Medicines Agency issued a new guideline regarding the BE assessment (3). In case of highly variable drugs, EMA suggests the potential application of a replicate crossover design with 3 or 4 periods. It is not specified in the guideline the exact type of the replicate design, however, it is underlined the fact that the R product should be administered twice. As in case of FDA$_s$, the latter allows the estimation of s_{wR} which can further be used to define scaled with s_{wR} bioequivalence limits, Eq. (2):

$$Upper/Lower\ BE\ limits = \exp\left(\pm k\cdot s_{wR}\right) \quad (2)$$

where k is a scaling factor that was set by the regulatory authorities equal to 0.760 (3).

According to the EMA$_s$ approach, the BE limits described by Eq. (2) should only be used when the CV_{wR} value of the study is between 30% and 50%. When CV_{wR} values are lower than 30%, the classic 0.80–1.25 limits should be used. Besides, for CV_{wR} values greater than 50%, the extreme values 1.4319 and 0.6984 are imposed for the upper and the lower limits, respectively (Fig. 1b). In other words, EMA$_s$ limits exhibit leveling-off properties. In addition, the EMA$_s$ method includes a GMR-constraint in the region 0.80–1.25. Nevertheless, it should be highlighted that the abovementioned EMA$_s$ approach applies only to C_{max}, whereas for *AUC*, the classic 0.80–1.25 limits remain the only option regardless of the level of variability (3,36).

The similarities and the differences between the EMA$_s$ and FDA$_s$ approaches are summarized in Table I.

Table I The Basic Features of the FDA_s and EMA_s Methods

	Bioequivalence approach	
	FDA_s [a]	EMA_s [b,c]
Clinical design	Full- or Semi-replicate	Full- or Semi-replicate
GMR-constraint	0.80–1.25	0.80–1.25
Scaling criterion	s_{wR}	s_{wR}
Scaling factor	$k'=\ln(1.25)/0.25\approx 0.893$	$k=0.760$
Switching criterion	CV_{wR} (%)	CV_{wR} (%)
Upper/Lower BE limit:		
$CV_{wR}<30\%$ [d]	0.80–1.25	0.80–1.25
$30\%\le CV_{wR}<50\%$ [e]	$=\exp\left(\pm\ln(1.25)\cdot\frac{s_{wR}}{s_{w0}}\right)$	$\exp(\pm k\cdot s_{wR})$
$CV_{wR}\ge 50\%$		0.6984–1.4319

[a] See Ref. (27,28)

[b] See Ref. (3)

[c] In case of EMA_s, the possibility to widen the acceptance criteria apply only to C_{max}. For *AUC*, the classic 0.80–1.25 approach should always be used (3)

[d] For EMA_s the switching criterion is: $CV_{wR}\le 30\%$

[e] For EMA_s the switching criterion is: $30\%<CV_{wR}<50\%$

Simulation Framework

Extreme GMR Accepted Values

Construction of the extreme GMR plots can serve as a tool to investigate the range of GMR values that can be accepted by a BE approach as a function of variability (31). This task can be achieved by setting the *Upper BE limit*, of either Eq. (1) or (2), to be equal to the *Upper 90% confidence interval*, namely:

$$Upper\ BE\ limit = Upper\ 90\%\ CI \tag{3}$$

Similarly, the lower BE limit can be combined with the lower part of the 90% CI.

Assuming that an equal number of subjects is included in each sequence, then the *Upper 90% CI* of a three sequence, three period crossover design (3x3) will be given by Eq. (4):

$$Upper\ 90\%\ CI = \exp\left(\ln(GMR) + t_{\alpha,df}\cdot\sqrt{s^2{}_w\cdot\frac{3}{2N}}\right) \tag{4}$$

where $s^2{}_w$ refers to the residual variability of the study, N is the total number of subjects in the BE study, and $t_{\alpha,df}$ is the t-Student statistic with $df=2N-3$ degrees of freedom (no *carry-over* effect is assumed) at $\alpha=0.05$ significance level. If the within-subject variability of the T product ($s^2{}_{wT}$) is equal to that of R, namely $s^2{}_{wT}=s^2{}_{wR}$, then the residual variability of the study will also be equal to $s^2{}_{wR}$. Thus, the $s^2{}_{wR}$ (instead of $s^2{}_w$) term can be used for the construction of the *Upper 90% CI* in Eq. (4).

Therefore, one can substitute into Eq. (3): i) the *Upper BE limit* proposed by the regulatory authorities, namely, Eq. (1) or (2) and ii) the *Upper 90% CI* described by Eq. (4). Besides, the GMR-constraint criterion of 0.80–1.25 is also considered for the estimation of the maximum and minimum GMR accepted values.

In case of the FDA_s method, when CV_{wR} is lower than 30%, the *Upper BE limit* should be set equal to 1.25, while for CV_{wR} values greater than or equal to 30%, the scaled BE limits defined by Eq. (1) should be used (Table I). For the EMA_s, a more complex procedure is proposed (Table I). The classic 0.80–1.25 limits are used when CV_{wR} is lower or equal to 30%, while for CV_{wR} variability values between 30% and 50%, scaled BE limits according to Eq. (2) should be applied. Finally, when CV_{wR} exceeds 50% the upper or lower BE limits should be set to their extreme values 1.4319 or 0.6984, respectively.

In this study, the maximum and minimum accepted GMR *versus* CV_{wR} plots were constructed in case of FDA_s and EMA_s assuming CV_{wR} values from zero to 80% and sample sizes of: 24, 36, 48, and 72 subjects.

Bioequivalence Limits

The bioequivalence criteria examined in this study include: i) the approach FDA_s proposed by the FDA scientists (27,28), ii) the newly introduced scaled procedure, EMA_s, in the EMA guideline (3), iii) a modified FDA approach where no GMR-constraint is applied (FDA_{nc}), and likewise iv) the EMA_s limits without the complementary GMR criterion which will be termed as EMA_{nc}.

Bioequivalence Simulations—Power Curves

Two-treatment, three period, three sequence crossover (3x3) bioequivalence studies, with equal number of subjects in each sequence, were simulated and evaluated using the FDA_s and EMA_s approaches. The R product was considered to be administered twice, while the T only once. Therefore, three possible sequences were derived: RTR, RRT, and TRR.

