International Journal of Pharmaceutics 456 (2013) 87–94

Contents lists available at ScienceDirect

International Journal of Pharmaceutics

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

The role of the upper sample size limit in two-stage bioequivalence designs

Vangelis Karalis*

Laboratory of Biopharmaceutics and Pharmacokinetics, Faculty of Pharmacy, National and Kapodistrian University of Athens, 15771 Athens, Greece

ARTICLE INFO

Article history:
Received 18 June 2013
Received in revised form 6 August 2013
Accepted 8 August 2013
Available online 14 August 2013

Keywords:
Bioequivalence
Two-stage designs
Adaptive designs
Pharmacokinetic simulations

ABSTRACT

Two-stage designs (TSDs) are currently recommended by the regulatory authorities for bioequivalence (BE) assessment. The TSDs presented until now rely on an assumed geometric mean ratio (GMR) value of the BE metric in stage I in order to avoid inflation of type I error. In contrast, this work proposes a more realistic TSD design where sample re-estimation relies not only on the variability of stage I, but also on the observed GMR. In these cases, an upper sample size limit (UL) is introduced in order to prevent inflation of type I error. The aim of this study is to unveil the impact of UL on two TSD bioequivalence approaches which are based entirely on the interim results. Monte Carlo simulations were used to investigate several different scenarios of UL levels, within-subject variability, different starting number of subjects, and GMR. The use of UL leads to no inflation of type I error. As UL values increase, the % probability of declaring BE becomes higher. The starting sample size and the variability of the study affect type I error. Increased UL levels result in higher total sample sizes of the TSD which are more pronounced for highly variable drugs.

1. Introduction

Classically, assessment of bioequivalence (BE) relies on the concept of average BE. In this case, two medicinal products are considered bioequivalent when the estimated ninety percent confidence interval (90% CI) for the difference of mean pharmacokinetic metrics lies within the predefined limits of acceptance (usually, 0.80–1.25%) (EMA, 2010; FDA, 2001, 2003). BE studies are in essence clinical trials and therefore their design obeys the same general principles of clinical studies. In this context, sample size estimation is of crucial importance in BE studies. This implies that one should have a prior knowledge of: (a) the expected difference in the mean values of the pharmacokinetic metric (e.g. AUC, C_{max}) between and test and reference formulation and (b) the within-subject variability of the active moiety. Wrong estimates of variability and/or difference in the pharmacokinetic parameters may lead to BE studies which are under- or over-powered. Both situations are not desired, since low statistical power results in inability to prove the alternate statistical hypothesis (i.e. bioequivalence), whereas over-powered studies lead to increased study costs and unnecessary exposure of humans to drugs.

In contrast, adaptive design methods can be used instead of typical single-stage studies in order to face the above mentioned problems. Adaptive methods allow modifications made to trial and/or statistical procedures of ongoing clinical trials. The concept of adaptive design was first considered back to 70s when the adaptive randomization and a class of designs for sequential clinical trials were introduced. Since then, several types of adaptive designs have been proposed in the literature, such as group sequential, sample – size re-estimation, drop loser, response – adaptive randomization, adaptive dose escalation, adaptive hypotheses, and seamless designs (Gallo et al., 2006; Dragalin, 2006; Chow and Chang, 2008). Even though, adaptive designs brought many advantages in clinical research, some difficulties also exist (Emerson and Fleming, 2010; Mehta and Pocock, 2011). Two-stage design (TSD) approaches rely on the basis that if BE cannot be demonstrated on the first stage of the study, then the sponsor can enroll more volunteers during the second stage of the study (Pong and Chow, 2011). At stage II of the study sample size re-estimation takes place based on the interim results of the first stage. Even though, adaptive designs offer many advantages in clinical research some problems may arise. For example, many adaptations of the study may lead to a significantly different trial. In addition, as the number of interim

Abbreviations: ANOVA, analysis of variance; BE, bioequivalence; CI, confidence interval; CVw, coefficient of variation of the within-subject variability of the bioequivalence metric; GMR, geometric mean ratio (test/reference) of the bioequivalence metric; N, total number of subjects participating in the study; N_1, starting sample size; N_2, additional number of subjects recruited at the second stage; TSD, two-stage design; TSD-1, first type of two-stage design used in this study; TSD-2, second type of two-stage design used in this study; UL, upper sample size limit; α, type I error of the nominal statistical hypothesis.

* Correspondence address: Laboratory of Biopharmaceutics and Pharmacokinetics, Faculty of Pharmacy, National and Kapodistrian University of Athens, University Campus, 15771 Athens, Greece. Tel.: +30 2107274267; fax: +30 2107274027.
E-mail address: vkaralis@pharm.uoa.gr

0378-5173/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.ijpharm.2013.08.013

analyses increases, there is an increased risk of inflation of type I error (i.e., the significance level α).

Recently, two-stage designs are allowed to be applied to BE studies. Several regulatory authorities worldwide recommend the use of TSD design for BE assessment. It is quoted in the European Medicines Agency (EMA) 2010 guideline that TSD methods can be used as alternatives to the standard single stage or the replicate designs (EMA, 2010, 2013). In the same vein, the US Food and Drug Administration (FDA) allows the application of two-stage group-sequential design approaches (FDA, 2012). Also other authorities, like World Health Organization, Health Canada, and the Japanese Pharmaceuticals and Medical Devices Evaluation Agency recommend the use of add-on designs in BE assessment (WHO, 2006; Health Canada, 2012; NIPHJ, 2012).

