

Sample Sizes for Designing Bioequivalence Studies for Highly Variable Drugs

Laszlo Tothfalusi¹ and Laszlo Endrenyi²

¹ Semmelweis University, Department of Pharmacodynamics, Budapest, Hungary.

² University of Toronto, Department of Pharmacology and Toxicology, Toronto, ON, Canada.

Received, August 29, 2011; Revised, November 10, 2011; Accepted, December 21, 2011; Published, December 22, 2011.

ABSTRACT. – Purpose. To provide tables of sample sizes which are required, by the European Medicines Agency (EMA) and the U.S. Food and Drug Administration (FDA), for the design of bioequivalence (BE) studies involving highly variable drugs. To elucidate the complicated features of the relationship between sample size and within-subject variation. **Methods.** 3- and 4-period studies were simulated with various sample sizes. They were evaluated, at various variations and various true ratios of the two geometric means (GMR), by the approaches of scaled average BE and by average BE with expanding limits. The sample sizes required for yielding 80% and 90% statistical powers were determined. **Results.** Because of the complicated regulatory expectations, the features of the required sample sizes are also complicated. When the true GMR = 1.0 then, without additional constraints, the sample size is independent of the intrasubject variation. When the true GMR is increased or decreased from 1.0 then the required sample sizes rise at above but close to 30% variation. An additional regulatory constraint on the point estimate of GMR and a cap on the use of expanding limits further increase the required sample size at high variations. Fewer subjects are required by the FDA than by the EMA procedures. **Conclusions.** The methods proposed by EMA and FDA lower the required sample sizes in comparison with unscaled average BE. However, each additional regulatory requirement (applying the mixed procedure, imposing a constraint on the point estimate of GMR, and using a cap on the application of expanding limits) raises the required number of subjects.

This article is open to **POST-PUBLICATION REVIEW**. Registered readers (see "For Readers") may **comment** by clicking on ABSTRACT on the issue's contents page.

INTRODUCTION

The evaluation of bioequivalence (BE) for highly-variable (HV) drugs and drug products frustrated the pharmaceutical industry for many years. It was difficult to demonstrate BE unless distressingly many subjects were included in the investigations (1).

Major regulatory authorities have sought remedies in recent years to the problem. The Committee for Medicinal Products for Human Use (CHMP) of the European Medicines Agency (EMA) issued in 2010 a substantially revised BE guideline (2) which included new approaches for the determination of BE for HV drugs. For the same purpose, members of a working group of the Food and Drug Administration (FDA) in the United States published a procedure (3).

The approaches of both EMA and FDA are based on the method of scaled average bioequivalence (SABE) (4) or its close variant. Nevertheless, there are meaningful deviations

between the implementations of the two procedures which will be considered below.

A practical issue arising from the new approaches involves the design of BE studies. Their main goal is to lower required sample sizes in comparison with the expectations of the customarily applied unscaled average bioequivalence (ABE). Sample sizes required by the use of ABE have been published (5-9). However, sample sizes are not available which are needed for using the new procedures.

The main purpose of the present communication is to provide sample sizes which will be useful for designing BE studies for highly variable drugs. These will be presented in four tables in the Appendix.

Corresponding Author: Laszlo Endrenyi; University of Toronto Department of Pharmacology and Toxicology 1 King's College Circle, Toronto, ON, M5S 1A8, Canada; E-mail: l.endrenyi@utoronto.ca

BACKGROUND REGULATORY REQUIREMENTS

The usually applied criterion for the determination of bioequivalence implements the two one-sided tests (TOST) procedure (10). Accordingly, the average logarithmic kinetic responses of the test and reference formulations (μ_T and μ_R , respectively) are contrasted. The acceptance of bioequivalence is stated if the 90% confidence interval for the difference between the estimated logarithmic means is between preset regulatory limits. The limits (θ_A) are generally symmetrical on the logarithmic scale, and usually equal $\ln(1.25)$. Consequently, the criterion for the determination of average bioequivalence (ABE) is, schematically:

$$-\theta_A \leq \mu_T - \mu_R \leq \theta_A \quad [1]$$

In a bioequivalence study, the individual kinetic responses are evaluated from the measured concentrations. The means of the logarithmic responses of the two formulations (m_T and m_R) are calculated; these sample averages estimate the true population means (μ_T and μ_R). The true values of the means are not known and therefore their estimates have to be used:

$$-\theta_A \leq m_T - m_R \leq \theta_A \quad [2]$$

Also the within-subject variance is estimated (s_w^2), in replicate-design studies with 3 or 4 periods, for each bioequivalence metric.