In each simulated crossover study, bioequivalence was declared if the following two conditions were satisfied: a) the 90% CI around the ratio of the geometric means for the two T and R products was between the BE limits (31) and b)

the true GMR of a simulated BE study lay between the 0.80–1.25 region. Alternatively, the method proposed by Tothfalusi and Enrendyi (23) could have been used. The latter is based on the classic 90% CI approach, but it utilizes a non-central *t*-distribution to estimate the confidence limits. In the current analysis, the first approach was used. The results derived from this method were in accordance with the data published from the FDA working group (33) and have also successfully been applied to previous works (24–26).

Log-normal distribution was assumed for the PK parameter under study. To be in agreement with the EMA guideline, this PK parameter could only refer to C_{max} and not to *AUC* (3). In all cases, the within-subject variability of the T was assumed to be equal to that of R product. Several levels of theoretical CV_{wR} values were considered for the simulations: 20%, 40%, 50%, and 70%. Each of these CV_{wR} values was selected to reflect a different situation that can potentially be encountered in practice: i) in the case of low variability (20%), all approaches (FDA_s, EMA_s, FDA_{nc}, and EMA_{nc}) are theoretically identical and no differences would be expected, ii) when CV_{wR} is set equal to 40%, all methods are based on scaling criteria which, however, utilize different expansion factors (3,27,28), iii) the 50% variability was selected since it refers to the point where the EMA_s and EMA_{nc} limits are transformed into the constant 0.6984 or 1.4319 limits, and finally iv) the CV_{wR}=70%, is used to assess the performance of the different methods when the drug exhibits high within-subject variability.

The entire analysis was done assuming several levels of sample size, such as: 18, 24, 30, 36, 48, 60, 72, 84, 96, 108, 120, and 150. Plausibly, some combinations of the above-mentioned conditions may not seem realistic; for example, no BE study with 70% variability would be conducted, if it included only 18 subjects. However, such types of cases were added in the simulations only for comparative and completeness purposes. Besides, this will allow us to unveil any possible trends of the behavior of the different BE criteria.

The theoretically true GMR value was gradually changed, from 1.00 to 1.60 using a step of 0.05. Under each condition, 40,000 BE trials were simulated and the percentage of accepted studies was recorded. This allowed the construction of *Power curves* by plotting the % acceptance values as a function of GMR. According to the EMA guideline (3), the scaling approach can only be applied to C_{max}. Thus, strictly speaking the power estimates shown in this study actually refer to the case of C_{max}. The entire programming work was implemented by developing the appropriate functions in MATLAB® (The MathWorks, Inc).

It is worth mentioning that the FDA_s and EMA_s are based on switching variability values (e.g., 30%) to define their BE limits. This implies that at regions close to the switching values, some simulated studies will follow one criterion and some others a different criterion. For example, close to the CV_{wR}=30%, the FDA_s and EMA_s limits will either be constant 0.80–1.25 or follow their scaled criterion defined by Eq. (1) and (2), respectively. Since Monte Carlo simulations were used to generate the data, the behavior which is depicted in the power curves as a single point, actually, reflects the overall performance.

RESULTS

Figure 1 presents the newly proposed BE limits of FDA_s and EMA_s as a function of within-subject variability of the R product. Despite the fact that both FDA_s and EMA_s include scaling with s_{wR}, these two methodologies exhibit a different pattern of widening. In case of the FDA_s approach (Fig. 1a), constant limits (0.80–1.25) are set for CV_{wR} values up to 30%. After this switching variability value, a scaled BE criterion is applied which is described by Eq. (1). It should be highlighted that the FDA_s limits endlessly expand with variability.

A deeper inspection of the FDA_s limits (Fig. 1a) reveals the discontinuity that exists at CV_{wR}=30%, an issue which was extensively discussed by Endrenyi and Tothfalusi (37,38). In Fig. 1a, this discontinuity is shown as a small "*jump*" from the 1.25 (or 0.80) value at CV_{wR}=30%. The reason for this abnormality arises from the fact that the preset regulatory standardized variation was set to be s_{w0}=0.25, namely, lower than the switching variability value 30% (27). In addition, it is important to note the fact that a CV_{wR} value equal to 30% does not exactly correspond to a 0.30 value for s_{wR}, but to a value of 0.2935, since CV_{wR} and s^2_{wR}, are linked by the relationship: $CV_{wR} = \sqrt{e^{s^2_{wR}} - 1}$. The specific assignment of values, made by FDA, for the switching variability and for s_{w0}, makes the BE limits discontinuous at CV_{wR}=30% value. However, this abnormality could have been avoided if, for example, these two values were set at the same level (37).

The behavior of the EMA_s limits is depicted in Fig. 1b. The EMA_s limits follow a pattern similar to that of the FDA_s at low variability values. However, EMA_s becomes very different from FDA_s at high variabilities. The EMA_s limits include two switching variation values: one at CV_{wR}=30% and another at 50%. When CV_{wR} is lower than 30% or greater than 50%, the upper BE limit refers to the horizontal line at 1.25 or 1.4319, respectively. In the same manner, the lines at 0.80 and 0.6984 refer to the lower limits. Besides, when CV_{wR} lies between 30% and 50%, an exponential increasing segment exists, according to Eq. (2).

Figure 2 presents the maximum and minimum GMR accepted values as a function of CV_{wR} in case of FDA_s (Fig. 2a) and EMA_s (Fig. 2b) assuming sample sizes of 24, 36, 48, and 72. As shown by these plots, two drug products can be declared bioequivalent if the corresponding GMR

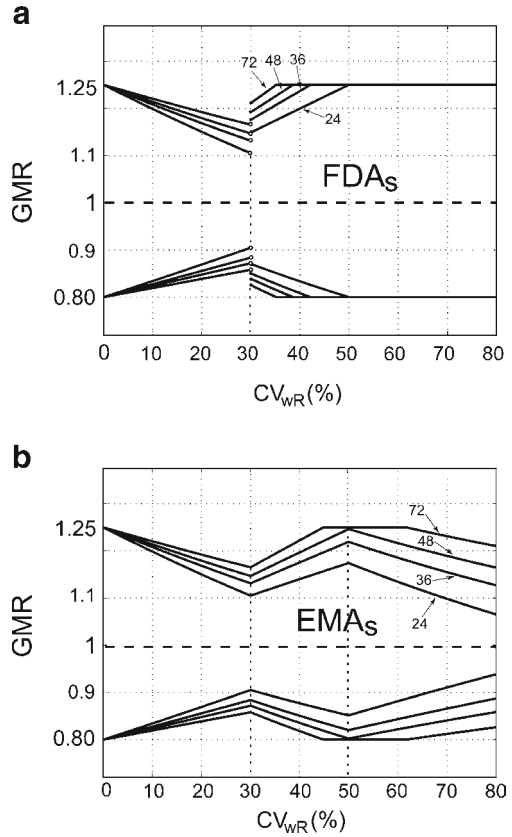

Fig. 2 Maximum and minimum accepted GMR values, for the FDA_s and EMA_s limits, as a function of within-subject variability. Each curve corresponds to a different number of subjects: 24, 36, 48, and 72. The "discontinuity" of the FDA_s limits becomes obvious at CV_{wR}=30%.

