Serial Versus Sparse Sampling in Toxicokinetic Studies

Francis L. S. Tse[1,2] and Jerry R. Nedelman[1]

Received February 8, 1996; accepted March 30, 1996

Purpose. Sparse sampling in rodent toxicokinetics usually involves the collection of a single blood sample on a given study day from each animal in a treatment group. The samples are allocated to different time points, often allowing some replicates, and statistical inferences are then made about the concentration-time behavior of the test compound. The present study compared the results of one such analysis with those obtained from serial sampling as might be applied using satellite animals.

Methods. Ten rats each received a single oral dose of tritium-labeled compound X. Blood concentrations in each rat at 10 time points postdose were determined by liquid scintillation counting. Individual peak concentrations of blood radioactivity (C_{max}) and peak times (t_{max}) were recorded, and area-under-curve (AUC) values were calculated by trapezoidal rule. The mean AUC and C_{max} of all 10 animals over all 10 time points, referred to as AUC_{true} and $C_{max,true}$, were used as points of reference. These values were then estimated using subsets of the data that simulated satellite-animal or sparse-sampling designs. First, several different sampling schedules of 5 bleeding times were simulated by taking subsets of the full data set. For analysis using satellite animals, serial blood concentrations from subsets of 3 or 4 rats were used to calculate point and confidence-interval estimates of AUC_{true} and $C_{max,true}$ by standard methods; all possible subsets of 3 or 4 of the 10 rats were considered. For sparse data analysis, a single concentration from each of the 10 rats was used to calculate both point and confidence-interval estimates of AUC_{true} by the Bailer-Satterthwaite method, and point estimates of $C_{max,true}$, according to several different designs of replication. Animals were randomly assigned to time points, and 1000 of over 50000 possible combinations were evaluated for each bleeding schedule. The average percent absolute errors of the point estimates were computed and, for the 95% confidence intervals, average widths were determined.

Results. For point estimates of AUC_{true}, sparse sampling yielded average percent absolute errors of 7–13%. Percent errors for 3 and 4 satellite animals were 6–12% and 5–10%, respectively. For 95% confidence intervals, sparse sampling yielded widths of 24–90% of AUC_{true}, whereas for 3 and 4 satellite animals widths were 37–50% and 24–34%, respectively. For $C_{max,true}$ point estimates from sparse-sampling and satellite-animal approaches had average percent absolute errors of 5–12% and 3–8%, respectively. The confidence-interval widths for $C_{max,true}$ from the satellite-animal approach were 15–24% of $C_{max,true}$, but coverage did not achieve the nominal 95% for some choices of sampling times.

Conclusions. By using proper study designs, one can limit the number of samples and the amount of blood drawn so as not to affect the animals' health status, yet still achieve the customary pharmacokinetic objectives in a toxicity study.

KEY WORDS: sparse sampling; toxicokinetics; rodent; Bailer's method.

[1] Department of Drug Metabolism and Pharmacokinetics, Sandoz Research Institute, Sandoz Pharmaceuticals Corporation, East Hanover, New Jersey 07936.

[2] To whom all correspondence should be addressed.

INTRODUCTION

The importance of monitoring drug levels in animals in toxicity studies in order to ascertain adequate drug exposure from all doses tested is now widely acknowledged (1). However, serial blood sampling normally cannot be applied to the main study animals on a rodent toxicity study because the trauma associated with frequent venipuncture and blood loss could cause adverse changes in physiology, thus jeopardizing the integrity of the toxicity trial. As a consequence, two primary strategies have been adopted by different laboratories in an attempt to circumvent this problem. Some investigators routinely carry satellite animals, which are groups of animals included in the design and conduct of the toxicity study and housed with the main-study animals, but used primarily for toxicokinetic sampling. This method inevitably requires a greater number of animals, more drug substance, and increased labor for dosing and veterinary care. Therefore, recent efforts in rodent toxicokinetics have focused on new experimental designs and associated statistical procedures for making inferences about concentration-time behavior, without the need for satellite animals (2,3). One new approach, commonly known as sparse sampling, has been applied successfully in determining drug exposure in several toxicity studies (4–6). Typically, it involves the collection of a single blood sample on a given study day from each main-study animal in a treatment group. The total number of samples are allocated to different predesignated time points, often allowing some replicates, and an appropriate statistical procedure is then employed to assess the population mean exposure to the compound, with some indication of the variance of the data. The present study was conducted to compare the results of one such analysis with those obtained from the conventional, serial sampling method. Comparisons were made on the basis of absolute error of point estimates of area under the curve (AUC) and peak concentration (C_{max}) relative to their true values, as well as by confidence-interval width. The former measure is important in toxicokinetics because point estimates are typically used in determining animal-to-human exposure ratios, whereas the latter is a conventional estimate of underlying precision.