Expression [2] implies that, for the declaration of bioequivalence, the difference between the estimated logarithmic means, together with their 90% confidence interval, should be within the regulatory limits of $\pm\theta_A$. The succeeding expressions have the corresponding interpretation.

The method of EMA substitutes the fixed θ_A regulatory constant with limits which depend on s_w^2 :

$$-ks_w \leq m_T - m_R \leq ks_w \quad [3]$$

This is still Average Bioequivalence but with Expanding Limits (ABEL) (11). Therefore, the TOST procedure of Schuirmann can be directly applied. Based on statistical recommendations (12), EMA proposed 0.76 as the value of the regulatory constant k (2).

An alternative, related definition of the regulatory constant, preferred by some, is:

$$\sigma_0 = \ln(1.25)/k \quad [4]$$

The expectation of EMA for the value of σ_0 is 0.294.

The European guideline prescribes additional conditions and requirements:

- 1- s_w must be evaluated using the data only of the reference drug. Accordingly the within-subject variation of the reference product (s_{WR}) should be used as s_w in Eq. 3 to calculate the limits. The determination of s_{WR} requires three- or four-period crossover designs.
- 2- Apply the so-called mixed procedure for average BE (12): if the estimated within-subject coefficient of variation (CV) does not exceed 30%, apply unscaled average BE as usual. If the variation is higher than 30% (corresponding to $s_{WR} > 0.294$) then apply ABEL.
- 3- GMR constraint: the point estimate for the ratio of the geometric means (GMR), $\exp(m_T - m_R)$, must be between 0.80 and 1.25.
- 4- A cap is placed on the expansion of the limits. If the CV exceeds 50% then the limits may not expand anymore but ABE should be applied again with limits of $\pm \ln(1.43)$.

FDA uses scaled average BE (SABE) which utilizes a rearranged form of Eq. 3:

$$k \leq (m_T - m_R)/s_w \leq k \quad [5]$$

(The paper of Haidar et al. (3) presents the squared form of Eq.5.) FDA proposes that the value of the regulatory constant be $k = 0.893$ or $\sigma_0 = 0.25$ (3).

The additional conditions and requirements of the FDA procedure are (3):

- 1- Substitute s_w in Eq. 4 by the estimated within-subject variation of the reference product (s_{WR}). Evaluate s_{WR} either in 3-period crossover studies in which the reference formulation is administered twice (RRT-RTR-TRR) or in 4-period investigations.

- 2- Apply the mixed procedure for average BE.
- 3- The confidence interval for SABE cannot be calculated analytically. The FDA recommends a numerical approach based on the approximate linearization of Eq. 5 (12, 13).
- 4- Apply the GMR constraint.

FDA does not impose further restrictions such as a cap on the application of SABE. Overall, the statistical properties of the methods proposed by EMA and FDA are rather complex as a result of the additional conditions and requirements (mixed procedure, GMR constraint, and (for EMA) a cap on the limits). Furthermore, the tests required by both EMA and FDA are dependent on each other which makes the theoretical treatment very complicated. Therefore, the required sample sizes were obtained by simulations.

METHODS

Matlab (Release 2011a, MathWorks, Natick, MA) was used for simulations on a PC with Intel Core i5-2500K processor and 8Gb RAM. Under each condition, ten thousand simulations were performed. The simulated random variables followed a log-normal distribution, thereby following the multiplicative model of BE analysis (14). Three-period partial replicate (TRR-RTR-RRT) and four-period replicate (TRTR-RTRT) trials were simulated assuming zero period and sequence effects.

The same within-subject variances were assumed for the test and reference products but s_{WR} was estimated only from the data of the reference product. Power was calculated as the success rate of passing the trial as the sample size increased from 12 to 200 subjects.

The Matlab program was validated by comparing the results with our previous Fortran program (4,12,16) and with that of the results of the FDA group (15). Several other properties of the results could be predicted by mathematical reasoning. We used also this kind of mathematical reasoning for validation. Finally, sample sizes simulated for 2-period studies at 80% power were compared with those calculated by Hauschke et al. (7) (Table 1). The agreement between the simulated and calculated numbers is satisfactory.