value is between its upper and lower boundary. Visual inspection and comparison of Fig. 2a with 2b reveals that initially FDA_s and EMA_s exhibit similar behavior, but after the switching point 30% a discrepancy in the performances becomes obvious. When CV_{wR} is lower than 30%, no scaling exists for FDA_s and EMA_s since both BE limits are equal to 0.80 or 1.25. In Fig. 2, this attribute is depicted as the first narrowing part of the curves.

However, after the CV_{wR}=30% value, scaled BE limits take effect for FDA_s and EMA_s, which impose the range of the acceptable GMR values to gradually increase (Fig. 2). However, in the case of EMA_s, this expansion does not last forever, but stops when CV_{wR} values approach the 50% variation. In case of FDA_s limits, a straight line at 1.25 suddenly becomes apparent close to CV_{wR}=50% (Fig. 2a). This finding can be attributed to the existence of the secondary GMR-constraint criterion which becomes effective. This is due to the fact that at these high variability values, the FDA_s limits have become very liberal due to scaling according to Eq. (1). The GMR-constraint, then, gets stricter and exerts the dominate role in the determination of bioequivalence. In other words, the GMR criterion imposes the extreme GMR values to be at the 1.25 (maximum) or the 0.80 (minimum) values. As the number of subjects participating in the study increases, the effect of GMR-constraint becomes more pronounced (Fig. 2a). In such cases, the boundary lines at 1.25 and 0.80 appear earlier, i.e., at lower CV_{wR} variabilities.

Another point of special importance is how the "discontinuity" of FDA_s limits is reflected on the extreme GMR *versus* CV_{wR} plots. Actually, Fig. 2a reveals that again a "*jump*" becomes apparent at the switching value 30%. In other words, the sudden widening of BE limits at CV_{wR}=30% (Fig. 1a) leads to an abrupt expansion of the extreme GMR values.

A different behavior is evident for EMA_s (Fig. 2b). In this case, the role of GMR-constraint is limited only at regions around CV_{wR}=50%. The underlying reason for this attribute should be due to the nature of the EMA_s procedure. The EMA_s limits scale with s_{wR}, but only up to CV_{wR}=50%. Thus, for greater variability values, the EMA_s limits remain constant to the values achieved at CVwR=50% (i.e., 1.4319 or 0.6984). This attribute imposes the EMA_s limits to be "self-confined" for high variability values. Therefore, the GMR-constraint appears to only be effective at CV_{wR} values around 50%. Again, as sample size increases, the role of the secondary GMR criterion is enhanced and can be observed at wider regions around CV_{wR}=50% (Fig. 2b).

Figure 3 shows the percentage of simulated studies in which BE is accepted *versus* the GMR of the study. The within-subject variability was assumed to be equal to 20%, while the sample size was equal to 24 (Fig. 3a) and 48 (Fig. 3b). Four different approaches are depicted in each plot: FDA_s, EMA_s, FDA_{nc}, and EMA_{nc}. Visual inspection of Fig. 3a reveals that all data are superimposed upon each other, namely, all limits exhibit the same ability to declare bioequivalence when CV_{wR} is 20%. In case of Fig. 3b, as the number of subjects increases, the percentages of acceptance also rise. Nevertheless, the relative performance of the four methods (FDA_s, EMA_s, FDA_{nc}, and EMA_{nc}) remains the same.

The comparative performance of FDA_s, EMA_s, FDA_{nc}, and EMA_{nc} is further depicted in Fig. 4, but in case of higher variability values, such as 40%, 50%, and 70%. Figure 4a reveals a discrepancy in the performance of FDA_s and EMA_s when CV_{wR} becomes equal to 40%; the FDA_s method leads to higher % acceptances than EMA_s. As the number of subjects increases (Fig. 4b), both methods become more permissive, but the difference in their abilities to declare BE is preserved.

In addition, comparison of Fig. 4a and b unveils an interesting role of the GMR constraint. At low sample sizes, the GMR criterion seems to be ineffective; FDA_s and EMA_s behave identically to their equivalents without the constraint, i.e., FDA_{nc} and EMA_{nc}, respectively (Fig. 4a). However, when a large sample size is assumed (as in case of Fig. 4b), then the GMR constraint becomes effective for the FDA_s method. Indeed, the approach without the GMR criterion (FDA_{nc})

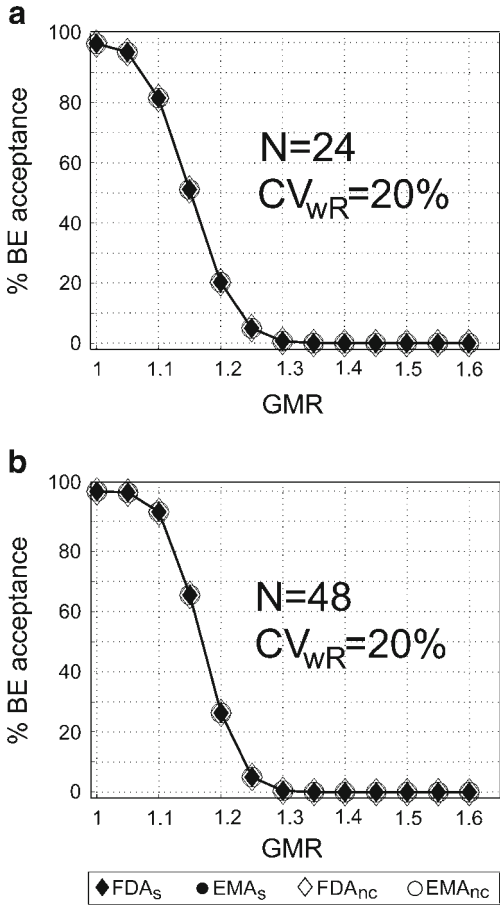

Fig. 3 Percentage of 3 × 3 bioequivalence studies accepted, by four different approaches: FDA_s, EMA_s, FDA_{nc}, and EMA_{nc} (see "Materials and Methods") at various GMR values. The within-subject variability of the R product was set equal to 20%, while sample size was assumed to be 24 (**a**) and 48 (**b**).

exhibits much higher % acceptance values than FDA_s. In case of the EMA limits (EMA_s and EMA_{nc}), no role of the GMR-constraint can be observed at this level of variability.