MATERIALS AND METHODS

Rat Study

The study adhered to the "Principles of Laboratory Animal Care" (NIH publication #85-23, revised 1985), and was approved by the Sandoz Animal Care and Use Committee. Ten naive male Sprague Dawley rats each weighing 200–300 g were used. They were housed in a room with controlled temperature (~22°C) and humidity (~50%), and were allowed free access to food and water. The test compound X, an acetylcholinesterase inhibitor, was labeled with tritium (11 μCi/mg) at a metabolically stable position, and the radiochemical purity was >95%. The dose of [^{3}H]X, 1 mg/kg, was prepared as a suspension in 1% carboxymethylcellulose/0.2% Tween 80 and administered by oral gavage. Blood (200 μl) was collected from the cut tail of each rat at 0.25, 0.5, 1, 2, 3, 4, 6, 8, 12, and 24 h postdose. The samples were air-dried and combusted in an oxidizer (Model 306, Packard), and the radioactivity was measured by

0724-8741/96/0700-1105$09.50/0

liquid scintillation counting (Packard Tri-Carb® Liquid Scintillation Spectrometer, Model 2500TR). Radioactivity concentrations in blood are given as ng equivalents of X per ml.

Individual peak concentrations of blood radioactivity (C_{max}) and peak times (t_{max}) were recorded, and AUC values were calculated by the linear trapezoidal rule. The mean AUC of all 10 animals over all 10 time points is referred to as AUC_{true}, and the mean C_{max} of all 10 animals is referred to as $C_{max,true}$, since they represent estimates of the true mean AUC and true mean C_{max}, respectively. Although these values are only estimates, they are useful as points of reference for assessing the performance of procedures using less data.

Choices of Sampling Schedules

Subsets of the full data set were drawn to simulate sampling schedules of five bleeding times. Four sampling schedules were chosen. One, the naive schedule, was a typical schedule that might be used in the absence of any prior information about the concentration-vs-time (Cxt) behavior of the compound. The times for the naive schedule were 0.5, 1, 4, 8, and 24 h postdose. The other three schedules were obtained by optimizing various criteria over all possible subsets of five times, using the data from all 10 animals. The three criteria were: A) minimum relative absolute error in AUC; B) minimum average of the relative absolute errors in AUC and C_{max}; and C) same as B but with no sampling time between 8 and 24 h postdose. The reason for a constraint such as that imposed for criterion C is the desire to avoid after-hours work for technicians.

Estimates of AUC_{true} and $C_{max,true}$ by Serial-Sampling (Satellite-Animal) Designs

Individual C_{max} and AUC were determined for each of the four different sampling schedules. Subsets of 3 or 4 rats were used to calculate the mean and confidence-interval estimates of AUC_{true} and $C_{max,true}$ by standard methods; all 120 or 210 possible subsets of 3 or 4 of the 10 rats were considered.

Estimates of AUC_{true} and $C_{max,true}$ by Sparse-Sampling Designs

Sparse-sampling designs were simulated by selecting one concentration value from each of the 10 rats according to three different replication designs. A replication design specifies how many rats are sampled at each time point. One replication design that was used was 2-2-2-2-2, i.e., 2 rats were sampled at each of the 5 time points. There are 113,400 ways to allocate 10 rats to such a design. Another replication design was 1-2-3-3-1, which was found to be useful in a previous study (4). There are 50,400 ways to allocate 10 rats to such a design. A third replication design was selected based on the patterns of variability among the concentrations of the 10 rats as previously described (4). For each of the three replication designs, 1,000 of the over 50,000 possible allocations of the 10 rats were randomly selected. From each such allocation, point and confidence-interval estimates of AUC_{true} were computed by the Bailer-Satterthwaite method (4), and point estimates of $C_{max,true}$ were computed as the maximum of the average concentrations over the five time points.

