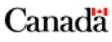# DISCUSSION PAPER

Bioequivalence Requirements: Highly Variable Drugs and Highly Variable Drug Products: Issues and Options

PLEASE NOTE: THESE PAPERS ARE INTENDED ONLY FOR DISCUSSION AT THE WORKSHOP AND EAC-BB MEETING ON JUNE 26-27, 2003 RESPECTIVELY. THEY SHOULD NOT BE INTERPRETED AS HPFB POLICY. UNTIL SUCH TIME AS A NEW OR UPDATED POLICY IS FINALIZED AND PUBLISHED, THE THERAPEUTIC PRODUCTS DIRECTORATE'S (TPD) PRACTICE WITH RESPECT TO BIOEQUIVALENCE REQUIREMENTS REMAINS UNCHANGED.

PLEASE NOTE THAT WHILE THE EAC-BB MAKES RECOMMENDATIONS TO THE DIRECTOR GENERAL, TPD, DECISION-MAKING RESPONSIBILITY REMAINS WITH THE TPD.

Canada

Outline of discussion paper

1. Brief overview of sources of variability in bioequivalence studies

2. Brief overview of the Two One-Sided Test

3. Highly variable drugs and highly variable drug products: Problems

4. Chlorpromazine, the archetypal highly variable drug

5. An example of a highly variable drug product

6. An example of a (statistical) subject by formulation interaction

7. A case of a non-linear highly variable drug

8. Possible methods of dealing with highly variable drugs or drug products

9. Conclusions

1. Sources of variability

Highly variable drugs (HVDs) have been defined as drugs in which the within-subject variability (WSV) in pharmacokinetics estimated from the ANOVA-CV equals or exceeds 30% (1). In traditional bioequivalence study designs based on 2-periods, the factors in the ANVOVA model are: Formulation, Period, Sequence and Subjects nested within Sequence. These factors account for all the identifiable sources of between subject variability. Within-subject variability is contained in the Residual Variance (also called the 'Error Mean Square or Error Term'). The residual variance is made up of several components: (i) WSV in absorption, distribution, metabolism and excretion (ADME) combined with a component of analytical variability, (ii) within-formulation variability (WFV), (iii) the subject by formulation interaction (S*F) and (iv) unexplained, random variability. The components of the residual variance cannot be subdivided further in a 2-period design. The hope is that the two products in a bioequivalence study are of good pharmaceutical quality so there is little within formulation variability and there is negligible subject by formulation interaction. An advantage of replicate designs, in which the test and reference formulations are each administered twice is that the subject by formulation interaction can be 'teased out' of the residual variance and it is possible to estimate within subject variabilities associated with the test (Swt) and reference (Swr) formulations. For example, common replicate designs with 2-sequences are (3-period) TRT vs RTR and (4-period) TRTR vs RTRT.

2. The Two One-Sided Test

The modern concept of bioequivalence is based on a survey of physicians carried out by Westlake in the 1970s which concluded that a 20% difference in dose between two formulations would have no clinical significance for most drugs. Hence bioequivalence limits were set at 80 - 120%. Plasma concentration dependent measures such as Cmax or AUC are not normally distributed; they are log normal, and hence bioequivalence limits became 80 - 125% (or $\pm\,0.225$ on the natural log scale). In traditional average bioequivalence based on the Two One-Sided Test (2), the 90% confidence interval around the geometric mean ratio of the test and reference formulations is therefore required to fall within bioequivalence limits of 80 - 125%. The width of the 90% confidence interval depends on the number of subjects in the study and the magnitude of the residual variance. The ANOVA-CV is simply the square root of the residual variance multiplied by 100.

3. Highly variable drugs and drug products: Problems

The problem with highly variable drugs is shown in Figure 1 which illustrates the results of two bioequivalence studies on formulations of drugs A and B. There are the same number of subjects in each study and the GMR is the same in both. As far as the Two One-Sided Test is concerned, the only difference between the two studies is the magnitude of the ANOVA-CV. Drug A has a low within-subject variability (ANOVA-CV 15%) and the 90% confidence interval falls comfortably within the bioequivalence limits of 80 - 125%. Drug B is highly variable, however, with an ANOVA-CV of 35%. The study fails because the lower bound of the 90% confidence interval falls below the

4

lower bioequivalence limit (80%). In other words, the study on drug B was underpowered, the simple remedy for which would be to repeat the study with a greater number of subjects. Highly variable drugs are usually safe drugs with flat dose response curves and application of the present preset bioequivalence limits of 80-125% amounts to imposition of unwarranted tougher bioequivalence requirements than for lower variability drugs.

