JPP JOURNAL OF Pharmacy and Pharmacology
ROYAL PHARMACEUTICAL SOCIETY
JPP Journal of Pharmacy And Pharmacology
Research Paper

On the statistical model of the two-stage designs in bioequivalence assessment

Vangelis Karalis* and Panos Macheras

Laboratory of Biopharmaceutics-Pharmacokinetics, Faculty of Pharmacy, National and Kapodistrian University of Athens, Athens, Greece

Keywords
adaptive design; bioequivalence; pharmacokinetic simulations; statistical method; two-stage design

Correspondence
Vangelis Karalis, Laboratory of Biopharmaceutics - Pharmacokinetics, Faculty of Pharmacy, National and Kapodistrian University of Athens, Panepistimiopolis, Athens 15771, Greece.
E-mail : vkaralis@pharm.uoa.gr

Received May 24, 2013
Accepted September 16, 2013

doi: 10.1111/jphp.12164

*To whom correspondence should be addressed

Abstract

Objectives Two-stage clinical designs are currently recommended by the regulatory authorities for the assessment of bioequivalence (BE). A specific statistical methodology was recently proposed by the European Medicines Agency (EMA). The aims of this article are to elaborate on the suggested statistical design from the EMA and to compare it with the existing statistical methods reported in the literature.

Methods Monte Carlo simulations were used to simulate the conditions of a two-stage BE design. The starting sample size was either 24 or 48, whereas the coefficient of variation of the within-subject variability was equal to 20% and 40%. Several geometric mean ratio levels of the BE metric were considered. Under each condition, 1 000 000 studies were simulated.

Key findings The overall performance, in terms of percentage of BE acceptance, is identical. The additional term, 'sequence × stage', suggested in the EMA method is in most cases nonsignificant. The same results were obtained regardless of the type (fixed or random) of the effect applied to the 'subjects' term.

Conclusions Any BE study either finished or in progress which relies on the existing literature methodology leads to the same percentage of BE acceptance as if it was analysed with the recently proposed EMA method.

Introduction

Nowadays, adaptive methods constitute a possibility of clinical design, and recently, they have attracted the attention of official regulatory authorities in case of bioequivalence (BE) assessment. The US Food and Drug Administration (FDA) allows the application of two-stage group-sequential design approaches.[1,2] In addition, the latest guideline issued by the European Medicines Agency (EMA) proposes a two-stage design (TSD) as an alternative to the standard 2 × 2 or the replicate BE designs.[3] The TSD procedures are generally based on the fact that if BE cannot be demonstrated on stage 1, the applicant is allowed to recruit more subjects and move on to a second stage.[4]

However, a specific and detailed description of a framework regarding the structure and the criteria of the TSD were not set by the regulatory authorities. Until now, a major contribution to the TSD BE design was provided by the articles of Potvin *et al.* and Montague *et al.*[5,6] In these two articles, four methods for sample size re-estimation were assessed, and the authors proposed recommendations on the appropriateness of each method. Very recently, two additions were made to the TSD methods in BE assessment. Firstly, our group published an article on the underlying properties of TSD designs and introduced a TSD design with an upper sample size limit where sample size re-estimation is based on the actual difference observed in stage 1.[7] Secondly, Fuglsang has focused on TSDs with increased power and controlled type I errors.[8]

Quite recently (February 2013), the EMA released a questions and answers guideline which, among others, refers for the first time to the statistical procedure that should be followed for the TSDs.[9] This guideline defines the statistical effects of analysis of variance (ANOVA) that should be included in the analysis of the combined data of the two stages of the study. Even though the EMA guideline quotes a specific method of statistical analysis, the recommended procedure is different from that already presented in

literature and most likely used in practice. Therefore, there is a question whether the new statistical methodology, proposed by the EMA, could influence the BE studies which are currently in progress or already completed. This concern is further enlarged because of the fact that TSDs are currently widely and increasingly used in BE assessment.