Comparisons of Results from Different Designs and Serial vs. Sparse Sampling

To assess the performance of each procedure, the average percent absolute errors (relative to the estimated mean values AUC_{true} and $C_{max,true}$) of the point estimates were computed. These measures estimate the actual imprecisions of the estimators. For the confidence-interval procedures the average width of the 95% confidence-interval estimate was determined as $t \times$ (s.e.), where t is the appropriate critical value from a t distribution and s.e. is the appropriate standard error. Note that this width is the distance from the center to one endpoint of the interval, which is half the total width of the interval. The confidence interval width reflects the observed imprecision of the inference.

RESULTS

Rat Study

The blood concentrations and pharmacokinetic parameters of [^{3}H]X are summarized in Table 1. Accordingly, AUC_{true} was estimated to be 1720 ng equiv.·h/ml and $C_{max,true}$ was estimated to be 240 ng equiv./ml.

Choices of Sampling Schedules

The four choices of sampling times, and their performance when applied to all 10 rats, are summarized in Table 2.

Choices of Replication Designs

It was shown previously (4) how more precise inferences, meaning narrower confidence intervals, for AUC_{true} can be obtained using the Bailer-Satterthwaite method if replicates are reallocated from times where $w_i^2\sigma_i^2$ are small to where values of that quantity are large, where w_i is the trapezoidal weight associated with the i'th sampling time and σ_i is the standard deviation of concentrations at that time point. Table 3 displays the values of these quantities, normalized to sum to one, with σ_i estimated by the sample standard deviations reported in Table 1.

From the patterns in Table 3, it appears that reallocating animals from the first two times to the last two times of each schedule might improve the replication design. Thus, a 1-1-2-3-3 design was used in addition to the 2-2-2-2-2 and 1-2-3-3-1 designs. The latter design was selected in a similar analysis of a different drug (4), but the patterns in Table 3 suggest that this replication design might not be appropriate for Drug X in the present study.

Comparisons of Results from Different Designs and Serial vs. Sparse Sampling

Table 4 summarizes the average relative absolute errors and confidence-interval widths for different sampling schedules with both the satellite-animal and the sparse-sampling approaches.

The average relative absolute errors for both AUC and C_{max} in Table 4 show that using 4 satellite animals instead of 3 provides only a small improvement in actual precision, and using sparse sampling from the main study animals results in

Table 1. Blood Concentrations and Pharmacokinetic Parameters of [^{3}H]X Following a Single Oral Dose (1 mg/kg) in the Rat

Rat no.	1	2	3	4	5	6	7	8	9	10	Mean ± SD
Body weight (g)	284	296	284	294	289	285	290	292	289	298	290 ± 4.79
Concentration (ng equiv./ml) at:											
0.25 h	164	145	156	170	209	257	149	195	78.5	209	173 ± 48.1
0.5 h	218	245	223	228	236	249	225	268	195	279	237 ± 24.6
1 h	196	254	208	216	189	215	208	253	207	257	220 ± 24.9
2 h	109	154	125	158	137	142	130	156	145	177	143 ± 19.3
3 h	65.4	94.0	71.5	98.0	79.7	86.6	78.9	105	84.0	86.9	85.0 ± 11.9
4 h	55.6	60.1	66.3	88.6	68.2	68.4	59.9	81.1	69.2	96.9	71.4 ± 13.3
6 h	39.8	50.7	54.0	79.7	65.9	55.8	48.8	73.3	57.8	55.9	58.2 ± 11.8
8 h	77.6	99.0	83.4	83.8	94.1	71.9	47.9	59.0	38.3	45.0	70.0 ± 21.3
12 h	77.1	50.5	22.8	55.9	59.6	40.8	82.7	33.4	82.7	44.4	55.0 ± 20.8
24 h	38.7	70.6	24.2	122	50.7	33.8	49.0	22.3	33.1	89.0	53.3 ± 31.7
t_{max} (h)	0.5	1	0.5	0.5	0.5	0.25	0.5	0.5	1	0.5	0.6 ± 0.2
C_{max} (ng equiv./ml)	218	254	223	228	236	257	225	268	207	279	240 ± 23.6
AUC (ng equiv.·h/ml)	1690	1880	1260	2270	1800	1510	1770	1450	1670	1900	1720 ± 279