When CV_{wR} gets equal to 50%, the performance of FDA_s further deviates from that of EMA_s (Fig. 4c). In other words, the approach proposed by FDA becomes much more liberal comparing to the EMA_s, as within-subject variability increases. The impact of GMR-constraint on the FDA_s limits is now apparent even in case of a low sample size (as in Fig. 4c). When more subjects are recruited in the study, the performance of EMA_s is closer to that of FDA_s (Fig. 4d). Actually, this attribute starts to become evident when the number of subjects is about 54.

Figure 4d also presents an interesting situation for the EMA_s approach. When CV_{wR} is in the region of 50% and a large number of subjects is used, the EMA_s limits, as described by Eq. (1), reach their most liberal value (in regard to variability). In turn, this triggers the secondary GMR-constraint to become effective. In other words, Fig. 4d depicts an uncommon situation where the GMR-constraint can be of value for the EMA_s approach.

A further increase of variability results in a higher discrepancy between the performances of FDA_s and EMA_s (Fig. 4e). In cases of very high within-subject variabilities, the scaled BE limits proposed by the FDA working group are very liberal compared to the scaled limits of EMA's guideline. As sample size increases, the permissiveness of EMA_s becomes similar to that of FDA_s (Fig. 4f). In such cases, the role of GMR criterion is very prominent for FDA_s, but it has no impact on the EMA_s. If not such a criterion was applied to the FDA_s, then the permissiveness would be very high. For example, even for two drug products which differ by 50% in their PK values (i.e., GMR = 1.50), the probability to be declared bioequivalent with the FDA_{nc} (namely, with no GMR constraint) would be more than 50% (Fig. 4f).

As it is already quoted in the "Materials and Methods" section, a sample size of 24 subjects for a 70% variability might be considered unrealistic. However, these kinds of results are presented in this analysis only for reasons of comparison in order to detect any potential trends. Further results for the % of studies accepted as a function of GMR for the recently introduced FDA_s and EMA_s approaches are presented in the Appendix.

DISCUSSION

An aim of this study was to explore the comparative performance of the newly proposed approaches FDA_s and EMA_s Table I (3,27). In order to accomplish this task, BE studies were simulated by using different levels of variability and sample size. The comparative performance of FDA_s and EMA_s is shown in Fig. 1. At low within-subject variabilities (i.e., when CV_{wR} does not exceed 30%), both methods are based on the classic 0.80–1.25 limits. However, after this point forward, the FDA_s limits scale limitlessly with variability. This attribute is depicted in Fig. 1a as a continuous expansion of the limits. On the contrary, the EMA_s scale with variability, but only up to a maximum variability value (50%), after which they remain constant and equal to the extreme values 1.4319 or 0.6984 (Fig. 1b).

In turn, the different properties of the FDA_s and EMA_s methods are reflected on the extreme GMR plots (Fig. 2). When CV_{wR} is lower than 30%, the extreme GMR acceptance range of FDA_s is depicted as a shrinkage, since the classic 0.80–1.25 limits are effective (Fig. 2a). After the 30% point, a continuous expansion, according to Eq. (1), is observed for FDA_s. However, this spreading out of the GMR acceptance range is suddenly interrupted due to the activation of the secondary GMR constraint. Without this constraint, the FDA_s limits would continuously increase, thus, allowing two drug products with large differences in their GMR to be declared bioequivalent.