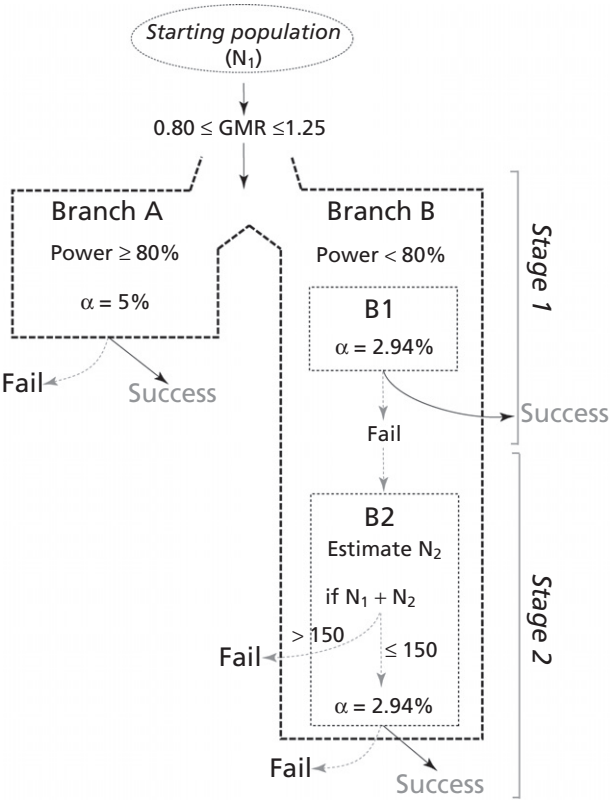
The aim of this study is to compare the statistical methods, either the existing (EX) or the newly proposed by EMA, applied to the analysis of two-stage BE designs.[5–7,9] The properties of these two methods are analysed using Monte Carlo simulations. The similarities and the differences between the EMA and EX methods as well as their impact on BE assessment are highlighted and discussed.

Materials and Methods

Two-stage design

The TSD analysed in this study is actually based on the TSD introduced in our previous work with a minor modification: the inclusion of an initial geometric mean ratio (GMR) criterion. This TSD was described in detail in our previous work, and for this reason, only a brief description is provided below.[7] The TSD is split into three branches: A, B1 and B2 (Figure 1). The first stage includes branches A and B1, whereas the second stage refers to branch B2. Each stage of this TSD consists of a two-treatment, two-sequence, two-period crossover design. Sample size re-estimation at branch B2 relies on the observed (actual) GMR and the coefficient of variation of within-subject variability (CVw) estimated at branch A. No increase of type I error rate (α) beyond 5% was observed when this TSD was applied.[7]

According to the design used in this study, the first step is the estimation of GMR using the data from branch A. If the point GMR value lies outside the 0.80–1.25 range, then BE failure is considered, and the TSD stops. Even though this criterion does not control the use of second stage, but the entire TSD algorithm, it can be considered as a hidden futility criterion. For GMR values within the 0.80–1.25 interval, the assessment continues with the estimation of the power on stage 1 setting $\alpha = 5\%$ and using the actual values of CVw and GMR. If the so-derived power is higher than or equal to 80%, then BE is assessed at branch A using $\alpha = 5\%$ level of significance. If the calculated power of the study is less than 80%, branch B of the TSD method is followed, and BE is assessed at α level of 2.94% (branch B1 in Figure 1). If BE is proved, then the assessment algorithm stops; if BE is not proved, the TSD method proceeds into branch B2 where sample size re-estimation takes place based on the CVw and the GMR estimates derived from stage 1 and setting $\alpha = 2.94\%$. The number of additional subjects (n_2) estimated during sample size re-estimation at branch B2 could range between 2 and 150.[7,9] Finally, the assessment of



N_1 : the starting sample size
N_2 : the additional number of subjects recruited at the second stage
GMR : the geometric mean ratio of the bioequivalence metric under study
α : the significance level
150 : the maximum allowed number of subjects from stages 1 and 2

Figure 1 The two-stage design used in this study.

BE is based on the data from both stages 1 and 2, assuming type I error is equal to 2.94%.

Statistical model

In case of stage 1, the terms used in the ANOVA model are always 'sequence', 'period', 'treatment' and 'subject(sequence)'.[3] After sample size re-estimation, at branch B2, statistical analysis uses data combined from stages 1 and 2. In this case, the EX statistical model traditionally appearing in the literature uses the following ANOVA effects: 'sequence', 'treatment', 'stage', 'period(stage)' and 'subject(sequence × stage)'.[3,5–7] All these effects are treated as 'fixed' factors.[3,9] The recently recommended ANOVA model by the EMA suggests the additional use of the 'sequence × stage' effect for the BE assessment at the second stage.[9] Unless mentioned differently in this study, the results refer to the situation where all effects are considered as 'fixed'.