The reported numbers represent the minimum number of volunteers when the power exceeds the desired level. That is, the results were rounded

upward and, at the given sample size, the power was at least as high as the stated level. The precision of the estimation was evaluated by running the simulations twenty times at twenty-two different conditions (different CV's, different GMR's and different designs). The expected power was either 80% or 90%. The standard deviation of the simulated powers was calculated under each condition; the mean of these standard deviations was 0.460. Thus, the precision of the power estimation is about $\pm 0.5\%$. In the worst case, this precision corresponds to a simulation error of ± 1 subject in the sample size tables. Such instability of the estimation could occur when the true power with the given sample size is very close to 80% or 90%. In this case if the simulated power is just below 80% or 90% then the number of subjects has to be increased by one in order to reach the required power.

RESULTS

Sample Sizes for Designing BE Studies

Sample sizes are presented in tables of the Appendix which are required for designing BE studies of highly-variable drugs and drug products. Tables A1 and A2 provide sample sizes with the EMA regulatory criteria whereas Tables A3 and A4 show sample sizes with the FDA criteria. Tables A1 and A3 consider the 3-period design of TRR-RTR-RRT whereas Tables A2 and A4 show results for the 4-period TRTR-RTRT design.

Sample sizes required for 80% and 90% statistical power are given within each table. The ratios of geometric means (GMR) are considered from 0.85 to 1.20. More than 200 subjects are always required when the GMR is outside this range. Within-subject variations of the reference product are shown between coefficients of variation (CV) of 30% and 80%.

All entries in these tables refer to total sample sizes in an investigation. Consequently, the number of subjects within a study sequence is either one-third or half of these figures.

Two general conclusions can be drawn. First, the 4-period design requires about 30% fewer subjects in comparison with the 3-period investigations. Second, the sample size requirement is lower with the FDA than with the EMA approach.

Table 1. Comparison of sample sizes obtained in simulations and by calculations

CV%	S i m u l a t e d			C a l c u l a t e d (7)			
	GMR	1.00	1.05	1.10	1.00	1.05	1.10
10		6	7	10	6	8	10
15		10	12	19	10	12	20
20		16	18	32	16	18	32
25		24	28	48	24	28	48
30		32	38	67	32	38	68

CV: Coefficient of variation; GMR: Ratio of geometric means

Change of Sample Size with Within-Subject Variation: Effects of Regulatory Requirements

As noted earlier, the regulatory conditions and requirements of both EMA and FDA are complicated and contain various stipulations. As a result, the change of sample sizes with increasing variation is also complicated. For instance, when GMR is removed from 1.00, the required sample size, typically, initially decreases with rising variation and later increases. Therefore, additional simulations were performed in order to elucidate the features of this behaviour. The ranges of GMR and CV were limited in these simulations.

For these comparisons, 3-period studies were chosen with the EMA conditions of ABEL analysis and a regulatory constant of $k = 0.76$ or $\sigma_0 = 0.294$ but the additional constraints were varied. Results obtained in these simulations can be easily extrapolated also to the FDA method because the essential mathematical features of the EMA and FDA procedures are the same.

Table 2 shows the required sample sizes without the additional regulatory conditions, that is, without the mixed procedure, GMR constraint and the cap on the regulatory limits. Table 2 shows that the required sample size does not change with

increasing CV when GMR = 1.00 and decreases with rising CV when GMR deviates from unity.

The explanation lies in the features of the scaled difference of the means which is shown in Eq. 4. The scaled difference follows the non-central t-distribution (12). When GMR = 1.00, the noncentrality parameter is zero. In this case, the scaled difference has a standard t-distribution. This means that the width of the confidence interval is independent of CV and depends only on the sample size. Consequently, with GMR = 1.00, the number of volunteers is also independent of CV. This is shown in Table 2 (apart of small, random fluctuations due to the simulation error).

When GMR deviates from 1.00 then the non-centrality parameter raises the upper limit of the confidence interval. The non-centrality parameter is proportional to $\log(\text{GMR})$ and reciprocally proportional to s_w . Thus, at a given sample size and $\log(\text{GMR})$, the rise in the confidence interval gets smaller as s_w (i.e., CV) increases. Correspondingly, the required sample size declines as s_w (i.e., CV) increases.

This relationship is shown in Table 2.

Table 2. Sample sizes with EMA conditions but without mixed procedure, GMR constraint, cap on the use of ABEL

CV	GMR	80% power			90% power		
		1.0	1.1	1.2	1.0	1.1	1.2
30%		26	54	>201	33	74	>201
35%		26	45	198	33	63	>201
40%		27	41	123	33	56	172
50%		26	34	73	33	47	104
60%		26	32	55	33	43	77
70%		27	30	46	33	40	64
80%		27	30	43	33	39	58

CV: Coefficient of variation; GMR: Ratio of geometric means

Similar considerations apply when the approach of ABEL is used since the power of average BE can be characterized by a bivariate noncentral t-distribution (6, 7).

