Bioequivalence of Long Half-Life Drugs - Informative Sampling Determination -Parallel Designed Studies

Ahmed El-Tahtawy1, Ferrin Harrison2, Jeanne Fourie Zirkelbach2 and Andre J. Jackson2*

1Pfizer Laboratories, Division of Global Clinical Pharmacology, Pharmacometrics, New York, NY, USA (A.E-T)
2Office of Clinical Pharmacology, Office of Translational Science, Center for Drug Evaluation and Research, Food and Drug Administration, Silver Spring, MD, USA (F.H. (Deceased), J.F.Z., A.J.J.)

Abstract

Objective: To determine if 72 hours is the most informative sampling duration for the bioequivalence (BE) determination for drugs with half-lives > 30 h when using a parallel study design.

Methods: Two-treatment parallel-designed BE studies were simulated. A one-compartment oral absorption model with half-lives of 30 h and 350 h (clearance = 0.224 or 0.019 L/h), distribution volume = 9.7 L, and inter-subject variability for clearance of 75-250% was simulated. The test/reference ratio for fraction available was investigated at 1.0 and 1.25, while the rate constants for absorption (Ka) were simulated at a test/reference ratio of 1 and 4. AUC values truncated at 12-360 h were calculated. Experimental parallel BE studies drugs were also investigated.

Key findings: Experimental BE data indicated a decrease and then an increase in the root mean square error (RMSE) or variability as a function of time. Simulations supported these findings with the highest probability of passing the CI being between times 24 and 120 h depending on Ka, half-life, and inter-subject variability. Based on this work, a reduction in the sampling duration of parallel-designed BE studies is recommended. Experimental BE data indicated a decrease and then an increase in the RMSE. The 30-h half-life simulations exhibited a minimum in RMSE that rose to a plateau at 350 h. There was an increase in the probability of rejecting BE with longer sampling times for the 30-h simulations showing a maximum near 300 h while the 350-h half-life simulations showed no maximum.

Conclusion: For parallel-designed BE studies, sampling beyond 120 h will not change the BE decision and therefore is unnecessary.

Keywords: Truncated Area under the curve; Parallel designed studies; Bioequivalence

Abbreviations: AUC0-∞: Area under the curve, time zero to infinity; AUCext: extrapolated area; BE: bioequivalence; CI: Confidence Interval; Cplqo: last measured concentration above the limit-of quantitation; Cpt(t): generated drug concentration at time t; CV: coefficient of variation; CV: Median: calculated using the range of data based on 75% and 25% percentile and median; F: fraction absorbed or bioavailability; FDA Food and Drug Administration; (FT/FR): test/reference ratio for fraction available; Ka: rate constant for absorption; ke: elimination rate; TKA: Test Ka/Reference Ka; RATF: Test F/Reference F; RMSE: root mean square error

Introduction

The U.S. Food and Drug Administration (FDA) Guidance for Industry (U.S. Food and Drug Administration, 2003) entitled, “Bioavailability and Bioequivalence Studies for Orally Administered Drug Products — General Considerations” discusses long half-life drugs and appropriate sampling times. The recommendation is stated, “For drugs that demonstrate low intra-subject variability in distribution and clearance, an AUC truncated at 72 hours (AUC0-72 h) can be used in place of AUC0-1 or AUC0-∞ (AUC time zero to infinity). For drugs demonstrating high intra-subject variability in distribution and clearance, AUC truncation warrants caution. In such cases, we also recommend that sponsors and/or applicants consult the appropriate review staff.”

