Considerations in the Attainment of Steady State: Aggregate vs. Individual Assessment

Walter W. Hauck,[1,4] Thomas N. Tozer,[2] Sharon Anderson,[1] and Frédéric Y. Bois[3]

Received April 20, 1998; accepted August 13, 1998

KEY WORDS: multiple-dose clinical pharmacology trials; steady state pharmacokinetics; aggregate evaluation; individual evaluation.

INTRODUCTION

Multiple-dose clinical pharmacology trials often include assessment of pharmacokinetics at steady state. In the analysis of these trials, the sponsor desires some assurance that plasma concentrations on the prespecified sampling day actually reflect steady state. The purpose of this note is to address statistical considerations for assessing attainment of steady state, beginning with general issues.

Steady-state, the condition in which the amount lost from the body during a dosing interval equals the amount input on a regimen of fixed dose and dosing interval is, in theory approached but never achieved. The question is how close one must be to declare practical achievement of steady state. From a clinical perspective, 90% of the theoretic steady-state value is often used as a practical definition, as the difference in response to a 10% difference in concentration can rarely be assessed (1). This definition is used throughout this paper, although other choices of a percentage are possible. To avoid redundancy, the phrase "attained steady state" should be taken to mean "attained at least 90% of steady-state value."

The approach to steady state can be evaluated for a group of individuals *as a group* (an aggregate evaluation) or it can be evaluated for each individual. Whether most or all subjects, or only the average value (across subjects) needs to have achieved at least 90% of the steady-state value is open to question.

The difference between individual and aggregate approaches becomes apparent when pharmacokinetic behavior is highly variable among subjects as shown in Fig. 1. Subjects reach their steady states at different times. The mean time required to reach 90% (14.7 days) does not equal the time (16.3 days, shown as the solid circle) that the mean concentration reaches 90% of the mean steady-state concentration. In this example, three of ten subjects reach steady state after the mean for the group does.

[1] Biostatistics Section, Division of Clinical Pharmacology, Thomas Jefferson University, Philadelphia, Pennsylvania 19107-5244.

[2] Department of Biopharmaceutical Sciences, School of Pharmacy, University of California, San Francisco, California.

[3] Lawrence Berkeley Laboratory, Berkeley, California.

[4] To whom correspondence should be addressed. (e-mail: w_hauck@lac.jci.tju.edu)

The appropriate designs and methods for assessing attainment of steady state depend on this fundamental choice between individual and aggregate approaches. Unfortunately, the choice is not an easy one. We favor individual evaluation; basing steady state on the average just does not control events at the individual level as seen in Fig. 1. However, there are difficulties, discussed later, with individual evaluation of steady state that lead to consideration of an aggregate approach.

KINETIC CONSIDERATIONS

One cannot assess steady state from observations during only one dosing interval (unless all absorbed drug is eliminated during the interval). One must have information on the kinetics of the drug either from prior studies or from the trial itself. Kinetic information to predict steady state may be obtained from observations following the first dose in the trial, during the course of administering the multiple doses, or during the washout of the last dose. Description of methods for doing so is beyond the scope of this note. Suffice it to say, they may be model based or nonmodel based. They may utilize the concentration-time profile, comparisons of single-dose to multiple-dose areas, the time to 90% of single-dose $AUC_{(0-\infty)}$, or 90% of $AUC_{(0-\infty)}$ after the last dose. In this paper we restrict attention to verification of steady state based on predictions of steady state AUC.

STATISTICAL CONSIDERATIONS

In this section, we consider statistical methods for comparing observed AUC's to predicted steady state AUC's and critique current statistical approaches.

Aggregate Evaluation, Current Methods

There are several ANOVA-based statistical approaches for aggregate assurance of steady state from studies without either a single-dose component or data following the last dosing interval (2). One approach is to employ the overall F-test for all days as the statistical test. If that test of the null hypothesis of the equality of the data on all sampling days is not statistically significant, steady state is concluded. If the overall F-test is rejected, it may be followed by multiple comparisons of the various days and a judgment regarding steady state achievement on the last sampling day.