Several articles have recently appeared in the literature focusing on the properties of TSD in case of BE assessment. Potvin et al. and Montague et al. published two articles which examined the performance of four two-stage methods; a naïve TSD approach and three modifications of adaptive methods assuming a pre-defined geometric mean ratio (GMR) value for the pharmacokinetic parameter (Potvin et al., 2008; Montague et al., 2012). In these articles, the authors did not use the actual GMR observed on stage I, but they considered a fixed prior value of either 0.90 or 0.95. Very recently, Fuglsang presented a study which focused on two-stage bioequivalence designs with increased power and controlled type I errors (Fuglsang, 2013).

At the same time, our group published a study where sample size re-estimation is not based on a prior GMR estimate, but on the use of the actual GMR observed in stage I (Karalis and Macheras, 2013). This situation is more realistic since the true GMR observed in stage I, apart from the actual within-subject variability, is used for sample size re-estimation. In the past, Cui et al. have noticed that increasing sample size based on an interim estimate of the treatment difference can lead to a substantial inflation of type I error (Cui et al., 1999). In order to deal with this issue, our method included also a pre-defined upper limit (UL) to the total sample size (N), namely, to the sample size occurring by adding the number of subjects in stage I (N_1) and those enrolled after sample size re-estimation in stage II (N_2).

The aim of this study is to elaborate on the necessity and the role of a pre-defined upper sample size limit in two-stage clinical designs which are based entirely on interim study results. Two TSD approaches are used in order to examine: (a) the impact of upper sample size limit on the percent of BE acceptance, (b) the relationship between UL and type I error, and (c) the effect of UL on the magnitude of the utilized total sample size from stages I and II. Conclusions about the role of UL and its appropriate setting in BE studies are derived. Monte Carlo simulations were used to investigate several different scenarios that may be encountered in practice.

2. Materials and methods

2.1. Two-stage designs

Two TSD methods were assessed in this study (Fig. 1). All of them were originated from the basic idea of the TSD approaches described by Potvin et al. (2008) and Montague et al. (2012). However, our TSD methods were changed appropriately in the following points:

(a) Sample size re-estimation is based on the actual GMR estimated at stage I rather than an assumed population GMR of 0.90 or 0.95.
(b) A pre-defined upper total sample size is used.
(c) An initial GMR criterion is used; if GMR lies only within the 0.80–1.25 region, the TSD design will be followed.

Each TSD was split into three segments: A, B1, and B2. Stage I of the study includes segments A and B1. Besides, segment B2 refers to stage II of the study (Fig. 1). Each stage of the TSD methods consists of a two-sequence, two-period (2×2) crossover design. Sample size re-estimation always takes place in stage II and it is based on the observed GMR and the coefficient of variation of the within-subject variability (CVw) for the active moiety calculated in segment A. The initial step of each TSD approach is GMR estimation relying on the data of segment A. When the point estimate for GMR, of the bioequivalence metric under study, falls outside the region 0.80–1.25, then the TSD stops and BE failure is declared. If the point estimate for GMR lies within the 0.80–1.25 interval then BE assessment using each TSD (TSD-1 or TSD-2) method continues as follows:

2.1.1. TSD-1

The statistical power of the study is estimated assuming $\alpha = 5\%$ and using the observed GMR and CVw values. If the estimated power is higher than or equal to 80%, then BE assessment is made at $\alpha = 5\%$ (Fig. 1A). The algorithm stops regardless of the BE outcome (pass or fail). In cases when the statistical power is lower than 80%, evaluation proceeds into segment B where initially an assessment of BE is made at $\alpha = 2.80\%$. Two alternatives are possible here; either the algorithm stops if BE is declared or estimation continues to segment B2 if BE was not shown earlier. In B2, sample size re-estimation takes place setting $\alpha = 2.80\%$ and using the CVw and GMR observed in stage I. The final step is assessment of BE using all data from stage I and II and setting $\alpha = 2.80\%$ (Potvin et al., 2008; Montague et al., 2012).

2.1.2. TSD-2

TSD-2 algorithm is quite different from TSD-1. In this case, after the initial GMR criterion, the process continues with BE assessment at $\alpha = 2.94\%$ (Fig. 1B). If BE is met, then the algorithm stops (segment A). When BE is not declared, then the statistical power of the study is estimated using the observed GMR and CVw values calculated in the previous step and setting $\alpha = 2.94\%$. If the so-derived power is higher than or equal to 80%, then the procedure stops (segment B1). However, in cases when the statistical power is lower than 80%, the algorithm continues to segment B2 where sample-re-estimation takes place. The latter is based on the actual values of GMR and CVw found in stage I and $\alpha = 2.94\%$. The final step of the algorithm involves estimation of BE on a type I error $\alpha = 2.94\%$ using all data from both stages I and II (Potvin et al., 2008; Montague et al., 2012).

2.1.3. Important points

Obviously, according to the utilized algorithms, TSD-1 or TSD-2, statistical power calculation and sample size re-estimation are based on the observed CVw and the GMR estimated of stage 1 rather than an assumed population GMR of 0.90 or 0.95. In addition, it should be highlighted that in this study an upper limit to the total sample size (the sum from stages I and II) was set. Several values of the UL were considered in this analysis (see Section 2.3 for more details). The minimum number of subjects recruited at stage II is two. Therefore, assuming that the total number of subjects from both stages is $N = (N_1 + N_2) \leq \text{UL}$, the number of additional subjects estimated at stage II could range from 2 to UL–N_1 (EMA, 2013).

In the utilized algorithm (either TSD-1 or TSD-2), the condition whether N is lower than or equal to UL was checked after estimation of N_2 at segment II. If the required N_2 resulted in a value of N which exceeded UL, then algorithm stopped and a BE failure was declared. Otherwise, the algorithm continued to the assessment of BE using data from both stages. To this point it should be mentioned that if one defines N_1 to be higher than UL, then it is implied that BE assessment does not proceed into stage II; in other words the TSD design reduces to a simple one-stage design.