However, the recent literature contains several studies for drugs with long half-lives which have investigated the use of AUC0-72 h as a surrogate for AUC0-∞. Work presented by Midha and colleagues found that for various drugs limiting the duration of sample collection did not increase the variation of AUC ratios (i.e., AUCtruncated/AUCinf) [1]. Endrenyi and Tothfalusi found that with intra and inter-individual variation coefficient of variation (CV) for clearance (CL) up to 25%, the resulting variability of the estimated truncated AUC ratios was generally reduced, as the duration of sample collection in simulated trials was shortened from 4 down to 2 half-lives following drug administration [2]. Their study concluded that the assessment of bioequivalence (BE) for long half-life drugs has undiminished validity when the duration of sample collection is shortened to, at most, 2 half-lives following drug administration. A further investigation into the use of truncated AUC for two-way crossover design experimental data, with median AUCinf intra-subject CVs up to 34% and inter-subject CVs up to 45 %, concluded that there was discordance in BE conclusions based upon AUCinf versus those with AUCs truncated at less than 1*Tmax [3]. However, for longer sampling times (i.e., those greater than 1*Tmax) there was better agreement. In the same study using simulated data it was found that intra-subject CVs of truncated AUCs...
changed as a function of time, being higher for AUCs truncated after the first few hours, but then decreasing rapidly to reach a minimum when AUCs were truncated at 96 to 144 h. An investigation using simulated data with half-lives from 40-172 h and % CV for clearance of 20% concluded that the truncated approach for the estimation of the AUC for long half-life drugs in bioequivalence studies may be useful, but also increases the probability of accepting drugs as being bioequivalent when they are not [4].

The present study was undertaken to investigate more thoroughly the statistical distributions of the data associated with parallel-designed studies where inter-subject, rather than intra-subject variability is of concern. The aim was to further investigate the accuracy of the FDA recommended 72 h sampling time for long half-life drugs when a parallel-design is used and to determine if there is a true “most informative time” to truncate AUC values in such BE studies [4]. Our primary interest was to investigate the use in parallel-designed studies, since this design is often used to study drugs with very long half-lives and also in most clinical pharmacology studies of hepatic and renal insufficiency.

Methods

Experimental data

Experimental data from two BE trials submitted to the FDA were used to determine if there was a similar BE conclusion for all observed truncated areas, compared to time infinity and the possible influence of inter-subject variability as measured by root mean square error (RMSE). Studies were parallel-designed single-dose trials in healthy volunteers conducted under fasting conditions. Study details are presented in Table 1.

For each BE trial, truncated AUC values were calculated for each sampling interval from 12 h to the last time point at which plasma sampling was obtained for the test and reference.

The following truncated areas were calculated for all studies: AUC(0-144h), AUC(0-120h), AUC(0-72h), AUC(0-56h), AUC(0-48h), AUC(0-36h), AUC(0-24h), and AUC(0-12h). AUCinf was determined (e.g., study 1) by regressing the time points near the limit-of-quantitation (loq) to obtain ke (elimination rate constant) based upon the highest R-square value with ke being positive, and calculated from at least 3 data points. The extrapolated area (AUCext) from the last measured concentration above loq (i.e., Cplq) was calculated based upon AUCext=Cpqloq/ke. An analysis of variance was performed using the natural logarithm (ln) of the truncated areas. The ANOVA model included subject and treatment. The ratio of geometric mean truncated AUC and its 90% confidence intervals (CI) were calculated using the least square means and the standard error of the estimate obtained from the ANOVA. RMSE was used as the estimate of inter-subject variability.

Simulations: Simulations were done to investigate the relationship between known experimental variables and times for determining truncated AUC values. Stochastic simulations were designed as parallel single-dose studies in 120 subjects to match the design of the experimental data from study 1.

Baseline model and parameter distributions: The standard one-compartment model with first-order absorption and elimination was used to simulate concentration-time data using the following equation:

\[ C_p(t) = \frac{Dose \times Ka}{V} \times \left( e^{-CL/V \times t} - e^{-ke \times t} \right) \]

where \( C_p(t) \) is the generated drug concentration at time t, Dose is the administered dose, F is the bioavailability, V is volume of distribution, CL is clearance and Ka is the first-order rate constant for absorption. The number of subjects simulated was 60 for test and 60 for reference. The mean half-lives of elimination were 30 h and 350 h, which are equivalent to study 1 and ~3x that for study 2.

