J. Pharm. Pharmacol. 1996, 48: 11–16
Received November 13, 1995
Accepted November 27, 1995

Bioequivalence and Generic Prescribing: an Industrial View*

ALWYN PIDGEN

Biometrics Department, Pfizer Central Research, Sandwich, Kent CT13 9NJ, UK

The topic of bioequivalence has stimulated much discussion over the past 25 years, resulting in numerous symposia, papers and changes to the regulatory guidelines. The focus of the discussions has changed over time (statistical, pharmaceutical, pharmacokinetic and clinical) but the constant theme has been to influence regulatory authorities and their thinking. In fact, the current regulatory guidelines have led to an almost standard approach to the design, analysis and interpretation of bioequivalence studies world-wide. However, because of governmental pressure (largely economic) the topic still continues to be the centre of much debate with calls for even more changes and refinements to the guidelines. In fact, this topic has probably generated more meetings and produced more papers than any other single topic in pharmaceutical medicine.

The impact of these meetings and papers has generally been to raise quality and awareness; however, we need to be careful to ensure that we are focusing on the important issues.

Regulatory Guidelines–the Standard Bioequivalence Trial

The standard bioequivalence trial is conducted according to a crossover design in between 12 and 24 healthy normal male adults with an appropriate washout period. The subjects are matched for age and weight. Single doses of a drug are administered after an overnight fast and food/fluid intake is controlled throughout the study. Blood samples are taken in order to obtain a concentration vs time profile in each individual for each formulation. Pharmacokinetic characteristics of these blood level curves notably C_{max}, T_{max}, AUC and k_{el} are then derived and compared between treatments using statistical procedures. Other kinetic parameters may be obtained if appropriate.

Multiple dose studies may be required for several reasons: there may be problems of assay sensitivity which prevent sufficiently precise plasma concentration measurements being obtained after a single dose; intra-individual variability in the plasma concentrations following a single dose may be inherently large; if a drug exhibits dose- or time-dependent pharmacokinetics; in studies on extended-release products.

A similar approach to that described above is used for multiple dose studies. Pharmacokinetic parameters obtained will generally include C_{max}, T_{max} (after last dose), AUCo-τ, C_{min}, C_{av} and the degree of fluctuation (DF).

Regulatory Guidelines–Statistical Issues

Much of the focus in the 1970s and early 1980s was on the statistical aspects of comparing two or more formulations in bioequivalence studies. In those days, almost all decision making in bioequivalence studies was based on an hypothesis testing (or significance level) approach. Unfortunately this approach resulted in a number of anomalies, such as large (clinically important) differences between formulations which were not statistically significant, and small (clinically unimportant) differences which were statistically significant.

These problems arose mainly because of the wide variations in sample sizes which were used in bioequivalence trials. If a small sample size was used (e.g. n = 6) in a trial on a drug with a large intra-subject variance, then any real differences between the formulations would not be detected statistically. On the other hand, the use of a large sample size (e.g. n = 24) in a trial on a drug with a small intra-subject variance would ensure that any differences between the formulations, however small, would be detected.

The challenge was then on to find a statistical technique which overcame these problems and several different approaches were proposed, including:-

80/20 power rule (FDA proposal)
comparing concentrations at individual time points;
split-plot analysis;
75/75 (or 75/125) rule (FDA proposal);
classical confidence intervals;
symmetrical confidence intervals (Westlake);
non-parametric confidence intervals;
Bayesian approaches;
2 one-sided *t*-tests approach.

The technique which best met the requirement of enabling more clinical relevance to be brought into decision making was the classical confidence-interval approach. This approach overcame many of the problems associated with the significance testing methods. In this approach one can determine with a specific level of confidence, that the bioavailability of the test formulation will fall within certain limits of the bioavailability of the reference formulation (e.g. ± 20%). There was some debate regarding the appropriate level of confidence which should be used (90 or 95%) until the authorities settled on a 90% confidence-interval approach. The two one-sided *t*-tests procedure (Schuirmann 1987) is operationally identical and may be used in place of the 90% confidence-interval approach.