Table 2. Sampling Schedules

Schedule	Time (h)	Average Relative Error for AUC (%)	Average Relative Error for C_{max} (%)
Naive	0.5, 1, 4, 8, 24	16.2	0.3
A: Best for AUC	1, 4, 8, 12, 24	1.6	7.9
B: Best for AUC and C_{max}	0.5, 3, 8, 12, 24	2.9	1.2
C: B with Time Constraint	0.5, 1, 3, 6, 24	10.2	0.3

acceptably small losses of actual precision relative to using satellite animals. Comparing across sampling schedule reveals that schedules selected for their optimality do indeed yield greater actual precision for AUC. Even schedule C, optimized under the constraint that no samples be taken between 8 and 24 hours post dose, had better precision than the naive schedule for AUC. For C_{max}, the three schedules containing t = 0.5 h did equally well.

The average confidence interval widths in Table 4 show that both sampling schedule and animal allocation can markedly

Table 3. Bailer-Satterthwaite Terms

Naive Schedule					
Time (h)	0.5	1	4	8	24
Normalized $w_i^2 s_i^2$	0.00	0.02	0.02	0.4	0.57
Schedule A					
Time (h)	1	4	8	12	24
Normalized $w_i^2 s_i^2$	0.03	0.03	0.10	0.36	0.48
Schedule B					
Time (h)	0.5	3	8	12	24
Normalized $w_i^2 s_i^2$	0.02	0.03	0.12	0.36	0.48
Schedule C					
Time (h)	0.5	1	3	6	24
Normalized $w_i^2 s_i^2$	0.00	0.01	0.01	0.16	0.82

Table 4. Performance of Estimates of AUC and C_{max}

Average Relative Absolute Error for AUC (%)

	Naive	A	B	C
3 Satellites	11.5	6.4	6.4	9.2
4 Satellites	10.5	5.1	5.2	7.4
2-2-2-2-2	12.1	8.3	8.4	9.2
1-1-2-3-3	10.7	6.7	6.7	6.6
1-2-3-3-1	12.8	9.5	9.4	12.8

Average Relative Absolute Error for C_{max} (%)

	Naive	A	B	C
3 Satellites	3.8	8.2	4.2	3.8
4 Satellites	3.1	8.1	3.4	3.1
2-2-2-2-2	4.7	8.9	5.5	4.7
1-1-2-3-3	7.5	12.0	7.9	6.4
1-2-3-3-1	6.4	12.0	7.9	6.4

Average AUC Confidence Interval Width (% of AUC_{true})

	Naive	A	B	C
3 Satellites	51	37	36	49
4 Satellites	34	24	24	33
2-2-2-2-2	82	53	53	90
1-1-2-3-3	33	25	25	34
1-2-3-3-1	27[a]	20[a]	20[a]	15[a]

Average C_{max} Confidence Interval Width (% of $C_{max,true}$)

	Naive	A	B	C
3 Satellites	22	24[b]	23	22
4 Satellites	15	16[b]	16	15

[a] Coverages were only 60%–83%.
[b] Coverages were only 71%–83%.

affect the observed precision of the inferences. The considerable reduction in width when going from 3 to 4 satellite animals reflects the sensitivity of the t-distribution to changes in degrees of freedom when degrees of freedom are small. The critical values used for the confidence intervals are 4.3 and 3.2 for 2 and 3 degrees of freedom (n = 3 and 4), respectively. For the sparse-sampling 2-2-2-2-2 design, the large confidence-interval widths were due to few degrees of freedom and large standard errors arising from the replication design. By taking advantage of the pattern of variability in the data, however, and changing the replication design to 1-1-2-3-3, a precision comparable to that obtained with 4 satellite animals was produced. On the other hand, incorrectly understanding the behavior of the data so that an inappropriate replication design is used, as 1-2-3-3-1 was here, can lead to misleading results. The confidence intervals for the 1-2-3-3-1 design were narrow, but the intervals covered the true value less than their nominal 95% of the time.