It should be mentioned that no poolability criterion was set to the utilized TSD approach, and data from stages 1 and 2 were always pooled even if a significant 'sequence × stage' effect was observed.

Simulations

The utilized algorithm of the simulations was in accordance with that used in our previous work.[7] In brief, simulated values for the pharmacokinetic parameter were generated from log-normal distribution. The pharmacokinetic estimates for each product, Test or Reference, were appropriately assigned to the two sequences and the two periods of stage 1 of the study. The latter was accomplished in a way that ensured randomness and balance with respect to sequence, period and treatment effects.

Two levels (20% and 40%) of theoretical CVw values of the initial population were considered in the simulations. In addition, two different starting sample size (N_1) values were considered: 24 and 48 subjects. The theoretical GMR value was gradually changed, from 1.00 to 1.25, using a step of 0.025. Under each condition, a number of 1 000 000 studies according to the TSD scheme (Figure 1) were simulated. In each study, BE was declared if the (1–2α)% confidence interval around the point GMR estimate of Test or Reference was between the BE limits (80.00–125.00).[10] The entire programming work was implemented in MATLAB (The MathWorks, Inc., Natick, MA, USA).

Results

Table 1 lists the degrees of freedom (df) for all statistical effects of ANOVA using the EX and the EMA method when combined data from stage 1 and 2 are analysed. Plausibly, the total number of df is the same. The difference in the df values is observed in the nested term 'subject(sequence × stage)', whereas the term 'sequence × stage' appears only in case of the EMA method with one df. Subsequently, the 'subject(sequence × stage)' effect exerts one df less for the EMA model, namely, $n-4$ instead of $n-3$ of the EX method. The df values for the remaining effects remain unaltered.

Table 1 Degrees of freedom for the ANOVA effects in case of the EMA and the EX

	Degrees of freedom	
Source	EX	EMA
Subject (sequence × stage)[a]	$n-3$	$n-4$
Sequence[a]	1	1
Period (stage)	2	2
Treatment	1	1
Stage	1	1
Sequence × stage[a]	–	1
Residual error	$n-3$	$n-3$
Total	$2^* n-1$	$2^* n-1$

ANOVA, analysis of variance; EMA, European Medicines Agency; EX, existing. [a]The estimated sum of squares for these terms differs between the EX and EMA method. The term 'n' refers to the total number of subjects from stages 1 and 2.

To estimate the sum of squares (SS) of each ANOVA effect, statistical analysis was also applied to the combined data from both stages 1 and 2 using the EX and EMA approaches. The results of the statistical analysis clearly reveal (results not shown) that the SS values slightly differ between the EX and EMA method in the case of the 'subject(sequence × stage)', 'sequence' and for the 'sequence × stage' terms. The SS for all other ANOVA effects are identical.

Figure 2 depicts the percentage of simulated studies in which BE is accepted versus the GMR of the study when the EX and EMA methods are applied to the same data. The theoretical CVw was set to 20%, and two different levels of starting sample size are shown: $N_1 = 24$ (Figure 2a) and $n = 48$ (Figue 2b). Visual inspection of Figure 2 reveals that the EX and EMA statistical approaches lead to identical results. As the number of subjects recruited in the BE study increases, the percentage of BE acceptance also increases.

Special emphasis was also given to the significance (P value) of the additional term 'sequence × stage' used in the ANOVA model proposed by EMA. For this reason, the significance (expressed as a %) of 'sequence × stage' is plotted versus the GMR of the study (Figure 3). Two different cases ($N_1 = 24$ and $N_1 = 48$) are shown within each plot, whereas CVw values are set to 20% (Figure 3a) and 40% (Figure 3b).

In almost all situations, the significance of the 'sequence × stage' term was found to be nonsignificant (i.e. values were greater than 5%). Only when GMR was close to the limit of 1.25 can the significance obtain lower values than the significance level 5%, and thus, the 'sequence stage' effect was declared significant. As the starting sample size or CVw values decrease, the significance values of this term rises (Figure 3).

Because of space limitation, the results from the remaining conditions studied are not shown. However, the results follow the general pattern unveiled in this analysis. Similarly, the results coming from studies where the 'subjects' effect was set as 'random' are not quoted, but similar findings were observed.