Table 3 shows *the effect of including the mixed strategy*. With the mixed strategy, the method of evaluation depends on the estimated CV of the reference product. When the true CV is 30% then half of the simulated trials are evaluated by the unscaled and the other half by the scaled approach. When the true CV is higher than 30%, say 34%, then the estimated CV is still below 30% in many studies which are then evaluated by unscaled average BE. Thus, instead of using ABEL (or SABE) which would have narrower limits (“goalposts”) in this region, unscaled average BE is applied with its wider, more relaxed 0.80-1.25 limits/goalposts. Consequently, on applying the mixed strategy, the required sample size becomes lower at and near CV = 30 % (compare Tables 2 and 3).

Table 4 presents sample sizes *in the additional presence of the GMR constraint*. (These results do not assume the cap on the use of ABEL which is an expectation of EMA.) Consequently, in comparison with Table 3, the effect of including the GMR constraint can be observed. The constraint raises the required sample size when the CV is high. This is understandable since at large CV's, the chances of observing very high (or very low) GMR is high. The effect of the GMR constraint is particularly conspicuous when the true GMR deviates from 1.00.

Table 5 shows sample sizes with similar conditions as Table 4 except that the value of the regulatory constant is $k = 0.893$ or $\sigma_0 = 0.25$ and thereby *corresponds to the requirements of FDA*. The number of subjects is lower with the FDA than the EMA requirements. The deviation is meaningful at moderately high variations but diminishes at still higher CV's.

Table 3. Sample sizes with EMA conditions including mixed procedure but without GMR constraint and cap on the use of ABEL

CV	GMR	80% power			90% power		
		1.0	1.1	1.2	1.0	1.1	1.2
30%		23	46	>201	28	64	>201
35%		25	45	>201	32	63	>201
40%		26	40	124	33	55	175
50%		27	35	73	33	47	103
60%		26	32	56	33	43	78
70%		26	31	48	33	41	64
80%		27	30	42	33	39	59

CV: Coefficient of variation; GMR: Ratio of geometric means

Table 4. Sample sizes with EMA conditions including mixed procedure and GMR constraint but without cap on the use of ABEL

CV	GMR	80% power			90% power		
		1.0	1.1	1.2	1.0	1.1	1.2
30%		23	46	>201	28	63	>201
35%		26	45	>201	32	62	>201
40%		26	41	138	33	56	>201
50%		27	36	137	33	51	>201
60%		27	34	199	35	52	>201
70%		28	37	>201	37	63	>201
80%		30	38	>201	42	72	>201

CV: Coefficient of variation; GMR: Ratio of geometric means

Table 5. Sample sizes with FDA conditions including mixed procedure and GMR constraint

CV	GMR	80% power			90% power		
		1.0	1.1	1.2	1.0	1.1	1.2
30%		21	38	>201	26	55	>201
35%		21	31	103	26	44	178
40%		21	28	96	26	39	>201
50%		21	27	137	27	40	>201
60%		22	28	194	29	48	>201
70%		24	31	>201	34	59	>201
80%		27	38	>201	41	76	>201

CV: Coefficient of variation; GMR: Ratio of geometric means

The results in Table 5 can be compared also with those given in Table A3. The sample sizes in Table 5 were obtained by using ABEL (in order to be able to compare them directly with those in Table 4) whereas those in Table A3 were computed with SABE as required by FDA. Therefore, small differences between entries in the two tables are due to the two algorithms.

DISCUSSION

Comparison of Sample Sizes Required by EMA and FDA

It was noted from the comparison of Tables A1 and A2 with Tables A3 and A4 in the Appendix and of Table 4 with Table 5 that regulatory requirements of FDA call for fewer subjects than those of EMA.

At first sight, the requirements of FDA are more favourable to sponsors than those of EMA. However, statistical reasoning supports the recommendations of EMA. Figure 1 illustrates the sample sizes required by the two agencies at and just above $CV = 30\%$. The figure shows also the sample sizes needed by unscaled average BE just below and up to this variation. For illustrative purposes, the results were calculated with the true CV values, without the mixing effect and without the GMR constraint.