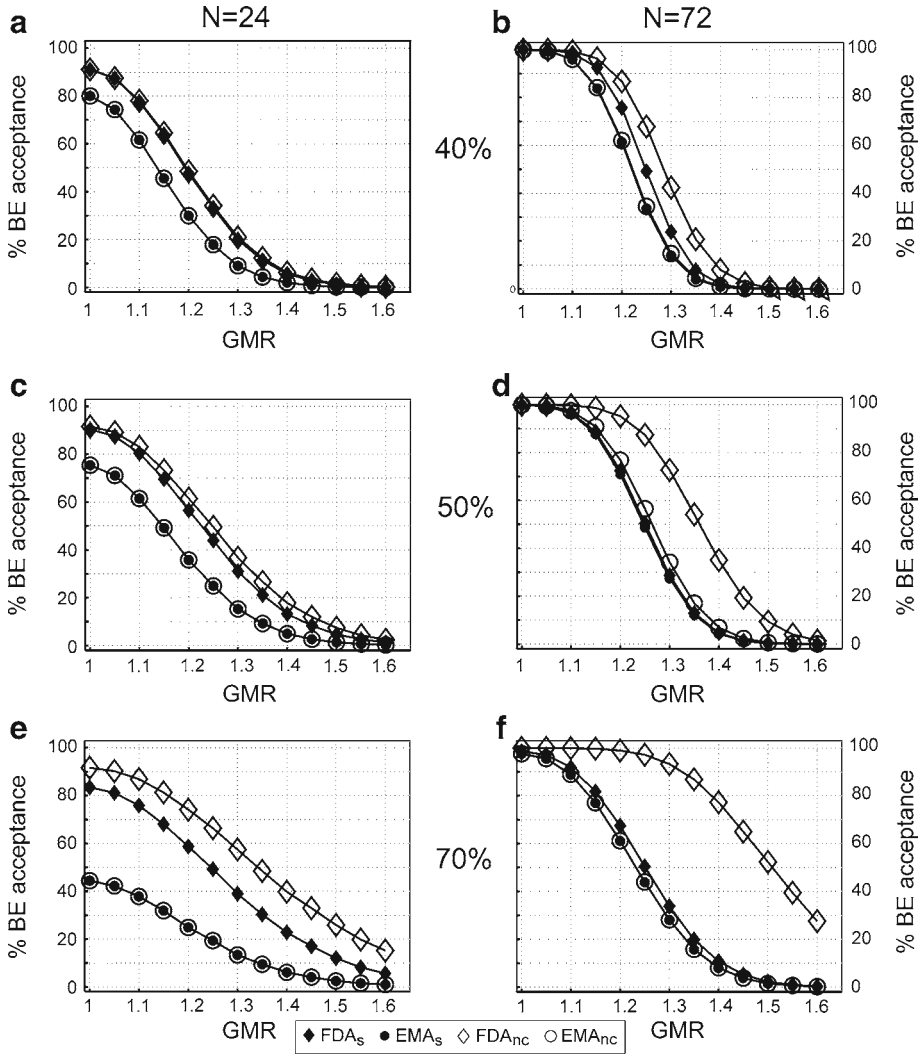

Fig. 4 Percentage of 3 × 3 bioequivalence studies accepted, by four different approaches: FDA_s, EMA_s, FDA_{nc}, and EMA_{nc} (see "Materials and Methods") at various GMR values. Three levels of within-subject of the R product were assumed: 40%, 50%, and 70%. Sample size was set equal to 24 and 72.

In case of the EMA_s limits (Fig. 2b), a reduction of the GMR acceptance range is initially observed, which is again followed by an extension after CV_{wR}=30%. However, this expansion of the GMR accepted range is now self-limited. This is due to the fact that the EMA_s limits do not scale continuously with variability, but for CV_{wR}>50% are confined to the extreme values 1.4319 or 0.6984. In case of the EMA_s limits, the impact of the complementary GMR criterion only becomes evident when a large number of subjects is included in the BE study (e.g., 72) and CV_{wR} is close to the value 50%. The GMR-constraint, which takes effect when the EMA_s limits become highly liberal, is depicted as the flat part of the curves in Fig. 2b. Nevertheless, at higher CV_{wR} variability values, the extreme GMR acceptance range for EMA_s does not stay constant, as in Fig. 2a, but it vanishes monotonously and converges towards unity (Fig. 2b). This finding can be attributed to the leveling-off properties of the EMAs limits, since for CV_{wR} values greater than 50%, the EMA_s limits are equal to the boundary values 0.6984 and 1.4319.

The results derived from power curves in Fig. 3 are in concordance with the findings already discussed in cases illustrated by Figs. 1 and 2. When CV_{wR} was set equal to 20%, namely, lower than the switching value 30%, all approaches exhibit identical performances, since at this level of variability the BE limits of all methods are equal to 1.25 or 0.80. However, as within-subject variability increases (Fig. 4), the FDA_s method is getting more liberal than the EMA_s. When CV_{wR} reaches the value of 40%, the scaled criteria of FDA_s and EMA_s become effective, in accordance with Eq. (1) and (2), respectively. However, the scaling factor of FDA_s is $\ln(1.25)/0.25 \approx 0.893$, while the corresponding value for EMA_s is 0.760. Therefore, a greater expansion rate with s_{wR} is anticipated and observed for FDA_s.

When CV_{wR} becomes equal to 50%, the effect of the different scaling factors becomes more pronounced. This is reflected on the percent values of acceptances for FDA_s and EMA_s. It is quite interesting to note that at this level of variability (CV_{wR}=50%) as few as 24 subjects seem to provide an 80% power of a BE study when the two products

differ by 10% (i.e., GMR=1.10) and are assessed with the FDA_s approach (Fig. 4c).

At higher variability values, for example 70%, the FDA_s and EMA_s exhibit a totally different performance when relatively few subjects are recruited in the BE study (Fig. 4e). In such cases, the probability of declaring bioequivalence with the FDA_s approach is considerably higher than the one with EMA_s. In other words, the FDA_s limits have now become very liberal due to their continuous expansion with s_{wR}. In case of the EMA_s limits, scaling had terminated earlier at CV_{wR}=50% and no further widening of the BE limits is now allowed. After the 50% variability value, the bioequivalence limits stay constant (1.4319 or 0.6984). Therefore, for any further increase of variability, it will be more difficult for two drug products to be declared bioequivalent with the EMA_s approach.

In general, an increase of sample size results in an enhanced ability to declare bioequivalence. This attribute is evident for both FDA_s and EMA_s approaches shown in Figs. 3 and 4. Obviously, this is anticipated because it is easier to declare bioequivalence when more subjects are recruited in a BE study. However, it should be highlighted that the increase of the percentages of acceptance is more prominent for EMA_s than for the FDA_s method. Thus, when more than 54 subjects are included in a BE study, the performance of EMA_s becomes very similar or even identical to that of FDA_s (Fig. 4b, d, and f).

According to the results shown in Fig. 4, the sample size also exerts another important role. As the number of subjects enrolled in the BE study is increased, the impact of the GMR-constraint on BE acceptance becomes more significant. For FDA_s, this finding is clearly apparent at all these conditions (Fig. 4b, d, f). Nevertheless, this attribute becomes evident for EMA_s only at regions close to the switching value of CV_{wR}= 50% (Fig. 4d). Presumably, as sample size rises, the width of the region around 50% gets wider.

However, one can argue that even though these major differences between the FDA_s and EMA_s are theoretically valid, in actual practice only minor differences can be detected. The basis of this argument relies on the fact that the greatest discrepancy, between FDA_s and EMA_s, is observed for CV_{wR} values higher than 50% (Fig. 2). Nevertheless, it is not impossible that a drug may exhibit within-subject variability even higher than 70%. For example, following oral administration of a risedronate tablet, the reported within-subject variability values were 74.3% and 82.2% for AUC and C_{max}, respectively (39). Thus, the discrepancy in the ability to declare bioequivalence would be even greater than the one shown in Fig. 4e and f. In other words, it would be very possible to encounter the situation where a risedronate tablet would be declared bioequivalent to an innovator's product following the FDA_s approach, but not if the applicant followed the EMA_s procedure.