Simulations were done to determine the effect the test product Ka and product half-life has on BE, as estimated by truncated AUC values at AUC(0-360h), AUC(0-300h), AUC(0-240h), AUC(0-216h), AUC(0-192h), AUC(0-168h), AUC(0-144h), AUC(0-120h), AUC(0-72h), AUC(0-56h), AUC(0-48h), AUC(0-36h), AUC(0-24h), and AUC(0-12h) for drugs with 30-h or 350-h half-lives. When the drug half-life was increased 10-fold, the rate constants of absorption for reference and test formulations were decreased to preserve the 4:1 ratio. Both rate constants reflect the longer time needed for absorption, with the \( T_{max} \) value for the reference observed at 38 h and that for the test at 8 h. All simulation parameters were based upon a lognormal distribution of the parameters except for F (bioavailability) values which were generated from a uniform distribution centered at 0.780 and 0.975 with a range of 0.01 units. Simulation parameters are presented in Tables 2 and 3 for the 30-h and 350-h simulations. The selected simulation parameters resulted in RMSE values vs. time for truncated AUC values of below 30% for low error, versus above 30% for high error.

Random assay error (i.e., standard deviation), both proportional and additive, were added to each generated concentration as \( \delta = (0.2 \times C_p + 1) \) where \( C_p \) is the generated value to give the observed concentration value, Conc. The limit of quantitation for the simulations was set to 0.1 ng/ml which was the same as that for the experimental data.

### Study Design

<table>
<thead>
<tr>
<th>Study Design</th>
<th>Number of subjects</th>
<th>Terminal half-life</th>
<th>Tmax</th>
<th>Plasma sampling times</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment 1 (300 mg)</td>
<td>60</td>
<td>29.1 h (28.0)</td>
<td>18.0 h</td>
<td>To 144 h post dose</td>
</tr>
<tr>
<td>Treatment 2 (200 mg+100 mg)</td>
<td>60</td>
<td>28.3 h (25.2)</td>
<td>20 h</td>
<td>To 144 h post dose</td>
</tr>
</tbody>
</table>

### Study 2 Design

<table>
<thead>
<tr>
<th>Number of subjects</th>
<th>Terminal half-life</th>
<th>Tmax</th>
<th>Plasma sampling times</th>
</tr>
</thead>
<tbody>
<tr>
<td>42</td>
<td>113 h (24.8)</td>
<td>32 h</td>
<td>To 648 h post dose</td>
</tr>
</tbody>
</table>

### Table 1: Summary of study designs for the BE trials. Values are arithmetic mean (%CV).

<table>
<thead>
<tr>
<th>Simulated Parameters</th>
<th>T/R Ratio</th>
<th>Mean</th>
<th>Standard Deviation (%CV)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Tmax</td>
<td>High Error</td>
</tr>
<tr>
<td>Ka-reference</td>
<td>1</td>
<td>0.58 h-1</td>
<td>20 h</td>
</tr>
<tr>
<td>Ka-test</td>
<td>4</td>
<td>2.3 h-1</td>
<td>6 h</td>
</tr>
<tr>
<td>CL 30-h half-life</td>
<td>--</td>
<td>0.192 Lh</td>
<td></td>
</tr>
<tr>
<td>V</td>
<td>--</td>
<td>9.7 L</td>
<td>0.97(10%)</td>
</tr>
<tr>
<td>F1</td>
<td>1</td>
<td>0.78</td>
<td></td>
</tr>
<tr>
<td>Fr</td>
<td>1.25</td>
<td>0.975</td>
<td></td>
</tr>
</tbody>
</table>

### Table 2: Final parameters for 30-h half-life simulation with Ka test/Ka reference ratio equal to either 1 or 4.
The ANOVA for the respective parameter included only subject and treatment effects.

Sixty subjects for both test and reference treatments were simulated 1000 times. The number of times the 90% CI were outside of the 80-125% range and the number of times the CI contained the true test/reference ratio were recorded.

Statistical parameters for simulated concentrations

Using all simulated observed concentration data for representative simulations at elimination half-lives of 350 h and 30 h (low and high error with Test/Reference ratio=1 or 4) as a function of time, the mean, standard deviation (S.D.), coefficient of variation (CV) and CV-Median (which is similar to CV but using the range of data based on 75% and 25% percentile and median) were calculated.