These ANOVA approaches, however, are based on the fallacy of concluding the null hypothesis on the basis of a difference that is not statistically significant. The inappropriateness of this has been learned in bioequivalence trials (3–4). One particular consequence of this fallacious approach is that the smaller the number of subjects in the study and the greater the intra-subject variability, the lower the power of the statistical tests, the easier it is to conclude steady state and, hence, the greater is the likelihood of a false assurance of steady state. The use of post hoc power calculations does not overcome these concerns and is problematic (5). Confidence intervals are better means of judging what can be concluded.

A second problem with the current aggregate approach occurs when only a limited period of observation is available to assess steady state. In Fig. 2A, the period of sampling is

0724-8741/98/1100-1796$15.00/0

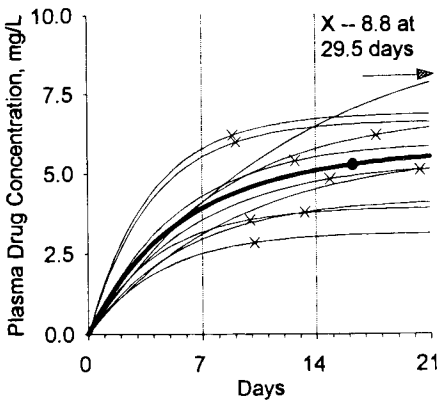

Fig. 1. The approach to steady state in a group of subjects in whom both clearance and volume of distribution are highly variable (coefficient of variation of 40% for each). The curves are simulated using a one-compartment model with mean values of clearance and volume of distribution of 2.5 L/day and 15 L, respectively, and a constant rate of input of 12.5 mg/day. The mean half-life is 4.41 days, thus steady state is expected to be achieved in 14.6 days. The time each subject and the mean data reached steady state are noted by a crossmark and filled circle, respectively.

equal to the presampling period. Steady state is just achieved on Day 15. However, variability in the observed concentrations (troughs in this example) may obscure a monotonic approach to steady state and lead one to conclude that steady-state is not yet achieved. In Fig. 2B the period of sampling is small relative to the length of the presampling period. There is little change in the concentration during this period, because it is much shorter than the half-life of the drug. Steady-state may be declared even though it is not achieved.

Aggregate Evaluation, Suggestions

For studies including single-dose data, we suggest a simple aggregate approach motivated by current methods for bioequivalence assessment. There are two steps. First we validate dose linearity. Then, we address attainment of steady state. Appropriate statistical tests should specify attainment of steady-state as the alternative hypothesis and "not-at-steady-state" as the null hypothesis (6). By setting attainment of steady-state as the alternative hypothesis, one is able to control the rate (probability) with which one falsely claims attainment of steady-state.

For each subject, based on some kinetic method, we obtain a predicted AUC for the final sampling period of the multiple-dose component from the single-dose data. If the pharmacokinetics are linear, then, on average (across subjects), the observed AUC on the final day should be "close" to the predicted value; any deviation from linearity should produce a systematic shift of predicted values from observed. Thus, we need to show that the observed values are sufficiently similar, or "equivalent," to the predicted values. We suggest a comparison of mean (or median) predicted values to the mean (or median) observed values. Assuming that each subject's prediction is statistically independent of those for the other subjects, as would be true if each individual's prediction is based solely on that individual's single-dose data, the statistical equivalence methods are based on paired-t tests or Wilcoxon signed rank tests, similar to those for bioequivalence (4,7). As in bioequivalence, a log scale