The role of GMR-constraint regarding the FDA_s approach was discussed by Endrenyi and Tothfalusi (37). However, in the current analysis, the focus was on the relative impact of this secondary GMR-criterion on the performances of both FDA_s and EMA_s. Based on these results, it can be concluded that the effect of GMR-constraint exerts a dominant role in determining power/sample size for the FDA_s approach when the drug under study exhibits high within-subject variability (Figs. 2 and 4). As CV_{wR} and especially when sample size increase, the impact of GMR-constraint on the BE acceptance of FDA_s becomes very prominent (Fig. 4d and f). Without the GMR-constraint in the FDA method, it is possible (over 50% probability) for two drug products with CV_{wR}=70% to be declared bioequivalent even when they differ 50% in their mean measure of bioavailability (Fig. 4f). In case of the EMA_s method, the role of GMR-constraint can only be observed under certain conditions due to the "*leveling-off*" properties of the new EMA_s bioequivalence limits (40). In particular, its impact becomes apparent only when CV_{wR} is in the region of 50% and a large number of subjects (e.g., greater than 60) is used (Fig. 4d).

Undoubtedly, the newly proposed methods FDA_s and EMA_s offer the opportunity to increase the power of the study without recruiting many subjects. However, it should be stressed that emphasis should not only be paid to the number of subjects, but also to the total human exposure to drugs. Even though, the FDA_s and EMA_s approaches require reduced sample sizes, they are based on replicate designs (e.g., the 3x3). In other words, the FDA_s and EMA_s include more periods of drug administration and, therefore, lead to an increased exposure of humans to drugs (41). Nevertheless, at high within-subject variability values, the new FDA_s and EMA_s methods are advantageous since they require fewer subject exposures.

CONCLUSION

The aim of this study was to compare the performances of the recently proposed methodologies by FDA and EMA for the assessment of BE in case of highly variable drugs (3,27,28). The relative performance of FDA_s and EMA_s was eloquently proven from both the extreme GMR plots and the power curves results. Some basic conclusions of this analysis are itemized as follows:

i) The FDA_s and EMA_s approaches are in essence *identical* only when variability is lower than 30%. Also, these methods can result in *similar* performances when a large number of subjects is used.
ii) In most of the cases, the FDA_s method is more permissive than EMA_s. This behavior becomes more evident as CV_{wR} increases. After CV_{wR}=30%, both scale with s_{wR}, but FDA_s includes a greater scaling factor than EMA_s, which gradually becomes more permissive. The major discrepancy, between FDA_s and EMA_s, is observed for

CV_{wR} values higher than 50% and N roughly lower than 50. This is due to the fact that the FDA_s limits continue to expand with variability, whereas the EMA_s exhibit leveling-off properties and are confined to their extreme values 1.4319 or 0.6984.

iii) The secondary GMR-constraint was found to be essential for FDA_s, especially in case of high variability. In such cases, if no GMR-constraint was in effect, the permissiveness of the FDA_s limits would be extremely high. However, the role of GMR-constraint is of less importance for the EMA_s. Its effect becomes evident only when a large sample size is used and the CV_{wR} of the drug is close to 50%. For example, the impact of GMR-constraint starts to become apparent when CV_{wR} is equal to 50% and 60 subjects are enrolled in the BE study.

iv) The increase of sample size affects the EMA_s limits more than does to FDA_s.

v) Plots of % BE acceptance *versus* GMR, for several sample sizes and variabilities, were constructed. These plots can guide those wishing to conduct three-period, three-sequence (3x3) bioequivalence studies using either FDA_s or EMA_s approach.

ACKNOWLEDGMENTS & DISCLOSURES

The authors wish to thank the reviewers for their constructive critique and helpful comments which improved the quality of this manuscript.

APPENDIX

Three period, three sequence crossover (3x3) bioequivalence studies, with equal number of subjects in each sequence, were simulated using the FDA_s and EMA_s approaches. The design of these studies was: TRR, RTR, RRT. In each simulated BE study, determination of bioequivalence was based on the 90% CI around the GMR of T and R drugs. Several levels of sample size were assumed: 18, 24, 30, 36, 48, 60, 72, 84, 96, 108, 120, and 150 subjects. Within-subject variability of the R was set equal to T. For the simulations, several levels of variability were considered: 20%, 40%, 50%, and 70%. The percentages of acceptance, for the FDA_s (27) and EMA_s (3) methods, were recorded and plotted as a function of GMR.

These power curves are shown in Figs. 5 and 6 for FDA_s and EMA_s, respectively. Only the region where the GMR lies between 0.80 and 1.25 is depicted in Figs. 5 and 6. This range of GMR was deliberately chosen, since it corresponds to the area defined by the complementary constraint criterion on GMR.