Discussion

Compared with the EX statistical method, the newly introduced EMA model includes one additional effect in the ANOVA model for the analysis of the combined data from stages 1 and 2.[5–7,9]

The overall performance in terms of percentage of BE acceptance of the TSD remains unaltered. Plausibly, no dif-

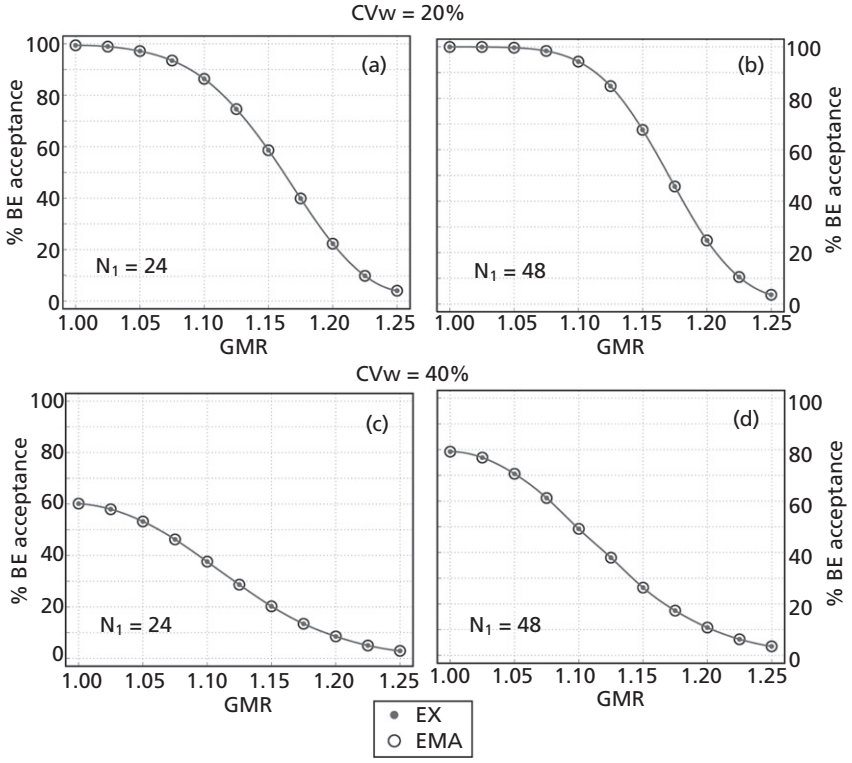


Figure 2 Percentage of bioequivalence studies accepted by the existing and European Medicines Agency methods versus geometric mean ratio (GMR). The coefficient of variation of the within-subject variability is equal to 20% (a,b) and 40% (c,d). Two different starting sample sizes (N_1) are assumed: 24 and 48.

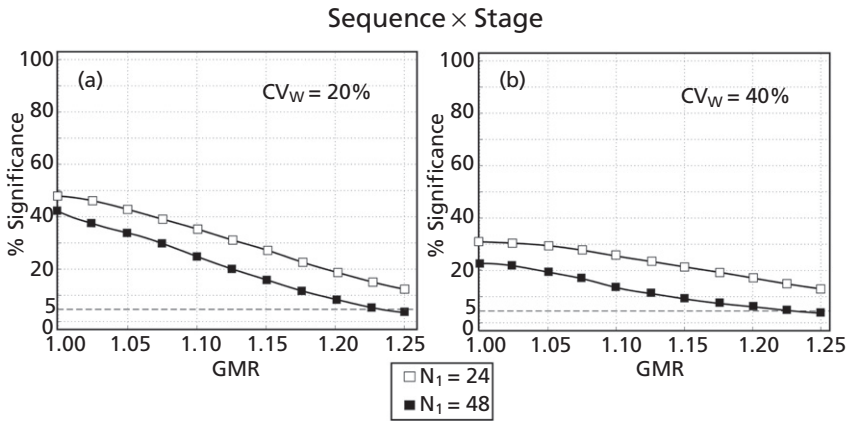


Figure 3 The percentage of significance of the 'sequence × stage' analysis of variance effect proposed by the European Medicines Agency method as a function of geometric mean ratio (GMR). Two starting sample size values ($N_1 = 24$ and $N_1 = 48$) are quoted inside each plot. Two levels of within-subject variability (20% (a) and 40% (b)) are assumed.

ference in both the df and the SS values is observed for the 'residual error'; the 'sequence × stage' is in essence a between-subject factor, whereas BE assessment is based on CVw. The only differences are ascribed to the existence of the additional term 'sequence × stage' and the df of the nested term 'subject(sequence × stage)' (see Table 1). In addition, the SS estimates for these two effects as well as for the related 'sequence' term are slightly different between the EX and EMA methods.