A highly variable drug product (HVDP) is a formulation of poor pharmaceutical quality in which the drug itself is not highly variable, but there is a big component of within formulation (tablet to tablet, capsule to capsule, patch to patch) variability (WFV). HVDPs pose a problem because they are cannot be detected in traditional 2-treatment, 2-period, 2-sequence cross-over design studies. Replicate designs, however, facilitate their detection because the within-subject variabilities of the test and reference formulations can be estimated separately. When they are very different, the probable explanation is that one of the formulations is a HVDP. Now if the brand is a HVDP, then a better quality test product should not be penalized by forcing it to meet the variability of the poor quality reference product.

4. Chlorpromazine: The archetypal highly variable drug.

Table 1. summarizes the ANOVA-CVs from three different types of studies on chlorpromazine conducted at different times by three different analysts by three completely different analytical methods as outlined in the table. Study-1 was a 3-period bioequivalence study carried out in 37 subjects in which the test formulation was administered once and the reference formulation was administered twice. Study-2 was a

5

3-period pharmacokinetic study (3) in which an oral solution was administered. Solution

data gives the best estimate of true pharmacokinetic within-subject variability since, in

the absence of a formulation, there is no subject by formulation interaction, and there is

no component of within formulation variability included in the estimate. Study-3 was a

2-period interaction study in which chlorpromazine was administered with and without

quinidine. The use of solution data in study-2 shows unequivocally that chlorpromazine

behaves as a highly variable drug in terms of both Cmax and AUC. There was

remarkable consistency of the ANOVA-CVs across the three studies (Table 1) despite the

differences in study design, analytical method and analytical personnel.

　　　　Table 2 summarizes the results of the bioequivalence study on two formulations

of chlorpromazine (unpublished data) which is a good illustration of the kind of problems

that beset bioequivalence studies on highly variable drugs. Despite the large number of

subjects (n=37), all comparisons of Cmax failed US-FDA conditions which require a

90% confidence interval around this measure to be within 80-125%. The most

interesting point here is that a reference to reference comparison also failed decisively.

The reason is illustrated in Figure 2 which shows stick plots of the parametric values of

ln Cmax (left panel) and ln AUClast (right panel) for each of the 37 subjects after the two

administrations of the reference formulation. The results for six subjects are shown in

Figure 2a and highlighted in Figure 2b from which the remaining 31 subjects have been

deleted. Deletion of data from these subjects produced a substantial reduction in the

ANOVA-CVs of both Cmax and AUClast. Since chlorpromazine is a genuine highly

variable drug, it is likely that any high quality formulation of it would respond in a

similar manner to the reference formulation in our study. At present the only recourse to

the bioequivalence dilemma is to increase substantially the number of subjects in the study.

5. An example of a highly variable drug product

After a traditional 2-treatment, 2-period, 2-sequence cross-over bioequivalence study on two formulations of the beta-blocker nadolol failed, it was decided to investigate the sources of variability by conducting a 4-period replicate study. Thus a 2-treatment, 4-period, 4-sequence cross-over study was conducted in 22 healthy volunteers. The results (Table 3a) indicated a failure of both AUClast and Cmax in terms of the 90% confidence intervals failing to fall within preset bioequivalence limits of 80 - 125%, although Cmax passed the Health Canada requirement for the GMR to fall within these bioequivalence limits. Examination of the two administrations of the test formulation (Table 3b) demonstrated that the drug itself was not highly variable in terms of either measure and the GMR for both was 97%, close to the ideal. In contrast, however, the reference to reference comparison (Table 3c) showed both Cmax and AUClast very highly variable, the GMRs of both measures was 87%, and the reference formulation failed when tested against itself. The reason for the problem is illustrated graphically in Figure 3a which shows stick plots of the two values of ln Cmax after administration of the test (left panel) and reference (right panel) formulations. Three subjects with the greatest differences in Cmax after the two administrations of the reference formulation were highlighted (Figure 3a, right panel), whereas the same subjects had very much smaller differences after the test formulation (left panel). Figure 3b is the same as Figure 3a with all but the three most variable subjects deleted. These three subjects made the

greatest contribution to the formulation variability (WFV) of the reference which behaved as a highly variable drug product. This phenomenon cannot be attributed to subject by formulation interaction because the subject means for each formulation were relatively close, despite the wider spread with the reference formulation. In fact, the subject by formulation interaction term was negligible for both measures (0% for Cmax and 8% for AUClast).