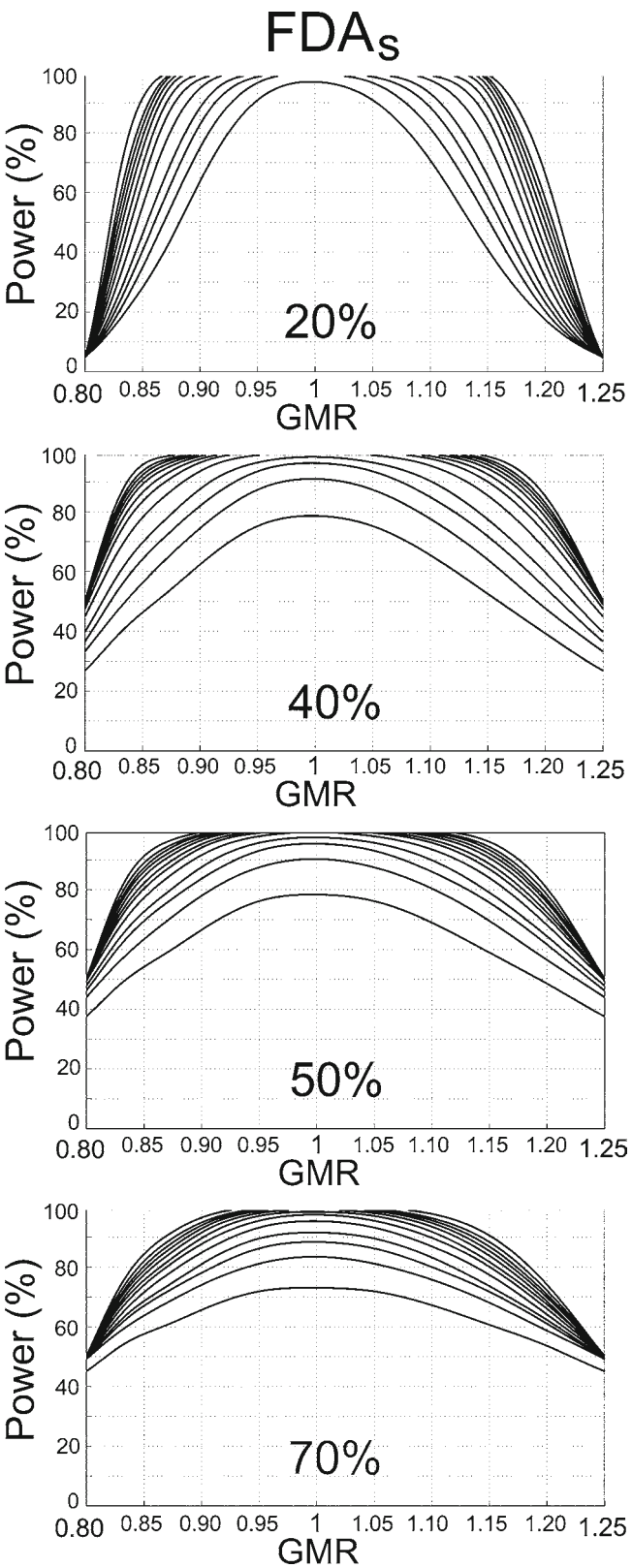

Fig. 5 Percent of studies accepted as a function of GMR for the FDA_s approach (27). Sample size (from bottom to top) is: 18, 24, 30, 36, 48, 60, 72, 84, 96, 108, 120, and 150. Four levels of within-subject variability are shown: 20%, 40%, 50%, and 70%.

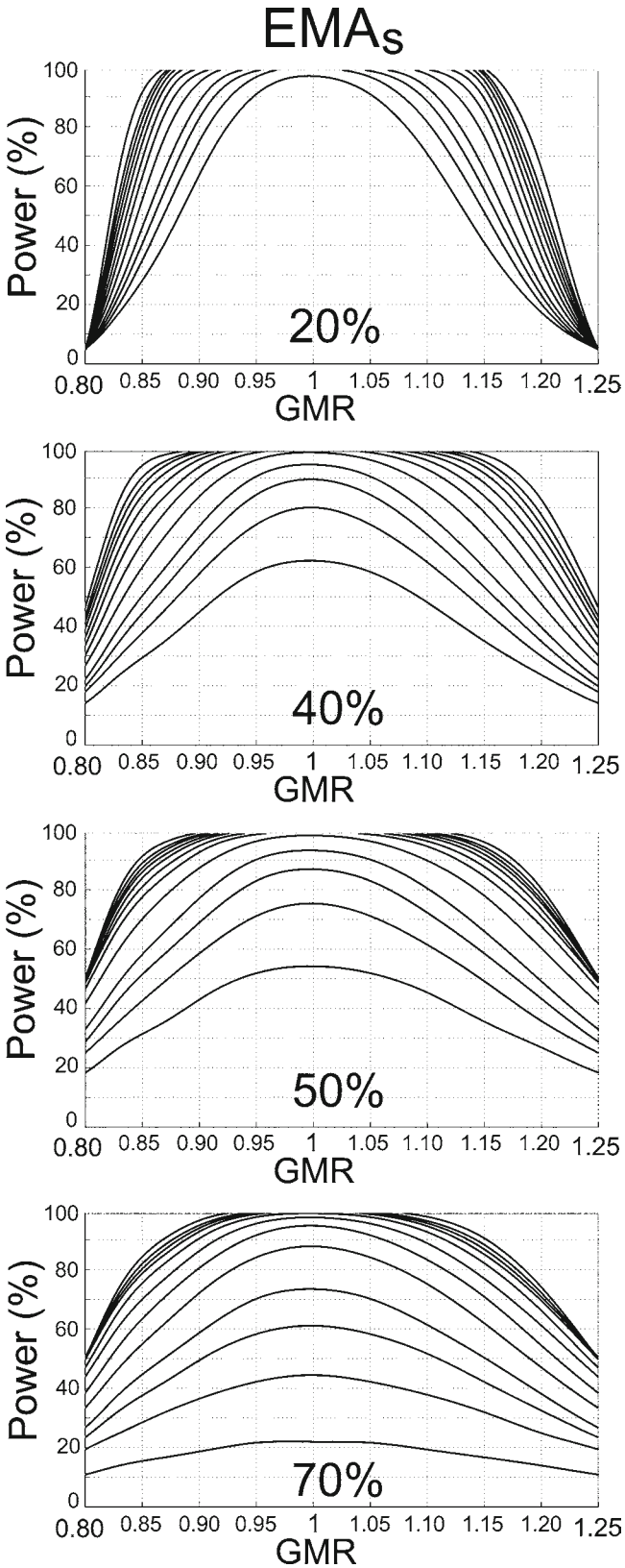

Fig. 6 Percent of studies accepted as a function of GMR for the EMA_S approach (3). Sample size (from bottom to top) is: 18, 24, 30, 36, 48, 60, 72, 84, 96, 108, 120, and 150. Four levels of within-subject variability are shown: 20%, 40%, 50%, and 70%.