As expected, the required sample size increases with rising variation when the results are evaluated by unscaled average BE. Above $CV=30\%$ and with the application of ABEL, the sample size is independent of the variation when the true $GMR = 1.0$. However, when the true GMR deviates from 1.0 (e.g., when $GMR = 1.1$), the required sample

size initially decreases with rising CV. This behaviour was discussed in connection with the results given in Table 2.

Importantly, the required sample size changes *continuously* around $CV = 30\%$ when the requirements of EMA are followed (i.e., with the regulatory constant of $k = 0.76$ or $\sigma_0 = 0.294$). In other words, at $CV = 30\%$ the same sample size is obtained regardless whether ABE or ABEL is applied.

In contrast, there is a *discontinuity* of sample sizes when the regulatory conditions of FDA are used (i.e., with the regulatory constant of $k = 0.893$ or $\sigma_0 = 0.25$). In other words, at $CV = 30\%$ larger sample size is required by applying ABE than ABEL. In fact, using ABEL within a range above $CV = 30\%$ requires a smaller sample than applying ABE within a range below $CV = 30\%$ (Figure 1).

By requiring smaller samples above than below $CV = 30\%$, the FDA regulatory condition could tempt some sponsors to prefer higher variations.

Other disadvantageous consequences of the proposed FDA regulation include the higher consumer risk for non-highly variable drugs. This aspect was already noted earlier (16).

Further features of sample sizes arising from the EMA and FDA requirements can be noted. The difference between the two sample sizes decreases as the variation gets higher, towards 40-50% (Tables in the Appendix and also from the comparison of Tables 4 and 5). The reason is that the GMR constraint influences the outcome of the BE decision sooner with the FDA than with the EMA condition (15, 16).