Results

Experimental data

Table 4 presents the results from the experimental data demonstrating the relationship of the truncated AUC(0-t) 90% CI to the point estimate and RMSE values as a function of truncation time and T_max. For study 1, the 90% CI showed a minimum range value between the 24-h and the 36-h sampling times corresponding to the smallest values for the RMSE. For study 2 the 90% CI, RMSE and estimate values appear to reach a minimum value at sometime between the 72 to 120-h sampling times.

Simulated data

The RMSE values from the simulation scenarios are presented in Figure 1. It was noted that the simulated RMSE values are similar to the experimental data with the RMSE for the 30-h half-life drug showing a very sharp minimum at 24 hours followed by a very rapid increase over time. The magnitude of the rate of the increase depended on the magnitude of the study error. Increasing the Ka (Test/Reference ratio) to 4 had minimal effect at the 30-h half-life, but there was a large effect for the 350-h half-life simulations. Values appeared to approach a maximum at 300 h for the 30-h half-life drug with the magnitude of RMSE reflecting the study inter-subject variability. For the 350-h half-life drug simulations the minimum for RMSE was more of a “minimum region” from 84-96 h and the rate of increase from the minimum was much slower than for the 30-h simulations. Also there was no apparent maximum for RMSE within the 350 h sampling as seen for the shorter half-life simulations. The increase in the RATKA (Test/Reference ratio) had the largest impact on the low error simulation at the 350-h half-life by slowing the increase in RMSE from the “minimum region.”

Figure 2 upper panel shows the results at a half-life of 30 h. For the 30-h half-life simulations, with the RATF (Test F/Reference F) and RATKA (Test Ka/Reference Ka) equal to 1 or the latter increased to 4, high error simulations showed a probability of not concluding BE, which approached 60% for the half-life of 30 h. The 30-h half-life data also showed a minimum probability of not being declared BE at times near 48 h for the high error simulations at both RATKA values. For low error at the 30-h half-life, no minimum was seen due to the low overall probability of not being declared BE (i.e., less than 0.5%). The middle panel shows the same results for the 350-h half-life simulation, which exhibited a lower probability of not being declared BE, (i.e., near 20%) up to the 200-h time point with a RATKA equal to 1 for both low and high error showing the same probability of failing to conclude BE. The bottom panel shows the results when RATF is equal to 1.25 and the RATKA equals 1. The probability of not being determined as BE was
95% for the 30- and 350-h half-life simulations for low and high error, which preserves the Type I error level.

For the 350-h half-life drug, the minimum probability of not being BE was observed from 84-120 h for high error, which was not influenced by increasing the RATKA ratio from 1 to 4. For low error RATKA=1 at the 350-h half-life, the probability was less than 0.5 % after 60 h; therefore, no minimum was observed. The unexpected increase in probability of being declared BE at low error for the RATKA=1, matches the pattern seen in RMSE in Figure 1 at the 350-h half-life (i.e., the lower the RMSE the lower the probability of not being declared BE). When the FT/FR ratio was increased to 1.25, there was a 95% probability of not being declared BE for all the truncated areas measured.

