Research Article

A Sequential Bioequivalence Design with a Potential Ethical Advantage

Anders Fuglsang^{1,2}

Received 18 October 2013; accepted 14 May 2014; published online 29 May 2014

Abstract. This paper introduces a two-stage approach for evaluation of bioequivalence, where, in contrast to the designs of Diane Potvin and co-workers, two stages are mandatory regardless of the data obtained at stage 1. The approach is derived from Potvin's method C. It is shown that under circumstances with relatively high variability and relatively low initial sample size, this method has an advantage over Potvin's approaches in terms of sample sizes while controlling type I error rates at or below 5% with a minute occasional trade-off in power. Ethically and economically, the method may thus be an attractive alternative to the Potvin designs. It is also shown that when using the method introduced here, average total sample sizes are rather independent of initial sample size. Finally, it is shown that when a futility rule in terms of sample size for stage 2 is incorporated into this method, i.e., when a second stage can be abolished due to sample size considerations, there is often an advantage in terms of power or sample size as compared to the previously published methods.

KEY WORDS: alpha; bioequivalence; power; sequential design.

INTRODUCTION

Diane Potvin and co-workers (1,2) introduced in 2008 some two-stage approaches for evaluation of bioequivalence and used trial simulations to investigate how the methods performed in terms of type I errors, power, and sample sizes. The methods involve initial trialing of N_1 subjects and evaluation of power or bioequivalence. It allows to stop after stage 1 when the products are bioequivalent or if the power at stage 1 is exceeding 80% and uses the variability observed at stage 1 to determine the sample size at stage 2. During my work with an agency in the ASEAN region, it came to my attention that while there is no specific guideline in that region that discusses two-stage approaches for evaluation of bioequivalence, the regulators expect two actual stages, meaning they do not expect the applicant to stop after stage 1 regardless of how the data look after stage 1. This in effect means that the Potvin designs are not desirable; in fact, there is to my knowledge not a single publication which shows the performance of any two-stage method where both stages are mandatory. The purpose of this work is therefore to introduce a variant of the Potvin designs which involves two mandatory stages and which controls type I errors at or below 5% and to use trial simulation to characterize it in terms of power and sample sizes, and to compare the performance with the

method introduced by Potvin *et al.* for a population geometric mean ratio (GMR) of 0.95. Finally, the method is also evaluated when it is combined with a futility rule in terms of stage II sample size, at a GMR of 0.95.

MATERIALS AND METHODS

“Method E”

Figure 1 shows an approach derived from Potvin's method C/D where two stages are mandatory. The principle is to use the variability at stage 1 to evaluate power at stage 1 after the initial N_1 subjects have been exposed to the investigational products. If the power is equal to or higher than 80%, then we include N_2 new subjects for stage 2 of the trial. If the power after stage 1 is lower than 80%, then we calculate a sample size N_2 for a target power of 80%, execute stage 2 with N_2 new subjects, and evaluate the pool of data from stage 1 and stage 2. The same alpha is applied for power and sample size calculation after stage 1 and for bioequivalence evaluation after stage 2. In this paper, I investigate how the method performs when N_2 is not lower than $\frac{1}{2}N_1$ ($N_{2,\min} = \frac{1}{2}N_1$); I will refer to this as “method E” to avoid confusion with Potvin's methods A, B, C, and D. Figure 2 illustrates how method E can be extended with a stopping rule in terms of sample size ($N_{\max}$). In this paper, method E with futility rules is also being evaluated; this is done using the futility criteria applied previously (see next section).

Scenarios, Evaluation, and Software

As in the works by Potvin and Montague (1,2), scenarios with $N_1 = 12, 24, 36, 48$, and 60 and coefficient of variations

Electronic supplementary material The online version of this article (doi:10.1208/s12248-014-9622-7) contains supplementary material, which is available to authorized users.

¹ Fuglsang Pharma, Hiort Lorenzens Vej 6C, 6100, Haderslev, Denmark.

² To whom correspondence should be addressed. (e-mail: a.fuglsang@ymail.com; a.fuglsang@ymail.com)

Fig. 1. Graphical description of method E. In contrast to the designs invented by Potvin *et al.*, this approach always has two stages and a single alpha value is used at both stages. In this paper, the performance is investigated for $N_{2,min} = \frac{1}{2}N_1$

(CVs) from 0.1, 0.2, 0.3 ... 1.0 were evaluated with 1,000,000 trial simulations per scenario and with balance between treatment sequences. The equations used to evaluate the data are the same as in the publication by Potvin *et al.* Software to investigate the performance of this method was written in C and compiled with MinGW version 4.7.2 under the Code::Blocks environment. To identify an appropriate alpha for method E, the linear interpolation method described elsewhere (3) was used.