One further change to the regulatory guidelines involved logarithmically (ln) transforming the data. There had long been a debate about the relevance of logarithmically transforming pharmacokinetic parameters before statistical analysis (Westlake originally raised this issue in the early 1970s). It is now recommended in both the US and European guidelines that C_{max} and AUC are logarithmically trans-

*Presented at the British Pharmaceutical Conference, University of Warwick, September 15-18, 1995.

formed before statistical analysis, resulting in acceptance limits for the test formulation of between 80 and 125% compared with the reference formulation.

No clear guidance is given for statistically analysing T_{max} (or k_{el}). If the assumption of a normal or log-normal distribution is considered doubtful then non-parametric procedures can be used throughout.

Further discussion on the topic of statistical aspects of bioequivalence studies can be found by reference to Pidgen (1992) and Chow & Liu (1994).

Open Questions

From the numerous symposia and papers on the topic of bioequivalence a number of questions remain to be answered:

- can the 80-125% acceptance criteria be routinely applied to the majority of drugs or should the authorities set limits on an individualized drug by drug basis?
- should there be a category of highly variable drugs and if so what criteria should be used?
- are the classical pharmacokinetic parameters of C_{max}, T_{max} and AUC sufficient to describe the rate and extent of absorption in bioequivalence studies? Do we need a fresh approach?
- should bioequivalence studies be performed using replicated-treatment crossover designs for comparing a test and reference product?
- should bioequivalence assessment continue to be based upon the concept of average bioequivalence or should we move to a concept based upon individual bioequivalence?

Bioequivalence–The Fundamental Issue

All of the above questions clearly have some relevance to the topic of bioequivalence based on satisfying the current regulatory guidelines.

However, there is a more fundamental issue which needs to be addressed first: what is the key objective in bioequivalence testing?

Is it purely to assure pharmaceutical quality with respect to drug release characteristics and subsequently drug absorption? Is it to assure clinical safety and efficacy? Or is it both?

To attempt to answer this, let us first examine the basic definition of bioequivalence (as given in the European guidelines), which states:-

two medicinal products are bioequivalent if they are pharmaceutical equivalents or alternatives and if their bioavailabilities (rate and extent) after administration of the same molar dose are similar to such a degree that their effects, with respect to efficacy and safety will be essentially the same.

Now, in order to unequivocally demonstrate that two or more drug products have similar safety and efficacy profiles we would need to perform a large-scale clinical trial. This would be expensive, time-consuming and possibly have a doubtful ethical motive. So, we replace this direct (clinical) approach with an indirect (pharmacokinetic) approach of the bioequivalence trial, based on the principle that:-

two (or more) formulations of a drug that give essentially equivalent concentrations of the active species in blood (viewed as a profile over time) will give essentially similar safety and efficacy profiles.

In other words, once the drug is absorbed into the systemic circulation, the dosage form can no longer influence the fate of the drug and the clinical response is formulation independent.

But, is this true? Does bioequivalence actually mean therapeutic and safety equivalence? Or put another way, is there a relationship between circulating plasma drug (and/or active metabolite) concentrations and clinical effect?

For the current bioequivalence criteria to be effective, then the above statements have to be true. However, for most classes of drugs, belief is more readily available than proof in this particular subject area. Although there is increasing interest and research underway in the area of pharmacokinetic-pharmacodynamic modelling there are still very few papers which show a well-defined relationship between drug (and/or active metabolite) concentration and therapeutic effect (or safety).

Therefore, the whole area of bioequivalence and its clinical consequences is based largely upon a hypothesis which for many classes of drug has neither been conclusively nor partially proved. Until this proof is obtained, or long-term clinical experience built up with a particular drug product, there will still be the concern that two or more formulations of the same drug may not be equally safe and effective even after completion of a successful (in regulatory terms) bioequivalence study.

The Open Questions

Can the 80-125% acceptance criteria be routinely applied to the majority of drugs or should the authorities set limits on an individualized drug by drug basis?

Before attempting to answer this question we first need to examine the clinical rationale behind the acceptance criteria used by the regulatory authorities in bioequivalence studies. It could be argued that the majority of drugs which make it to the market-place have wide therapeutic windows and as such will not be affected by small or moderate changes in the rate and extent of absorption. Also, drugs with narrow therapeutic windows, either never make it to the market-place or will have narrower acceptance criteria, and will be subject to careful therapeutic monitoring (e.g. cyclosporin, phenytoin). So where is the issue? Why is one acceptance criterion not sufficient for all drugs?