Confidence intervals for $C_{max,true}$ obtained with the satellite-animal approaches were honest and comfortably narrow when the sampling schedule included the most common t_{max} value of 0.5 h. When the earliest sampling time was 1 h, the intervals covered the true value less than the nominal 95% of the time. No method is currently available for computing confidence intervals for $C_{max,true}$ from the sparse-sampling approach.

DISCUSSION

In a previous study (4), computer simulations with completely randomly generated data were used to assess the performance of the Bailer-Satterthwaite method applied to sparse sampling. One objection that can be levelled against such simulations is that they make assumptions about the behavior of toxicokinetic data that may not be valid in practice. The purpose of the present study was to generate real data with which to assess sparse-sampling methods. Since the concentration-time profile and pattern of variability are drug and assay-method specific, the results reported herein may not be readily extrapolated to other data sets. Nonetheless, it is only by such analysis of real data that the validity of sparse-sampling methods can be demonstrated.

In order to facilitate the collection of 10 serial samples that contain measurable concentrations of the analyte within 24 h, it was decided to measure total radioactivity instead of the parent compound. The resulting data had two features not commonly observed when measuring parent drugs. Firstly, the radioactivity exhibited a relatively long terminal phase, with secondary peaks. Secondly, the standard deviations of the concentrations were approximately constant over time rather than varying proportionally with concentrations. One consequence of this behavior of the data was that the best reallocation of animals to reduce the width of the Bailer-Satterthwaite confidence intervals was different from what was found for the compound described in the previous study (4). The reallocation used in that study (4), applied here to compound X, was decidedly worse even than the naive, uniform allocation in that it led to dishonest confidence intervals, i.e., intervals that did not attain their nominal coverage.

Finding a better replication design for the sparse-sampling approach yielded confidence-interval widths that were uniformly narrower than and as narrow as those obtained from 3 and 4 satellite animals, respectively. The actual precision, as estimated by the average relative absolute error, obtained from the sparse-sampling approach was never unacceptably worse than that obtained from the satellite-animal approach. The better replication design even produced comparable actual precision for AUC.

CONCLUSIONS

Sparse sampling designs can provide statistically adequate information to achieve the customary pharmacokinetic objectives in a toxicity study. They avoid the use of satellite animals with the attendant costs of animal care and drug substance, and yet with only one sample per main-study animal per day they do not entail unacceptable risk to the main toxicity objectives of the study. To derive even greater precision from the sparse designs, prior information about the pharmacokinetic behavior of the test compound is useful for the determination of the sampling schedule and the replication design.

ACKNOWLEDGMENTS

The authors wish to thank Ms. Deborah Labbadia for technical assistance.

REFERENCES

1. L. F. Chasseaud. The importance of pharmacokinetic/toxicokinetic and metabolic information in carcinogenicity study design. *Drug Inf. J.* **26**:445–455 (1992).
2. J. van Bree, J. Nedelman, J.-L. Steimer, F. Tse, W. Robinson, and W. Niederberger. Application of sparse sampling approaches in rodent toxicokinetics: a prospective view. *Drug Inf. J.* **28**:263–279 (1994).
3. J. R. Nedelman, E. Gibiansky, F. L. S. Tse, and C. Babiuk. Assessing drug exposure in rodent toxicity studies without satellite animals. *J. Pharmacokinet. Biopharm.* **21**:323–334 (1993).
4. J. R. Nedelman, E. Gibiansky, and D. T. W. Lau. Applying Bailer's method for AUC confidence intervals to sparse sampling. *Pharm. Res.* **12**:124–128 (1995).
5. S. Pai, S. Fettner, G. Hajian, M. Cayen, and V. Batra. Characterization of AUCs from sparsely sampled populations in toxicology studies. *Pharm. Res.* **12**:S354 (1995).
6. F. Tse, J. Nedelman, D. Lapadula, and K. MacKenzie. The use of sparse sampling for toxicokinetic analyses in nonclinical studies. American College of Toxicology, Vienna, Virginia, November 12–15, 1995.