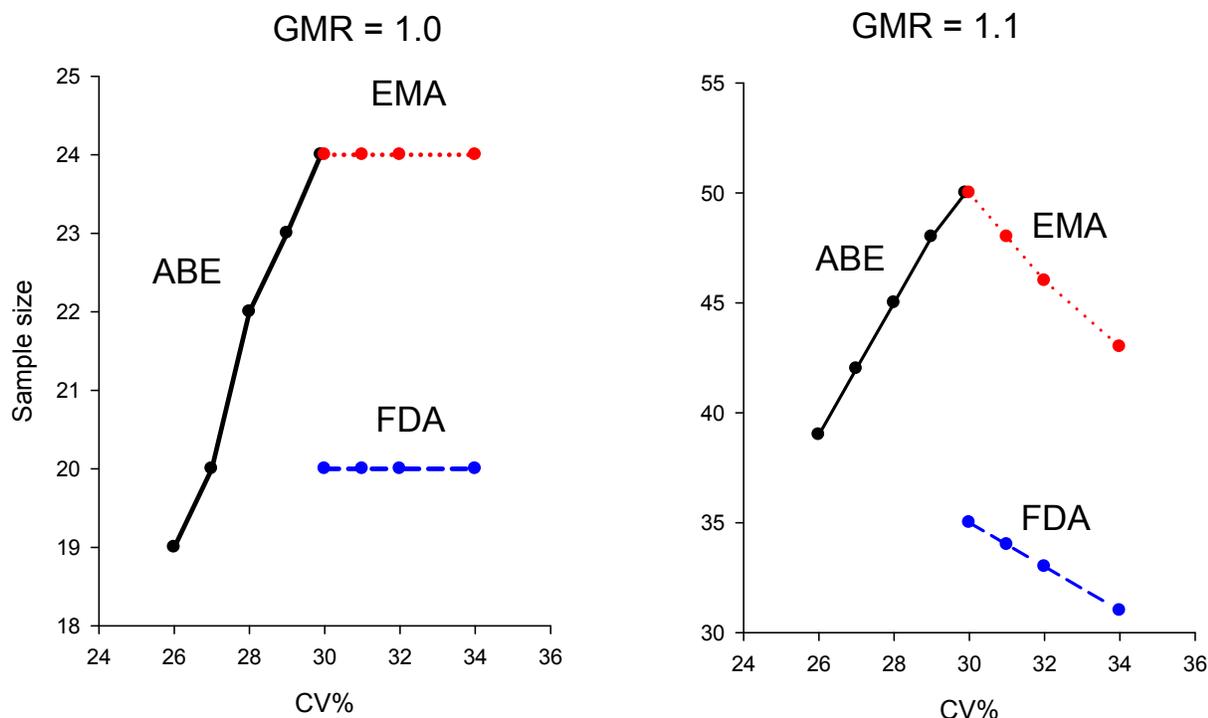


Figure 1. Sample sizes, around a within-subject coefficient of variation (CV) of 30%, required when applying unscaled average bioequivalence (ABE) and the regulatory expectations of EMA and FDA. Until 30% variation, unscaled ABE was considered, from 30% the use of either scaled ABE or ABE with expanding limits was assumed. The additional regulatory conditions and requirements of mixed model, constraint on GMR and a cap on the application of scaling were not considered and true (rather than estimated) variations were applied. The two regulatory authorities differ mainly in their recommended regulatory constants, $k = 0.76$ or $\sigma_0 = 0.294$ by EMA, and $k = 0.893$ or $\sigma_0 = 0.25$ by FDA. – GMR: Ratio of geometric means.

At still higher variations, at $CV > 50\%$, the required sample size increases more rapidly with rising CV with the EMA than with the FDA requirements. The reason is the cap on the use of ABE at $CV = 50\%$ which EMA imposes. FDA does not apply a similar cap. Consequently, with the EMA requirements, unscaled average BE is used again at CV's exceeding 50% thereby leading to stricter study requirements.

Designing BE studies for highly variable drugs

Sample sizes for designing BE studies which involve non-highly variable drugs are typically estimated by assuming a within-subject (or a residual) variation and using a sample-size table such as that of Hauschke et al. (7). The sample size is usually selected at a 5% deviation between the means, i.e. at a true $GMR = 1.05$.

Larger absolute differences between the two logarithmic means can be noted in the various BE studies when the within-subject variation is higher. Therefore, it is recommended that a 10% deviation

between the means, i.e. a true $GMR = 1.10$, be considered when the sample size tables in the Appendix are used.

With the approach of FDA, the minimum number of subjects with the 3-period partially replicating design is 28 and 40 subjects for 80% and 90% power, respectively. With a 4-period replicated design, these numbers are 21 and 30 for 80% and 90% power, respectively. Haidar et al. (3) suggested that the inclusion of at least 24 subjects would be needed. The suggestion may be considered as an absolute minimum.

With the procedure of EMA, the minimum number of subjects with the 3-period partially replicating design is 37 and 51 subjects for 80% and 90% power, respectively. With a 4-period replicated design, these numbers are 27 and 36 for 80% and 90% power, respectively.

The estimated sample sizes depend on the within-subject variation of the test product. If it is lower than that of the reference formulation then fewer volunteers are needed to achieve the stated

power. Conversely, if the variability of the test formulation is higher than that of the reference product then more subjects are needed than shown in the tables in the Appendix. In practice, however, the samples are too small to make judgments with adequate power about the relative variances of the two products and, consequently the assumption of identical variabilities is generally reasonable.

In view of the consequences of the mixed approach, it could be judicious to consider larger numbers of subjects at variations fairly close to 30%.

Both EMA and FDA developed the approaches for highly variable drugs in order to reduce the regulatory burden, i.e. to lower the required number of subjects in BE studies. The sample size tables in the Appendix demonstrate that both authorities achieve this goal.

CONCLUSIONS

Tables of sample sizes are provided for BE studies involving highly variable drugs. These investigations are evaluated either by the approach of scaled average BE or by its close variant, average BE with expanding limits. Sample sizes are shown for the differing regulatory requirements of EMA and FDA.

When the two drug products have truly the same kinetic metrics ($GMR = 1.00$) and without the additional regulatory conditions, the required sample size is independent of the within-subject variation of the reference formulation. In other words, the producer risk is independent of the variation.

Each of the additional regulatory conditions and requirements yields complications in the relationship between sample size and variation when the true GMR deviates from 1.00. Use of the mixed strategy lowers the sample size near $CV = 30\%$. A constraint on the point estimate of GMR increases the required sample size at higher variations. A cap on applying ABEL at a variation of 50% raises the sample size.

The sample sizes are lower with the requirements suggested by FDA than by those of EMA. However, as illustrated in Figure 1, the regulatory constant of FDA leads to a discontinuity of the required sample size and to statistically inconsistent behaviour around the switching variation of 30%.

ACKNOWLEDGMENT

The authors are grateful to Dr. Yu Chung Tsang for his insightful comments and suggestions.