In reality, the lack of suitable data on concentration-effect relationships makes it difficult to judge whether changes in the rate and extent of absorption would adversely affect clinical safety, or effectiveness, or both. It is likely to depend upon the characteristics of the individual drug and patient. In their presentation to the FDA at the 1986 bioequivalence hearing in Washington, the Pharmaceutical Manufacturers Association (PMA) proposed dropping the general rule in favour of an individualized drug by drug approach, whereby the acceptance limits would be determined by the regulatory authority at the expiry of the patent of the innovator drug. So far, this proposal has not been taken up by the regulators.

Surely, in order to better understand the clinical consequences of generic substitution and remove speculation we must make more progress in the area of pharmacokinetic-pharmacodynamic modelling. The pharmaceutical industry as a whole has a responsibility here and could itself derive considerable benefit by both generating and publishing more of this type of data on its drug products.

A good example of the value of pharmacokinetic-pharmacodynamic modelling was presented by Daley-Yates et al (1994) when comparing two prodrugs of methylprednisolone. In the classical pharmacokinetic context the prodrugs were found to be bio-inequivalent. Although the AUC fell within the 80-125% criteria, C_{max} and T_{max} did not. However, by utilizing pharmacokinetic-pharmacodynamic modelling techniques, the authors were able to demonstrate statistically that the changes observed in rate of absorption had no clinical consequences.

Although not their primary role, we also need to ask what use is made by the regulatory authorities of the information they hold on drug products which are submitted by the original innovator companies. In order to approve drugs for marketing the regulators have to thoroughly review these data and form their own opinions on key issues. One of these must be to assess the therapeutic window and its potential clinical consequences. In these days of electronic submissions, relevant information could be extracted from these population databases and (at the expiry of the patent) be published for the benefit of the whole industry. Regulators are in the unique position of not only obtaining information on a particular drug substance but also on drug classes.

It should not be difficult for regulators to obtain published estimates of the inter- and intra-subject variability in the pharmacokinetics of a drug in volunteers and patients. At the same time, regulators could also publish their assessments of therapeutic windows and the likely impact on issues such as bioequivalence acceptance criteria for C_{max}, T_{max} and AUC. However, in order to go down this route, the regulators (particularly the FDA) may have to accept the principle of having different acceptance criteria for different drug classes, rather than across-the-board criteria.

Should there be a category of highly variable drugs and if so what criteria should be used?

A proposal has been made for regulatory authorities to allow different acceptance criteria for drugs classed as having highly variable pharmacokinetics. The proposal is to class those drugs which exhibit an intra-subject variability of more than 30% as highly variable drugs.

Clearly there are difficulties when assessing the bioequivalence of highly variable drug products, particularly for the parameter C_{max}. One key consequence is the increased sample size which would be necessary to meet the acceptance criteria of 80 to 125%. An example is shown in Table 1, for sample sizes needed for 80% power to fall within 80-125%. Are we seriously suggesting that companies should use these large subject numbers? Clearly, there is an issue here as to where to draw the line. There is a real danger in dealing with this issue on purely statistical rather than on clinical grounds.

In order to be able to answer this question properly we have to return again to the fundamental issue: does a drug (or drug product) which exhibits high intra-subject variability in its pharmacokinetics also exhibit high variability in its clinical response?

Table 1. Sample sizes needed for 80% power to fall within 80-125%.

	Ratio of means (%)				
	90	95	100	105	110
CV = 30%	80	40	32	38	68
CV = 40%	140	70	56	66	120

In other words do we know the relationship between concentration and effect ?

It may be that for some drugs 30, 40 or even 50% variability in pharmacokinetics will have little or no clinical consequences and the current regulatory acceptance limits could be widened. However, for other drugs, this variability may be critical and to apply this 30% criterion across the board as a general principle without knowledge of the therapeutic window is potentially risky. Also, since the variability between subjects is generally much greater than that within subjects, should we not be looking for a different approach to assess bioequivalence?

Are the classical pharmacokinetic parameters of C_{max}, T_{max} and AUC sufficient to describe the rate and extent of absorption in bioequivalence studies? Do they have clinical relevance?