Fig. 2. Method E with a futility rule; the study is declared failed before the execution of stage II if the total sample size exceeds a threshold

Futility rules are investigated under $GMR=0.95$ and are as follows:

$$\begin{aligned}
 N_{\max} &\leq 2 \times N_1 \\
 N_{\max} &\leq 3 \times N_1 \\
 N_{\max} &\leq 4 \times N_1 \\
 N_{\max} &\leq 60 \\
 N_{\max} &\leq 80
 \end{aligned}$$

RESULTS AND DISCUSSION

Figure 3 shows how the appropriate alpha has been identified. The critical scenario (the scenario where the method displays the highest type I error) for $GMR=0.95$ is $N_1=12$ and $CV=0.3$; in this scenario, the type I error rate is controlled at 0.05 when the alpha is 0.0446, corresponding to derivation of a 91.08% confidence interval.

It is not possible to show all performance results in this paper; performance results not shown in this paper are available as [Supplementary material online](#). In this paper, examples of performance are shown as figures.

Figure 4 shows power for method E as function of N_1 for $CV=20\%$ and $CV=40\%$. Figure 5 shows the type 1 error for method E as function of N_1 for $CV=20\%$ and $CV=40\%$. Figures 6 and 7 show the corresponding plots for method E with the stopping rule $N_{\max}=3 \times N_1$.

On comparison of the results obtained here with the results published by Potvin *et al.*, sample sizes, e.g., the average sample sizes, are actually somewhat lower with method E as compared to method C at $GMR=0.95$ when the variation is relatively high and the initial sample size is relatively low. In these scenarios, this translates into a clear ethical advantage since the total exposure to investigational medicinal products is lower. There may be a minute trade-off in terms of power. As a rule of thumb, in terms of sample sizes, method E outperforms method C in scenarios that have a high tendency to go into a second stage, and conversely,

Fig. 3. Type I error rate as function of alpha for method E at $N_1=12$ and $CV=0.3$. The line indicates the linear regression line; at approximately $\alpha=0.0446$, the type I error is 0.05

Fig. 4. Power as function of N_1 for method E without a stopping rule, for CV=20% (black circle) and CV=40% (white circle)

Fig. 6. Power as function of N_1 for method E with the stopping rule $N_{\max}=3 \times N_1$, for CV=20% (black circle) and CV=40% (white circle)

method E is inferior in terms of sample sizes for scenarios where method C has a low chance of going into a second stage. Interestingly, method E may be associated with another advantage: From all the tested combinations of N_1 and CV, it is clear that average sample sizes are relatively independent of initial sample sizes, and from this perspective, there is thus no good reason to start with high initial sample sizes, albeit low initial samples are associated with lowest power for a given CV and GMR. This contrasts strongly with the methods of Potvin where the lowest total sample size is often observed for the highest initial sample sizes ($N_1=60$).

When futility rules are applied, we see similar patterns when the results are compared to Fugsang (4).

In this work, $N_{2,\min}=\frac{1}{2}N_1$ has been investigated, but other ways to define $N_{2,\min}$ exist and deserve attention. Finally, it should be emphasized that even though the two-stage approaches are useful for situation when variabilities are unknown, applicants may in cases of high intra-subject variabilities (CV>30%) have a good

alternative by using reference-scaled average bioequivalence

CONCLUSION

- Method E controls type I errors at or below 5% when $\alpha=0.0446$ for GMR=0.95.
- Method E generally has a sample size advantage over method C/D in scenarios where method C/D have a high chance of going into stage 2. Method E is therefore an ethical and economical alternative to these methods. Similar conclusion applies when futility rules are applied.
- In terms of average total sample size, with method E, there is generally no apparent reason to start with a high sample size.

Fig. 5. Type I error rate as function of N_1 for method E without a stopping rule, for CV=20% (black circle) and CV=40% (white circle)

Fig. 7. Type I error rate as function of N_1 for method E with the stopping rule $N_{\max}=3 \times N_1$, for CV=20% (black circle) and CV=40% (white circle)

REFERENCES

1. Potvin D, DiLiberti CE, Hauck WW, Parr AF, Schuirmann DJ, Smith RA. Sequential design approaches for bioequivalence studies with crossover designs. *Pharm Stat.* 2008;7:245–62.
2. Montague TH, Potvin D, Diliberti CE, Hauck WW, Parr AF, Schuirmann DJ. Additional results for 'Sequential design approaches for bioequivalence studies with crossover designs'. *Pharm Stat.* 2012;11:8–13.
3. Fuglsang A. Controlling type I errors for two-stage bioequivalence study designs. *Clin Res Regul Aff.* 2011;28:100–5.
4. Fuglsang A. Futility rules in bioequivalence trials with sequential designs. *AAPS J.* 2014;16:79–82.