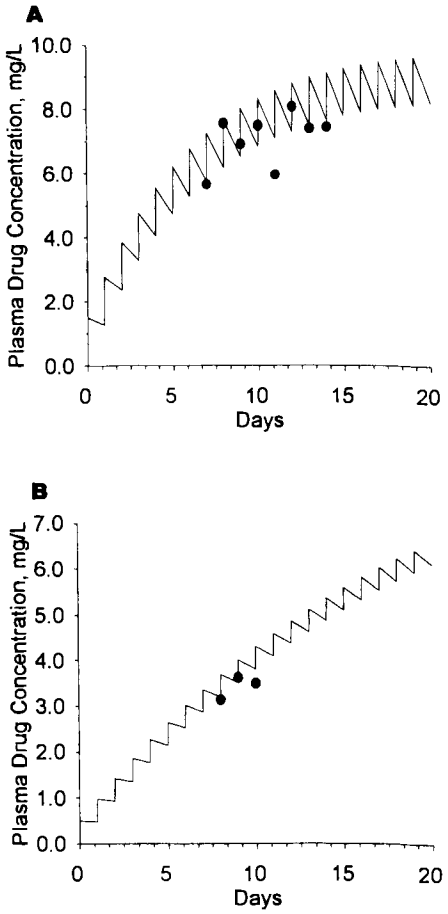

Fig. 2. How the evaluation of the approach to steady state depends on the time frame of observations with respect to drug kinetics. On a fixed-dose, fixed-interval multiple-dose regimen, the peak and trough plasma concentrations rise toward limiting steady-state values as shown by the jagged line. The concentration-time profiles are simulated for drugs with one-compartment characteristics as follows: A: ($k = 0.163$ days^{-1}, Dose/V = 1.5 mg/L); B: ($k = 0.0513$ days^{-1}, Dose/V = 0.5 mg/L). Observations are simulated by multiplying the mean by a factor with a 10% CV. A. Mean trough (sampled just before next dose) observations on Days 7–14 (circles) reveal a rise toward steady state during this period. The concentration of the last sample is theoretically about 90% of the steady state value (8.5 mg/L). B. Mean trough measurements on days 8–10 on this regimen show small differences among the values, but, steady state is not yet achieved.

analysis may be appropriate, but this needs to be verified. We do not know what the equivalence criterion should be to declare linearity. This is an area for further work. For now, study sponsors need to make their own judgment.

If the first step demonstrates sufficient linearity, the second is to compare the observed values on the final dosing day to 90% of the predicted steady state values. If average (or median) observed levels on the final sampling day are significantly greater than 90% of the average (or median) predicted steady state level, then one concludes attainment of steady state. This would be a one-sided paired-t or Wilcoxon signed rank test. Since one must satisfy both this and the linearity test, both can be at the 5% level. Note that the paired test approach takes into account the uncertainty of the prediction (both from the estimation of the model, if applicable, and the extrapolation to

multiple dosing) as well as the variability in the final day's levels. Most importantly, as discussed earlier, the alternative statistical hypothesis should be attainment of at least 90% of steady state and the null hypothesis that at least 90% of steady state is not yet attained.

Individual Evaluation

Statistically, it is possible to obtain a standard error for each subject's predicted steady state value that can be a basis for statistical testing of the hypothesis of attainment of steady state for that individual. Realistically, this is not likely to be a fruitful approach for two reasons. First, attainment of a large percentage of the steady state value is a high hurdle, even without consideration of within-subject variability. For example, even if each subject was at 90% of his/her average steady state value, because of assay error and within-subject variability we would only expect half of them to have observed concentrations above their 90% values. Second, estimated variability in AUC for an individual is based primarily on the assay errors in the measurements. Since assay variability is only a small part of true variability, the propagated standard errors for the AUC's are likely to be too small. Since this would lead to false assurance of attainment of steady state, we do not recommend this approach. To obtain an estimate of variance that includes sampling and time-to-time variability requires multiple single-dose components with adequate washout for each subject, each with its own estimate of steady state level. The number of single-dose components per subject would need be comparable to the typical number of subjects in such trials in order to have reasonable statistical power, making the full trial very large. Thus, while we might prefer methods based on assessment of steady state at the individual level, we are forced to consider aggregate measures for statistical control of error rates.