The regulatory guidelines recommend derivation of pharmacokinetics as shown in Table 2 for bioequivalence studies.

The relevance of using the standard pharmacokinetic parameters of C_{max}, T_{max} and AUC as metrics for assessing the rate and extent of absorption has been questioned by a number of authors including Endrenyi et al (1991), Tozer & Bois (1995) and Lacey et al (1995). Some of the issues raised and recommendations made are summarized below.

Rate of absorption. C_{max} is considered to be a consistently poor estimator of rate of absorption, but is of value in assuring safety of drug products. It is, however, confounded by the extent of drug absorption.

T_{max} is often used when there are clinical indications that the rate may be important. However, there are some problems, in that T_{max} is heavily dependent upon sampling times, it can often be a highly variable parameter (one of the reasons for recommending non-parametric approaches) and can lead to very wide confidence intervals.

Table 2. Recommended pharmacokinetic parameters for bioequivalence testing.

European
C_{max}, T_{max}, AUC_{τ}, AUC_{inf}, Ae, Ae_{inf}, dAe/dt (single dose), Css_{max}, Css_{min} and AUC_{τ}(multiple dose)
FDA
C_{max}, T_{max}, AUC_t, AUC_{inf}, $t_{1/2}$ (single dose), AUC_{τ}, C_{max}, T_{max}, C_{min}, C_{av}, DF (multiple dose)

From a pharmacokinetic point of view, a sensitive measure of changes in rate of absorption was found to be C_{max}/T_{max}, although it may be highly variable.
C_{max} /AUC has also been proposed for immediate release dosage forms and is accepted as a good measure of pharmaceutical quality, but has no clinical relevance and would be difficult to interpret (i.e. setting relevant acceptance criteria). It has therefore been concluded that there is no rate parameter which allows products to be compared for both pharmaceutical quality and clinical safety and efficacy.

Extent of absorption. AUC (measured to the last quantifiable concentration (t) is recommended as the most acceptable measure of extent of absorption. Some problems still remain; for example, area under the curve extrapolated to infinity (AUC_∞) behaves poorly when the assay sensitivity is low. For drugs with a very long half-life then AUC can be calculated to the last measured time point or to a predetermined time point as described by Urso & Aarons (1983).

Comparing concentration-time curves. A question which has been raised recently is: should we be deriving pharmacokinetic parameters in the first place or should we actually be comparing the two (or more) concentration vs time curves?

An alternative definition of bioequivalence reflecting this was proposed at a 1994 Bio-International conference in Munich: two pharmaceutical products are considered to be essentially the same when their concentration vs time profiles are so similar that they are unlikely to produce clinically relevant differences in therapeutic or adverse events.

This definition has the advantage that it makes no mention of which pharmacokinetic parameters are used to assess rate and extent, but raises the question as to whether we actually have statistically valid techniques to enable these curves to be compared. The only author to propose an approach for comparing the concentration vs time curves statistically was Westlake with his split-plot analysis (Westlake 1973). However, the approach failed to gain acceptance partly for technical statistical reasons (high correlation of early time points) and partly because there was no agreed clinically relevant criteria on which to accept or reject bioequivalence. This issue was also discussed in a recent paper by Salmonson (1995) although the author came up with no new ideas to resolve this problem.

Tozer & Bois (1995) raised the issue of how to measure the shape of the curve. They stated that optimal measures of shape may depend upon whether the input-time profile is important to any therapeutic use of the drug. If rate is important, for a condition requiring a rapid onset of effect, then the concentration profile to T_{max} may be of particular concern. A concentration near the mean residence time of the reference formulation may also be useful.

Should bioequivalence studies be performed using replicated-treatment crossover designs for comparing a test and reference product?

There has been a generally increasing interest in the use of replicated treatment crossover designs for comparing a test and reference product in bioequivalence trials. The whole issue was put into perspective by Donald Schuirmann of the US Food and Drug Administration at a recent AAPS/FDA workshop in March 1995. In replicated treatment studies at least some of the subjects receive at least one of the products more than once. These designs are considered to have a number of advantages over standard two-treatment, two-period, two-sequence crossover studies. However, under current regulatory requirements in the US the only regulatory advantage of these studies is that they permit the sponsoring firm to obtain more observations for the same number of subjects. This single regulatory advantage is balanced by a number of disadvantages including, higher cost, longer duration, increased risk of dropouts and more complicated statistical analysis.