REFERENCES

1. L. Tothfalusi, L. Endrenyi, and A. Garcia Areta. Evaluation of bioequivalence for highly variable drugs with scaled average bioequivalence. *Clin. Pharmacokin.* **48**: 725-743 (2009).
2. European Medicines Agency. Guideline on the investigation of bioequivalence. London, United Kingdom, January 20, 2010. http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2010/01/WC500070039.pdf.
3. S.H. Haidar, B. Davit, M.-L. Chen, D. Conner, L. M. Lee, Q. H. Li, R. Lionberger, F. Makhlof, D. Patel, D.J. Schuirman, and L.X. Yu. Bioequivalence approaches for highly variable drugs and drug products. *Pharm. Res.* **15**: 237-241 (2008).
4. L. Tothfalusi, L. Endrenyi, K.K. Midha, M.J. Rawson, and J.W. Hubbard. Evaluation of the bioequivalence of highly-variable drugs and drug products. *Pharm. Res.* **18**: 728-733 (2001).
5. K. F. Phillips. Power of the two one-sided tests procedure in bioequivalence. *J. Pharmacokin. Biopharm.* **18**: 137-144 (1990).
6. E. Diletti E, D. Hauschke, and V. W. Steinijans. Sample size determination for bioequivalence assessment by means of confidence intervals. *Int. J. Pharm. Ther. Toxicol.* **29**: 1-8 (1991).
7. D. Hauschke, V. W. Steinijans, E. Diletti, and M. Burke. Sample size determination for bioequivalence assessment using a multiplicative model. *J. Pharmacokin. Biopharm.* **20**: 557-561 (1992).
8. J. P. Liu and S. C. Chow. Sample size determination for the two one-sided test procedure in bioequivalence. *J. Pharmacokin. Biopharm.* **20**: 101-104 (1992).
9. D. Hauschke, V. Steinijans, and I. Pigeot. Bioequivalence Studies in Drug Development: Methods and Applications. Chichester: Wiley; 2007.
10. D. Schuirman. A comparison of the two one-sided tests procedure and the power approach assessing the equivalence of average bioavailability. *J. Pharmacokin. Biopharm.* **15**: 657-680 (1987).
11. A.W. Boddy, F.C. Snikeris, R.O. Kringler, G.C.G. Wei, J.A. Opperman, and K.K. Midha. An approach for widening the bioequivalence acceptance limits in the case of highly variable drugs. *Pharm. Res.* **12**: 1865-1868 (1995).
12. L. Tothfalusi, and L. Endrenyi. Limits for the scaled average bioequivalence of highly variable drugs and drug products. *Pharm. Res.* **20**: 382-389 (2003).

13. <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM209294.pdf>
14. V. W. Steinijans and D. Hauschke. Update on the statistical analysis of bioequivalence studies. *Int. J. Clin. Pharmacol. Ther. Toxicol.* **28**: 105-110 (1990).
15. S.H. Haidar, F. Makhlof, D.J. Schuirmann, T. Hyslop, B. Davit, D. Conner, and L.X. Yu. Evaluation of a scaling approach for the bioequivalence of highly variable drugs. *AAPS J.* **10**: 450-454 (2008).
16. L. Endrenyi and L. Tothfalusi. Regulatory conditions for the determination of bioequivalence of highly variable drugs. *J. Pharm. Pharmaceut. Sci.* **12**: 138-149 (2009).

APPENDIX

Table A1. Sample sizes for the requirements of EMA in 3-period studies

CV	80% POWER								
	GMR	0.85	0.90	0.95	1.00	1.05	1.10	1.15	1.20
30%		194	53	27	22	26	45	104	>201
35%		127	51	29	25	29	45	84	>201
40%		90	44	29	27	30	42	68	139
45%		77	40	29	27	29	37	57	124
50%		75	40	30	28	30	37	53	133
55%		81	42	32	30	32	40	56	172
60%		88	46	36	33	36	44	63	>201
65%		99	53	40	37	40	50	71	>201
70%		109	58	45	41	45	56	80	>201
75%		136	67	50	46	50	62	89	>201
80%		144	72	54	51	55	68	97	>201

CV	90% POWER								
	GMR	0.85	0.90	0.95	1.00	1.05	1.10	1.15	1.20
30%		>201	74	36	28	36	62	147	>201
35%		181	70	39	32	39	63	117	>201
40%		130	61	38	33	39	57	94	>201
45%		132	55	37	33	38	51	85	>201
50%		158	55	39	34	38	51	84	>201
55%		178	59	41	37	41	53	97	>201
60%		199	64	45	41	46	60	112	>201
65%		>201	72	51	46	51	67	125	>201
70%		>201	82	57	52	57	76	141	>201
75%		>201	93	66	58	64	85	161	>201
80%		>201	100	70	63	71	93	176	>201