This study raises some interesting questions that came to light solely because of the nature of the replicate design. After the earlier 2-period bioequivalence study in which within formulation variation and subject by formulation interaction term were inseparable components of the residual variance, it was tempting to attribute failure to subject by formulation interaction. The four period design, however, showed clearly failure was attributable to a very different problem. One could say that the 'right' answer was that the test formulation was not bioequivalent with the reference formulation simply because their variances were so different. Should this mean then, that a superior quality reference product can never reach the market?

6. An example of a (statistical) subject by formulation interaction

A bioequivalence study on two percutaneous patch formulations of nitroglycerin was carried out in 37 subjects in a 2-treatment, 4-period, 4-sequence cross over design. Serial blood samples were collected over 12 hours after which the patch was removed to facilitate measurement of the elimination phase. Table 4a shows the test formulation met bioequivalent requirements in Cmax and AUClast, despite high variability in both measures. This was a consequence of the study being adequately powered with a total of

148 observations in the data set. The test to test (Table 4b) and the reference to reference (Table 4c) comparison also met bioequivalence requirements.

Examination of the variabilities associated with the test and reference formulations (Tables 4b and 4c) showed them to be roughly comparable, but there were large subject by formulation interactions associated with both Cmax (28%) and AUClast (21%). These are depicted in Figures 4a and 4b in which the four measurements for each individual are plotted, with the two test values slightly to the left of the two reference values. In figure 4a, eight subjects contributing to the subject by formulation interaction in Cmax are highlighted, while in Figure 4b, five subjects contributing to the interaction in AUClast are highlighted. Thus it would appear that a subset of the population may respond differently to the two formulations, although the observation was not confirmed by repetition of the replicate study in the same individuals, or in a second sample of the population.

Two important points arise from this study. One is the lack of sensitivity of average bioequivalence to a substantial subject by formulation interaction. Individual bioequivalence (not shown) is very sensitive to the interaction and failed the study. The second point concerns the clinical significance of a subject by formulation interaction with any given drug, which is a difficult question to study prospectively. Problems in judging the clinical significance of subject by formulation interaction in general may have contributed to the demise of individual bioequivalence at US-FDA.

7. A case of a non-linear highly variable drug

Two bioequivalence studies were carried out on two formulations of propafenone which is a non-linear, highly variable drug. Both studies were traditional 2-treatment, 2-period, two sequence cross-over studies. In the first study, 74 healthy subjects were dosed after an overnight fast, and in the second, 25 healthy subjects were dosed after a standardized high fat breakfast. The first study was successful in that both measures met bioequivalence requirements despite high within-subject variability (Table 5a) The Fed study was not powered sufficiently for the 90% confidence interval around ln Cmax to fall within bioequivalence limits of 80-125% (Table 5b). The question remains, however, should it be necessary to subject 74 healthy volunteers to two doses of propafenone in order to test two formulations of the drug for bioequivalence?

8. Possible methods of dealing with highly variable drugs and drug products

a) No confidence interval for Cmax of HVDs/HVDPs

Cmax is often the most variable of the two measures, partly because it is a single point determinant which is dependent upon an adequate blood sampling schedule around tmax. A simple method of dealing with highly variable drugs would be to treat them as 'uncomplicated drugs' and not require a 90% confidence interval around Cmax. Each of the four examples discussed in this paper were also highly variable in terms of AUClast but its variability was less than Cmax (in test versus reference comparisons) which means a smaller number of subjects (observations) would have been required to achieve adequate statistical power.

b) Arbitrarily broaden the bioequivalence limits for HVDs

Since highly variable drugs are generally safe drugs with shallow dose response

curves, it is reasonable to tolerate greater than 20% differences between test and

reference formulations. The EU-CPMP guidelines, for example, permit a sponsor

prospectively to justify broadening the bioequivalence limits from 80 - 125% to, say, 75 -

133%.

c) Broaden the bioequivalence limits according to the within-subject variability of the

reference formulation

The bioequivalence limits can be scaled to the within-subject variability of the

reference formulation by the use of the residual variance in a two-period design or the

within-subject variance associated with the reference formulation in a replicate design.

The fundamental concept is shown in Equation 1. which implies the 90% confidence

interval around the difference between the log transformed means of the test and

reference formulation must fit between bioequivalence limits of θ_{ABE} which is set by the

drug regulatory body. The commonly accepted bioequivalence limits are set at 80 -

125% (0.8 - 1.25) which is ±0.225 on the natural log scale. In scaling, the

bioequivalence

$$-0.223 \leq (\mu_t - \mu_r) \leq 0.223 \qquad \therefore (\mu_t - \mu_r)^2 \leq 0.223^2 \qquad \text{Equation} \quad 1.$$

limits are divided by the within-subject standard deviation at which the limits are to be

permitted to be broadened (σ_{w0}). The latter is to be set by a drug regulatory agency. The

left hand side of Equation 1 is divided by the within subject standard deviation of the reference formulation (σ_{WR}) Equation 2. Rearrangement of Equation 2 gives Equation 3 which broadens the bioequivalence limits for highly variable drugs in a systematic way.