One supposed advantage of these designs, the ability to obtain an estimate of pure intra-subject variance and use this to obtain narrower confidence intervals, is a fallacy. This is because the pure intra-subject variance estimate is not the appropriate error term for computing confidence intervals for these designs.

Another supposed advantage is when computing a confidence interval for the reference vs reference. It is mistakenly believed that if this interval is wider than the standard acceptance limits (80-125%) then the confidence interval for the test vs reference does not have to meet the usual acceptance limits. Under current (March 95) regulatory requirements, if the study is unable to establish the equivalence of identical reference products (they will come from the same lot) it means that the study is inadequate to establish bioequivalence under current requirements.

Should bioequivalence assessment continue to be based upon the concept of average bioequivalence or should we move to a concept based upon individual bioequivalence?

Another topic of debate centres around the current procedures for assessing the bioequivalence of two formulations using the concept of average bioequivalence. That is, they assess whether the average responses between individuals on the two formulations are similar. Anderson & Hauck (1990) considered that average bioequivalence was not sufficient to guarantee that an individual patient could be expected to respond similarly to two (or more) formulations. To have reasonable assurance that an individual patient could be expected to switch from a therapeutically successful formulation to a different formulation (e.g. generic substitute) these authors proposed a different notion of bioequivalence, which they referred to as individual (or within-subject) bioequivalence. Anderson & Hauck (1990) also proposed a simple, valid statistical procedure for assessing individual bioequivalence. The basic idea behind individual bioequivalence is that the bioavailability of the new formulation will be sufficiently close to that of the current formulation in most individuals.

Average bioequivalence. Two formulations are average bioequivalent when the bioavailability of the new formulation, averaged over some appropriate population, is sufficiently close to the average of the reference formulation. The corresponding statistical hypothesis is:

$$1 - R_A \leq \mu_T/\mu_R \leq 1 + R_A$$

This is the type of bioequivalence that is currently assessed. (T and R are the population average (log) bioavailabilities for the test and reference products respectively. Typically $1-R_A = 80\%$ and $1+R_A = 120\%$ (or 125% following logarithmic transformation). Average bioequivalence is a special case of population bioequivalence, namely the similarity of the distributions in the population of responses to the two formulations.

Individual bioequivalence. Two formulations are deemed individual bioequivalent if the bioavailability of the new formulation is sufficiently close to that of the reference formulations in most individuals. The basic idea is that most individuals will be expected to have similar bioavailabilities on the two formulations in order to call them bioequivalent.

Let T denote the test and R the reference formulations and let X_{ij} be the measure of bioavailability for formulation i given to subject j. The most general idea underlying switchability is that the test and reference values should be similar in a given subject. The corresponding hypothesis is:

$$Ps = Pr[C_1 < X_{Tj}/X_{Rj} < C_2]$$

where Ps should be high for bioequivalent formulations and C_1 and C_2 are limits that need to be determined. First take $C_1 = 1/C_2$ so that the interval will be symmetrical on the log scale. Second, pick C_2 ($= C$) by considering a comparison of the reference to itself. The required criterion for the test formulation is one the reference formulation can satisfy when compared with itself.

The approach taken is to pick C to attain a specified value of Ps for the reference compared with itself and then see how much Ps drops when comparing test with reference. In particular, Ps is set at 0.9 for a comparison of the reference to itself. Under the assumption of bivariate normality in the log scale, C is approximated by $C^* = 10CV_{WR}$, where CV_{WR} is the within-subject coefficient of variation of the reference formulation.

The next step is to determine the conditions that drop Ps below an acceptable level. Assume $Ps \geq 0.8$ to be acceptable. The following conditions are necessary to ensure that $Ps > 0.8$. In all cases it is assumed that the means of the two formulations are equal.

Condition 1. The within-subject standard deviation (log scale) of the test formulation cannot be more than 50% greater than that of the reference. This is one-sided, since only increased within-subject variability can decrease switchability.

Condition 2. The absolute difference of the between-subject standard deviations cannot be more than 114% of the reference's within-subject standard deviation. This is two-sided, since different between-subject variabilities in both directions decreases switchability, although lower between-subject variability of the test has less effect on switchability than higher between-subject variability.