If the problems of evaluating steady state at the individual level are overcome, there is a question as to whether all or only some proportion (such as 50% or 75% or 90%) of subjects need to attain steady state. Inclusion of subjects not at steady state introduces bias in estimating steady-state pharmacokinetic parameters. Conversely, exclusion of subjects who are not at steady state introduces a bias as well by dropping a distinct subset of subjects from the analysis (i.e., those with lower clearances and/or larger volumes of distribution). There is clearly a tradeoff here. One could envision a variety of decision rules. For example, one might require that a high percentage, say at least 95%, of subjects be at steady state and that data for all subjects be used; otherwise the trial is judged not to be at steady state. The relative merits of different rules will vary depending on the pharmacokinetic scenario and the intent of the study. Clearly there is a benefit for designing the study so that all subjects are at steady state. Aggregate evaluation hides but does not eliminate these considerations. Aggregate approaches lead to inclusion of all subjects, but not all subjects need have attained steady state.

SUMMARY

Formal statistical approaches for assessing attainment of steady state are difficult to delineate due to inherent pharmacokinetic variability. A proper statistical assurance of attainment of steady state may require larger than typical sample sizes or may involve a comparison with so much variability that reasonable statistical assurance is not possible. Variability may be overcome with a larger sample size if the goal is aggregate assurance of steady state. However, for assurance of steady state for each individual, the approach we theoretically prefer, increasing the number of subjects does not help. Formal statistical approaches of individual evaluation lead to consideration of multiple single-dose components to the study in order to have multiple estimates of time to attainment of 90% of steady state for each subject and thus an estimate of the most appropriate variance.

Our best recommendation for a statistical approach is a combination of the aggregate and individual evaluations. First, evaluate the trial at the aggregate level as described above, including assessment of linearity. If the trial does not pass this assessment, conclude that steady state is not attained for the trial. This views aggregate attainment of steady state as a necessary, but not sufficient, condition. As seen earlier (Fig. 1), average and individual assessments need not agree. Our perspective is that a proper statistical approach for aggregate evaluation is possible and thus should be done in order to provide some control over statistical errors. What of the individual subjects? We further recommend that the trial be conducted in such a way that the time to steady state or the steady state level can be estimated for each subject in the trial. By simply comparing 90% of the estimates of steady state to what is obtained on the multiple dose test day, or comparing the time of the test day to the estimated times to attainment of steady state, the sponsor can make a judgment as to whether the results of the statistical aggregate assessment are overly optimistic.

ACKNOWLEDGMENTS

The National Heart, Lung and Blood Institute supported a portion of this work with a grant (#HL51401) to Jefferson Medical College.

REFERENCES

1. M. Rowland and T. N. Tozer. *Clinical Pharmacokinetics: Concepts and Applications*, 3rd ed., Williams and Wilkins, Media, PA, 1995, Ch. 6.
2. S. C. Chow and J. P. Liu. *Design and Analysis of Bioavailability and Bioequivalence Studies*, Marcel Dekker, New York, 1992, Section 12.4.
3. W. W. Hauck and S. Anderson. A new statistical procedure for testing equivalence in two-group comparative bioavailability trials. *J. Pharmacokin. Biopharm.* **12**:83–91 (1984).
4. D. J. Schuirmann. A comparison of the two one-sided test procedure and the power approach for assessing the bioequivalence of average bioavailability. *J. Pharmacokin. Biopharm.* **15**:657–80 (1987).
5. S. N. Goodman and J. A. Berlin. The use of predicted confidence intervals when planning experiments and the misuse of power when interpreting results. *Ann Intern. Med.* **121**:200–6 (1994).
6. R. A. Fisher. *The Design of Experiments*, Oliver and Boyd, London, 1935.
7. D. Hauschke, V. W. Steinijans, and E. Diletti. A distribution free procedure for the statistical analysis of bioequivalence studies *Int. J. Clin. Pharmacol. Ther. Toxical.* **28**:72–8 (1990).