Statistical analysis of simulated concentrations

Figures 3 and 4 show that the mean concentration data for the 350-h half-life high error simulation had larger tail values than those observed for the 30-h half-life low error simulations. However, the standard deviation for 350-h half-life data showed different patterns than did the 30-h half-life data. For the 30-h half-life high-error simulations, the standard deviation went up quickly then down and back up between 0 and 48 h, then back up, with a minimum at 86 h and then down slowly again. The low-error simulations (data not shown) followed a similar pattern. In contrast, for the simulations at the 350-h half-life, the standard deviation went up, then down, then back up but never going down again reaching a plateau near 86 h with the 30-h simulations decreasing and the 350-h simulations increasing (results for 350 h low error simulations exhibited the same pattern - data not shown). Nevertheless, for all cases, there appeared to be either a small minimum or plateau in the range of 72 to 86 h which seems to be the most informative times. To make a fair comparison for the deviations of the data among all time points, both CV and also CV-Median were used. As shown in all graphs, the CV and CV-Median showed an upward trend, starting from 24 h. In graphs for the 30-h half-life, the upward trends appear to be more prominent than those for the 350-h half-life. Between Figures 3 and 4, it was noted that using the CV-Median gives a stronger signal (e.g., increasing after 36 h for the 30-h half-life simulations and no change after 36 h for the 350-h simulations) than using CV. This is most apparent when the half-life, the standard deviation went up, then down, then back up but never going down again reaching a plateau between 70 and 86 h. In other words, the data appeared to show a similar pattern with a final inflection point near 86 h with the 30-h simulations decreasing and the 350-h simulations increasing (results for 350 h low error simulations exhibited the same pattern - data not shown). Nevertheless, for all cases, there appeared to be either a small minimum or plateau in the range of 72 to 86 h which seems to be the most informative times. To make a fair comparison for the deviations of the data among all time points, both CV and also CV-Median were used. As shown in all graphs, the CV and CV-Median showed an upward trend, starting from 24 h. In graphs for the 30-h half-life, the upward trends appear to be more prominent than those for the 350-h half-life. Between Figures 3 and 4, it was noted that using the CV-Median gives a stronger signal (e.g., increasing after 36 h for the 30-h half-life simulations and no change after 36 h for the 350-h simulations) than using CV. This is most apparent when the
difference between data generated with high errors and low errors are either not distinguishable except the longer plateaus in cases of high errors simulations.

The accuracy of the simulations based upon the number of times the true test/reference ratio was within the 90% CI was 90% or greater for the ratio at FT/FR=1.00 and at the FT/FR=1.25 (data not presented).

Discussion

A finding of the current work using a parallel-design study to determine BE was that the most informative time was impacted by half-life, error level and rate constant of absorption. Time ranges appeared to be more informative than a single time. For 30-h half-life simulations it was the times from 72 to 86 h, while for 350-h simulations it was from 86 to 96 h, depending on RATKA. Experimental data ranged from 36 to 120 h, differing from the 72 h recommended by FDA for two-treatment two-period crossover design studies.

Experimental data for both types of studies showed some distinctive differences. The data presented in Table 2 for two-treatment parallel BE studies show concordance on the BE decision related to truncated AUC, irrespective of length of sampling times with respect to Tmax for all studies. However, experimental data for a two-treatment two-period crossover design study show a discordance of 47% in the BE decision based upon truncated AUC at 1x Tmax vs. AUCinf [3] with the rationale for the difference being the higher intra-subject CV for AUC at 1x Tmax compared to later times. A primary question has been how long to sample after Tmax. Midha and colleagues suggested that use of the truncated AUC corresponding to twice the AUC, irrespective of length of sampling times with respect to Tmax for all studies. These authors based their conclusion on experimental absorption data for amitriptyline and nortriptyline.

Clinical data for the drugs in the current parallel studies showed a pre-Tmax variability of 17-31%, similar to that found for published crossover design studies [3]; however, the variability did not impact the CI, so there was no discordance in the values at any time periods, as observed in the crossover design studies. In the same study it was also reported that values for AUC truncated at 4x Tmax values performed the same as AUCinf for two-way crossover design studies. This was similar to what was found in the current parallel-design studies where there was close agreement for these parameters throughout sampling times. Nonetheless, this may be related to the power of the larger experimental cross-over study data base used (i.e., 123 studies) to distinguish such differences relative to that of the small parallel study data base in the current investigation (N=2). In addition, few parallel designed studies for BE sample beyond 72 h which limits the length of truncation that could be studied [4].