The difference in the significance values of the shared effects 'subject(sequence × stage)' and 'sequence', between the EX and EMA approaches, is rather small. This finding

was validated by plotting the significance of these terms versus the theoretical GMR (results not shown). The results obtained from the two approaches, EX and EMA, were almost super-imposable. Furthermore, the same results were obtained regardless of the type (fixed or random) of the effect applied to the 'subjects' term. In most of the cases, the significance value of the additional term 'sequence × stage' used in the EMA statistical model was found to be nonsignificant. Only when the theoretical GMR was higher than 1.20 was the significance of this term less than 5%. In any case, the EMA guideline does not clarify what the consequence would be if the 'sequence × stage' is statistically significant.[11]

Finally, this study unveils that the BE studies, which rely on the EX method and are already finished or currently in progress, lead to the same results as if these were analysed with the EMA method. It should be highlighted that even though our analysis was based on the TSD depicted in Figure 1, similar findings can also be observed using other TSDs.[5,6] The same is also true for TSD methods with no power estimation at stage 1, like the so-called 'B method' where α is adjusted already in the first stage.[5,6] In any case, the findings of this study do not depend on the utilized TSD but on the statistical method applied.

Conclusions

This study focuses on the statistical methods applied to the analysis of two-stage BE designs. The newly proposed method by the EMA was compared with the EX methods quoted in the literature. The basic conclusions derived from our analysis include the following: (1) The statistical approach recommended by the EMA leads to identical percentage of BE acceptance with the statistical methods already in use. (2) The additional term 'sequence × stage' is in most cases nonsignificant. (3) No difference in the overall percentage of BE acceptance can be observed if the effect of 'subjects' is set as 'fixed' or 'random'.

Any BE study either finished or in progress which relies on the EX method leads to the same percentage of BE acceptance as if it was analysed with the recently proposed EMA method. Plausibly, the probability of accepting a two-stage BE study, which is analysed according to the FDA guideline (i.e. EX method), will be the same with the EMA method.

Declarations

Conflict of interest

The authors declare that they have no conflicts of interest to disclose.

Funding

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

References

1. FDA (Food and Drug Administration). 2012 Draft guidance on Dexamethasone – Tobramycin (Rev. June 2012).
2. Polli J *et al.* Summary workshop report: facilitating oral product development and reducing regulatory burden through novel approaches to assess bioavailability/bioequivalence. *AAPS J* 2012; 14: 627–638.
3. EMA (European Medicines Agency). *Committee for Medicinal Products for Human Use, CHMP*. London: Guideline on the Investigation of Bioequivalence, 2010.
4. Pong A *et al.*, ed. *Handbook of Adaptive Designs in Pharmaceutical and Clinical Development*. Boca Raton, FL: CRC Press, Taylor and Francis group, 2011.
5. Potvin D *et al.* Sequential design approaches for bioequivalence studies with crossover designs. *Pharm Stat* 2008; 7: 245–262.
6. Montague T *et al.* Additional results for 'Sequential design approaches for bioequivalence studies with crossover designs. *Pharm Stat* 2012; 11: 8–13.
7. Karalis V, Macheras P. An insight into the properties of a two-stage design in bioequivalence studies. *Pharm Res* 2013; 30: 1824–1835.
8. Fuglsang A. Sequential bioequivalence trial designs with increased power and controlled type I error rates. *AAPS J* 2013; 15: 659–661.
9. EMA (European Medicines Agency). Committee for medicinal products for human use, CHMP. Questions & answers: positions on specific questions addressed to the pharmacokinetics working party. (EMA/618604/2008 Rev. 7) 2013.
10. 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: 657–680.
11. García-Arieta A, Gordon J. Bioequivalence requirements in the European Union: critical discussion. *AAPS J* 2012; 14: 738–748.