$$\frac{(\mu_t - \mu_r)}{s_{wr}} \leq \frac{0.223}{s_{w0}} \qquad \text{Equation 2.} \qquad (\mu_t - \mu_r) \leq \frac{0.223}{s_{w0}} \times s_{wr} \qquad \text{Equation 3.}$$

The method of Boddy and Co-workers (4) for widening the bioequivalence limits for highly variable drugs/products is a slightly different approach to the same concept. Here the bioequivalence limits are set as a fixed multiple (k) of the standard deviation (σ_{WR} or σ_{RES}). In other words, ($\theta_{ABE} = k\sigma_{WR}$), such that k represents the number of standard deviations by which the means are allowed to vary. Thus, when k is set at ($0.223/\sigma_{w0}$) the relationship becomes exactly the same as Equation 3.

Scaling the ABE metric amounts to the same thing as scaling the bioequivalence limits, since the relationships shown in Equations 1-3 are used for both methods. For the purposes of illustration, the point at which the bioequivalence limits are permitted to be broadened by scaling (σ_{w0}) was set at 0.20 or 0.25. The results are shown in Figure 5, and some specific examples with Swt = 0.3, 0.5 or 0.7 are shown in Table 6. The value σ_{w0} = 0.20 was selected by US-FDA during the decade long debate on individual bioequivalence, although for scaled ABE, we prefer the more conservative σ_{w0} = 0.25.

9. Conclusions

a) Scaled versus unscaled average bioequivalence

Unscaled average bioequivalence is very sensitive to differences between the means whereas scaled bioequivalence is much less sensitive. Scaling can allow the geometric mean ratio to rise to unacceptably high levels when the reference variance is very high, and it may be wise to place a limit on the maximum GMR for bioequivalence. For example, the GMR could be required to fall within 80 - 125%. If reference scaling is to be allowed, TPD must set σ_{w0}. We suggest $\sigma_{w0} = 0.25$ since $\sigma_{w0} = 0.20$ appears to be too liberal. (However, σ_{w0} can be set at any reasonable value)

It is our view that average bioequivalence based on replicate designs has merit because valuable evidence on the pharmaceutical quality is provided in that disparities between the test and reference variances are made apparent. Substantial subject by formulation interactions, if present, are also detected, but TPD may prefer to ignore them.

b) Subject by formulation interaction

As illustrated by the example of the bioequivalence study on the two patch formulations of nitroglycerin, average bioequivalence is not sensitive to the presence of a substantial subject by formulation interaction whereas individual bioequivalence is very sensitive. When a study is conducted by replicate design, however, a substantial subject by formulation interaction will be detected, even though average bioequivalence is not sensitive to it. If an interaction is deemed to be important clinically, then a regulatory agency can take any action, including failing the study because of it.

c) Highly variable drug products.

        As stated earlier, an advantage of replicate design is that data are provided on the within-subject variabilities of the test and reference formulations. If the test formulation is shown to be a highly variable drug product, then it should be allowed to fail average bioequivalence without resort to scaling. When an innovator formulation behaves as a highly variable drug product, the problem is more difficult to resolve because that was the formulation linked to pivotal clinical trials during the therapeutic and toxicological evaluation of the drug substance. By the time generic formulations appear, the innovator's highly variable drug product will have been on the market for several years, presumably without major untoward incident. One could surmise then that introduction of a better quality, less variable generic product would pose no hazard to the patient. Therefore we would recommend reference scaling be allowed when the reference formulation is a highly variable drug product.

1.   H. H. Blume, and K. K. Midha. Bio-International '92, Conference on Bioavailability, Bioequivalence and Pharmacokinetic Studies. *Pharmaceutical Research* **10**:1806-1811 (1993).

2.   D. J. Schuirmann. A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *J. Pharmacokinet. Biopharm* **15**:657-680 (1987).

3.   P. K.-F. Yeung, J. W. Hubbard, E. D. Korchinski, and K. K. Midha. Pharmacokinetics of chlorpromazine and key metabolites. *European Journal of Clinical Pharmacology* **45**:563-569 (1993).

4.   A. W. Boddy, F. C. Snikeris, R. O. Kringle, G. C. G. Wei, J. A. Oppermann, and K. K. Midha. An approach for widening the bioequivalence acceptance limits in the case of highly variable drugs. *Pharmaceutical Research* **12**:1865-1868 (1995).