CV: Coefficient of variation; GMR: Ratio of geometric means

Table A2. Sample sizes for the requirements of EMA in 4-period studies

CV	80% POWER								
	GMR	0.85	0.90	0.95	1.00	1.05	1.10	1.15	1.20
30%		127	35	19	15	18	30	68	>201
35%		88	34	20	18	20	31	57	140
40%		64	31	20	18	20	28	47	98
45%		57	29	21	19	21	27	41	90
50%		54	28	22	20	21	27	38	100
55%		55	30	23	21	23	28	40	116
60%		60	32	25	23	25	31	44	124
65%		74	37	28	26	28	33	49	155
70%		78	40	31	28	31	38	55	167
75%		85	45	34	32	34	42	61	186
80%		95	50	38	35	37	46	66	>201

CV	90% POWER								
	GMR	0.85	0.90	0.95	1.00	1.05	1.10	1.15	1.20
30%		180	49	25	19	24	42	95	>201
35%		123	48	27	22	27	43	80	>201
40%		93	42	26	23	26	39	66	165
45%		90	40	27	24	27	37	59	181
50%		102	39	27	25	27	36	60	>201
55%		123	41	29	26	29	38	63	>201
60%		139	45	32	29	31	41	71	>201
65%		159	51	36	32	35	46	81	>201
70%		172	55	40	36	40	52	97	>201
75%		195	62	43	39	44	58	106	>201
80%		>201	69	49	45	49	62	113	>201

CV: Coefficient of variation; GMR: Ratio of geometric means

Table A3. Sample sizes for the requirements of FDA in 3-period studies

CV	80% POWER								
	GMR	0.85	0.90	0.95	1.00	1.05	1.10	1.15	1.20
30%		145	45	24	21	24	39	82	>201
35%		74	37	24	22	25	34	54	109
40%		60	33	24	22	24	31	47	104
45%		59	31	23	22	24	29	43	116
50%		66	30	24	22	23	28	41	133
55%		80	30	24	22	24	28	44	172
60%		88	31	24	23	24	30	50	>201
65%		98	32	25	24	25	31	53	>201
70%		106	35	26	25	26	31	62	>201
75%		136	38	27	26	27	34	70	>201
80%		144	40	29	27	29	37	76	>201

CV	90% POWER								
	GMR	0.85	0.90	0.95	1.00	1.05	1.10	1.15	1.20
30%		>201	65	33	26	32	55	122	>201
35%		106	51	32	28	32	47	77	186
40%		99	45	31	28	31	43	68	>201
45%		128	43	30	28	30	40	69	>201
50%		158	45	31	28	30	40	79	>201
55%		178	50	31	28	31	42	96	>201
60%		199	54	33	30	34	50	112	>201
65%		>201	61	35	32	36	53	125	>201
70%		>201	68	39	34	37	61	141	>201
75%		>201	80	43	37	41	68	161	>201
80%		>201	83	48	41	47	75	176	>201

CV: Coefficient of variation; GMR: Ratio of geometric means

Table A4. Sample sizes for the requirements of FDA in 4-period studies

CV	80% POWER								
	GMR	0.85	0.90	0.95	1.00	1.05	1.10	1.15	1.20
30%		96	30	17	15	17	27	55	200
35%		54	26	18	16	18	24	39	79
40%		43	24	18	16	17	22	33	72
45%		44	23	18	16	17	21	32	82
50%		45	22	17	17	17	21	31	99
55%		52	22	18	17	17	21	31	116
60%		58	23	18	17	18	21	34	124
65%		74	24	19	18	18	22	36	155
70%		75	24	19	18	19	23	44	167
75%		81	26	20	19	20	24	47	186
80%		95	29	21	20	20	25	51	>201

CV	90% POWER								
	GMR	0.85	0.90	0.95	1.00	1.05	1.10	1.15	1.20
30%		152	44	23	18	22	38	81	>201
35%		80	38	23	20	23	34	55	128
40%		70	32	22	20	22	30	48	158
45%		84	32	22	20	22	30	49	181
50%		102	32	23	20	22	30	54	>201
55%		123	34	23	21	22	31	61	>201
60%		139	38	24	22	24	33	71	>201
65%		159	44	26	23	25	35	81	>201
70%		172	46	26	24	27	43	97	>201
75%		195	53	29	26	29	48	106	>201
80%		>201	60	33	28	31	51	113	>201

CV: Coefficient of variation; GMR: Ratio of geometric means