REFERENCES

1. FDA (Food and Drug Administration). Center for Drug Evaluation and Research (CDER), Bioavailability and Bioequivalence Studies for Orally Administered Drug Products. General Considerations, Rockville, MD. 2003.
2. EMA (European Medicines Agency). Evaluation of Medicines for Human Use, CPMP. Note for Guidance on the Investigation of Bioavailability and Bioequivalence, London. 2001.
3. EMA (European Medicines Agency). Committee for Medicinal Products for Human Use, CHMP. Guideline on the Investigation of Bioequivalence, London. 2010.
4. Blume H, Midha K. Bio-International '92, Conference on Bioavailability, Bioequivalence and Pharmacokinetic Studies. J Pharm Sci. 1993;82:1186–9.
5. Blume H, Elze M, Potthast H, *et al.* Practical strategies and design advantages in highly variable drug studies: multiple dose and replicate administration design. In: Blume H, Midha K, editors. Bio-international 2: Bioavailability, Bioequivalence and Pharmacokinetic studies. Stuttgart: Medpharm Scientific Publishers; 1995. p. 117–22.
6. Midha K, Shah V, Singh G, Patnaik R. Conference report: Bio-International. J Pharm Sci. 2007;96:747–54.
7. Shah V, Yacobi A, Barr W, *et al.* Evaluation of orally administered highly variable drugs and drug formulations. Pharm Res. 1996;13:1590–4.
8. Van Peer A. Basic Variability and impact on design of bioequivalence studies. Clin Pharmacol Toxicol. 2010;106:146–53.
9. Benet L. Bioavailability and bioequivalence: definitions and difficulties in acceptance criteria. In: Midha K, Blume H, editors. Bio-International: bioavailability, bioequivalence and pharmacokinetics. Stuttgart: Medpharm Scientific Publishers; 1995. p. 27–35.
10. Benet L. Individual bioequivalence: an overview. AAPS International Workshop on Individual Bioequivalence: Realities and Implementation, Montreal, Quebec. 1999. Aug.30–Sep.1.
11. Midha K, Rawson M, Hubbard J. The bioequivalence of highly variable drugs and drug products. Int J Clin Pharmacol Ther. 2005;43:485–98.
12. Anderson S, Hauck W. Consideration of individual bioequivalence. J Pharmacokinet Biopharm. 1990;18:259–73.
13. Endrenyi L, Amidon G, Midha K, *et al.* Individual bioequivalence: attractive in principle, difficult in practice. Pharm Res. 1998;15:1321–5.
14. FDA (Food and Drug Administration). Center for Drug Evaluation and Research (CDER), Statistical Approaches to Establishing Bioequivalence. Rockville, MD. 2001.
15. Patnaik R, Lesko L, Chen ML, Williams R. Individual bioequivalence: new concepts in the statistical assessment of bioequivalence metrics. Clin Pharmacokinet. 1997;33:1–6.
16. Schall R, Williams R. Towards a practical strategy for assessing individual bioequivalence. J Pharmacokinet Biopharm. 1996;24:133–49.
17. Midha K, Rawson M, Hubbard J. Individual and average bioequivalence of highly variable drugs and drug products. J Pharm Sci. 1997;86:1193–7.
18. EMA (European Medicines Agency). Evaluation of Medicines for Human Use, CHMP efficacy working party therapeutic subgroup on pharmacokinetics: Questions & Answers on the Bioavailability and Bioequivalence Guideline, London. 2006.
19. Hauck L, Parekh A, Lesko L, *et al.* 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. 2001;39:350–5.
20. Tothfalusi L, Endrenyi L, Midha L. Scaling or wider bioequivalence limits for highly variable drugs and for the special case of C_{max}. Int J Clin Pharmacol Ther. 2003;41:217–25.

21. Boddy A, Snikeris F, Kringle R, *et al.* An approach for widening the bioequivalence acceptance limits in the case of highly variable drugs. Pharm Res. 1995;12:1865–8.
22. Midha K, Rawson M, Hubbard J. Bioequivalence: switchability and scaling. Eur J Pharm Sci. 1998;6:87–91.
23. Tothfalusi L, Endrenyi L. Limits for the scaled average bioequivalence of highly variable drugs and drug products. Pharm Res. 2003;20:382–9.
24. Karalis V, Symillides M, Macheras P. Novel scaled average bioequivalence limits based on GMR and variability considerations. Pharm Res. 2004;21:1933–42.
25. Karalis V, Macheras P, Symillides M. Geometric Mean Ratio–dependent scaled bioequivalence limits with leveling-off properties. Eur J Pharm Sci. 2005;26:54–61.
26. Kytariolos J, Karalis V, Macheras P, Symillides M. Novel scaled bioequivalence limits with leveling-off properties based on variability considerations. Pharm Res. 2006;23:2657–64.
27. Haidar S, Davit B, Chen ML, Conner D, Lee L, Li Q, Lionberger R, Makhlouf F, Patel D, Schuirmann D, Yu L. Bioequivalence approaches for highly variable drugs and drug products. Pharm Res. 2008;15:237–41.
28. Haidar S, Makhlouf F, Schuirmann D, Hyslop T, Davit B, Conner D, Yu L. Evaluation of a scaling approach for the bioequivalence of highly variable drugs. AAPS J. 2008;10:450–4.
29. FDA (Food and Drug Administration). Office of Generic Drugs, Draft Guidance for Industry on Bioequivalence Recommendations for Progesterone Capsules (February 2011). Available from: http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM209294.pdf.
30. Davit B. Highly Variable Drugs: Reference-scaled average bioequivalence and sequential design studies. AAPS Workshop on Facilitating Oral Product Development and Reducing Regulatory Burden through Novel Approaches to Assess Bioavailability/Bioequivalence, Washington DC, 2011, October 23.
31. Schuirmann DJ. A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. J Pharmacokinet Biopharm. 1987;15:657–80.
32. Davit B. Highly variable drugs—bioequivalence issues: FDA proposal under consideration. Meeting of FDA Committee for Pharmaceutical Science, Rockville, MD. 2006, October 6.
33. Haidar S. Evaluation of the scaling approach for highly variable drugs. Meeting of FDA Committee for Pharmaceutical Science, Rockville, MD. 2006, October 6.
34. Davit B. Highly variable drugs—bioequivalence issues: FDA proposal under consideration. AAPS/FDA Workshop on BE, BCS, and Beyond, North Bethesda, MD. 2007, May 22.
35. Haidar S. BE for highly variable drugs—FDA perspective. AAPS/FDA Workshop on BE, BCS, and Beyond, North Bethesda, MD. 2007, May 22.
36. Morais J, Lobato Mdo R. The new European Medicines Agency guideline on the investigation of bioequivalence. Basic Clin Pharmacol Toxicol. 2010;106:221–5.
37. Endrenyi L, Tothfalusi L. Regulatory conditions for the determination of bioequivalence of highly variable drugs. J Pharm Pharmaceut Sci. 2009;12:138–49.
38. Tothfalusi L, Endrenyi L, Arieta AG. Evaluation of bioequivalence for highly variable drugs with scaled average bioequivalence. Clin Pharmacokinet. 2009;48:725–43.
39. Mitchell D, Barr W, Eusebio R *et al.* Risedronate pharmacokinetics and intra-and inter-subject variability upon single-dose intravenous and oral administration. Pharm Res. 2001;18:166–70.
40. Karalis V, Symillides M, Macheras P. On the leveling-off properties of the new bioequivalence limits for highly variable drugs of the EMA guideline. Eur J Pharm Sci. 2011;44:497–505.
41. Karalis V, Symillides M, Macheras P. Comparison of the reference scaled bioequivalence semi-replicate method with other approaches: focus on human exposure to drugs. Eur J Pharm Sci. 2009;38:55–63.