Condition 3. The within-subject correlation required to assure switchability increases with the between-subject variability. This is particularly noticeable if the within-subject variability is low.

The two notions of bioequivalence correspond to two distinct clinical contexts. In the first context (called prescribability) a patient is started on a new drug. Here average bioequivalence is deemed appropriate since the clinician has no information on that individual's response. In the second context (called switchability), the patient who has been taking a particular product is switched to a new formulation by the clinician (or perhaps by the pharmacist in areas permitting generic substitution). Here it is necessary to have reasonable assurance that the patient will get the same efficacy from the new formulation, thus requiring individual bioequivalence.

This issue is being examined closely by the FDA.

Summary

The whole concept of bioequivalence is based upon the existence of a clear relationship between drug concentration and clinical effect.

To date there are insufficient data available in the form of publications to support this concept.

Both the pharmaceutical industry and the regulatory authorities could do more to promote this issue and publish relevant information.

The pharmaceutical industry could provide more information on concentration-effect relationships in volunteers and patients.

Upon expiry of the patent, regulators could provide estimates of the inter- and intra-subject variability in the pharmacokinetics of a drug in volunteers and patients, asessment of therapeutic windows for drugs and drug classes and their impact on bioequivalence acceptance criteria.

Current regulatory guidelines refer to rate and extent of absorption BUT there is no rate parameter which allows products to be compared for both pharmaceutical quality and safety and efficacy.

References

Anderson, S., Hauck, W. W. (1990) Consideration of individual bioequivalence. J. Pharmacokinet. Biopharm. 18: 259–273

Chow, S.-C., Liu, J.-P. (1994) Recent statistical developments in bioequivalence trials - a review of the FDA guidance. Drug Information J. 28: 851–864

Daley-Yates, P. T., Gregory, A. J., Field, L. A., Brooks, C. D., Parker, T. J. (1994) Pharmacokinetic and pharmacodynamic assessment of bioavailability for two prodrugs of methlyprednisolone. In: Danhof, M., Peck, C. C. (eds) Measurement and Kinetics of In-vivo Drug Effects. Leiden/Amsterdam Centre for Drug Research

Endrenyi, L., Fritsch, S., Yan, W. (1991) C_{max}/AUC is a clearer measure than C_{max} for absorption rates in investigations of bioequivalence. Int. J. Pharmacol. Ther. Toxicol. 29: 394–399

Hauck, W. W., Anderson, S. (1995) Author, not in text? Delete?Individual vs average bioequivalence: when is average not sufficient? In: Blume, H. H., Midha, K. K. (eds) Bio-International 2; Bioavailability, Bioequivalence and Pharmacokinetic Studies. Medpharm Scientific Publishers, Stuttgart

Lacey, L. F., Keene, O. N., Bye, A. (1995) Evaluation of different metrics as measures of rate of drug absorption. In: Blume, H. H., Midha, K. K. (eds) Bio-International 2; Bioavailability, Bioequivalence and Pharmacokinetic Studies. Medpharm Scientific Publishers, Stuttgart

Pidgen, A. W. (1992) Statistical aspects of bioequivalence - a review. Xenobiotica 22: 881–893

Salmonson, T. (1995) Measurement of bioequivalence and its relevance to the clinical situation. In: Nimmo, W. S., Tucker, G. T. (eds) Clinical Measurement in Drug Evaluation. John Wiley and Sons Ltd, pp 195–204

Schuirmann, D. J. (1987) 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

Tozer, T. N., Bois, F. Y. (1995) Metrics of absorption: simulation approach. In: Blume, H. H., Midha, K. K. (eds) Bio-International 2; Bioavailability, Bioequivalence and Pharmacokinetic Studies. Medpharm Scientific Publishers, Stuttgart

Urso, R., Aarons, L. (1983) Bioavailability of drugs with long elimination half-lives. Eur. J. Clin. Pharmacol. 25: 689–693

Westlake, W. J. (1973) The design and analysis of comparative blood level trials. In: Swarbrick, J. (ed.) Dosage Form Design and Bioavailability. Lea and Febiger, Philadelphia, pp 149–179