Although the experimental data base was small, the simulations supported the results found in the representative parallel study data base. For example, for both the 30-h and 350-h half-life simulations, the RMSE initially decreased to a distinct minimum (i.e., 30-h half-life simulations) or a broader trough, (i.e., 350-h half-life simulations) then either rapidly or slowly increased with time, Figure 1. This pattern was similar to that seen for the experimental data in Table 2. These observed minima and troughs were consistent with the changes in plasma S.D., CV, and CV-Median presented in Figure 3 and Figure 4 for the 30-h and 350-h simulations. The high-error simulations and statistical parameter minimums for the 350-h half-life simulations in Figure 1 and Figure 3 for high error at RATKA=4 were more representative of a trough which extended for a time with values as follows: (RMSE= 26.7%-48 hours; RMSE = 25.9%-60 hours; RMSE = 25.6%-72 hours; RMSE = 25.6%-84 hours and 26%-96 hours. Values for RMSE reflect the changes in the study statistics. The crossover design studies also reported an increase in the intra-subject variability following the observed trough, with highly-variable drugs exhibiting a later decrease. In contrast to the parallel-design, which measures inter-subject variability, the RMSE values decreased to a minimum and then increased but did not show a final decrease similar to that observed for intra-subject variability within the 350-h sampling period. The simulations for the 30-h half-life drug did show a maximum for inter-subject variability at 300 h post dose.

An investigation of the utility of truncated AUC for the determination of BE in two-sequence, two-treatment, crossover design studies has been investigated by several authors [2,3,5-7]. These studies have suggested that indeed there exists informative sampling times for the length of sampling for truncated AUC that can best predict the results for AUC at time infinity. Endrenyi and Tothfalusi were one of the first to suggest that truncated AUC be considered a viable surrogate and showed that this metric would give results equivalent to the established metrics of area-under-the curve to time t (i.e.,AUC(0-t)) and AUCinf [2]. Gaudreault and colleagues, whose comprehensive investigation was based upon experimental and simulated data for drugs with low and high variability, concluded from their results that it would be reasonable to limit the sample collection time to 72 h [3]. This rationale for truncated AUC was based upon gastrointestinal transit time that indicated that it would not be possible for absorption to extend beyond 48 to 72 h. Similar conclusions were reached by Sathe and colleagues based upon one-compartment and two-compartment models whenever kd/kel, Ka/kel, Ka/β and kd/β (i.e., the ratio of the dissolution and absorption rate constant to elimination) were large and absorption essentially complete (Sathe et al., 1999). In their studies they found that truncated AUC values from 72 to 168 h passed the 90% confidence intervals. However, these authors did not distinguish between low and high variability drugs.

The work by Mahmood pointed out that using truncated AUC for long half-life drugs in BE may be useful, however, it increases the probability of accepting drugs as BE when they are not [6]. Our results do not support those findings related to extent of absorption as seen in Figure 2. When Ft/FR=1.25, the probability of rejection was always 95%. In addition, if one samples to the truncated time or beyond there was no evidence found in the experimental or simulated data that the BE decision would change from BE to nonequivalent or vice versa. In addition, it was illustrated how the differences in power between different truncated AUC values vary depending on the model, and all seemed to maintain a 5% false positive rate [3].

To conclude, simulated data and experimental data indicated that there existed a most informative time range depending on inter-subject variability, half-life and Ka. However, the current practice of sampling to 72 h for parallel BE studies [4] is not in opposition to the current results. This is especially true for low-error, long half-life drugs, since a trough occurs in the RMSE, probability of rejection, CV, S.D., and CV-Median regions between 24 and 86 h. The current FDA guidance states [4], “For drugs demonstrating high intra-subject variability in distribution and clearance, AUC truncation warrants caution.” In such cases, we also recommend that sponsors and/or applicants consult the appropriate review staff.” This would also seem to be applicable to parallel-designed BE studies with high inter-subject variability. These studies may require sampling to 120 h.
Although the focus of this investigation was on BE, one should be aware that using informative sampling for parallel-designed studies has implications for clinical pharmacology. Of the many available study designs, a parallel-design is suggested for dedicated drug-drug interactions studies [8]. Often, the analysis uses CI to determine the apparent magnitude of the interaction [9]. The parallel-design is used for all intrinsic factor studies such as for renal and hepatic insufficiency. Therefore, it appears that the most informative data in these important clinical pharmacology studies should also be obtained using truncated AUC values dependent upon the drug’s half-life and level of inter-subject variability.

**Acknowledgement**

The authors would like to acknowledge the help of Yeh-Fong Chen and Larry Ouderkirk for their help in the preparation of this manuscript.

**References**