Pharmaceutical Research, Vol. 13, No. 9, 1996

Commentary

Performance of Bailer's Method for AUC Confidence Intervals from Sparse Non-Normally Distributed Drug Concentrations in Toxicokinetic Studies

Sudhakar M. Pai,[1,4] Jerry R. Nedelman,[2] Gerald Hajian,[3] Ekaterina Gibiansky,[2] and Vijay K. Batra[1]

Received April 12, 1996; accepted June 21, 1996

In a paper published in 1988, Bailer (1) proposed a method to determine a confidence interval for the difference between two AUCs (e.g., from two doses) in rodents when blood samples are obtained by destructive sampling. Recently, Nedelman et al (2) extended the method to estimate a confidence interval for a single AUC (e.g., at an individual dose level) in sparse sampling situations. This was done by replacing a certain z statistic with a t statistic, and estimating the degrees of freedom by Satterthwaite's approximation ("Bailer-Satterthwaite method"). While evaluating the Bailer-Satterthwaite method we discovered that indiscriminately applying the method to sparse sampling can yield confidence intervals (CIs) that are so wide to be of no practical utility. This is because the degrees of freedom (df) play an integral part in CI estimation, and attempts must be made to maximize the df in appropriate regions of the concentration-time (Cp-t) profile. We were also concerned about the use by Nedelman et al of Cp-t simulated assuming normal theory. These closely associated factors were delineated by means of a simulation experiment as follows.

Non-normally distributed Cp-t were simulated, using a one-compartment model, based on parameters randomly selected from a generalized lambda distribution (3). Using arbitrarily chosen mean values and coefficients of variation (CV) of 16% and 24%, generalized lambda distributions were generated for the apparent volume of distribution (V_d), and the absorption and elimination rate constants (k_a and k_e, respectively). We assumed that V_d was independent of k_a and k_e, but k_a and k_e were weakly correlated ($r^2 = 0.5$). Two types of experiments were simulated. In the first, Cp-t as those obtained by destructive sampling were simulated in 2 or 4 rodents at each of 5 time-points (0, 1, 4, 8, and 24 hr). In each of the 10 or 20 rodents, a Cp-t was simulated using randomly selected values of V_d, k_a, and k_e. Then, a normally distributed measurement error with 10% CV was added to each concentration. In the second experiment, Cp-t as those obtained by serial sampling were simulated in 2 or 4 rodents at the 5 time-points, by selecting values of V_d, k_a, and k_e, and adding a 10% normally distributed measurement noise. In the first experiment, the mean AUC was estimated by a point estimate and Bailer-Satterthwaite CIs as described by Nedelman et al. In the second, a point estimate and CI for the mean AUC were computed from the 2 or 4 individual AUCs by standard methods (4). In both experiments, AUCs were computed by the linear trapezoidal rule (5). The experiments were simulated 1000 times with V_d, k_a and k_e having 16% CVs, and 1000 times with 24% CVs.

The distribution of the pharmacokinetic parameters from generalized lambda distribution runs are shown in Table 1, and the AUC estimates and the CI widths are summarized in Table 2. The coverage rate is the percent of runs where the true AUC is covered by the confidence interval, and the width is the width (lowest to highest) of the 95% confidence interval. The mean AUCs for the 16% and 24% CV runs were virtually identical whether the concentrations were from serial sampling or from destructive sampling; also the CVs of the point estimates were reasonably small in magnitude for both sets of simulations,

Table 1. Profiles of Pharmacokinetic Parameters Generated Using the Generalized Lambda Distribution

Parameter	Mean[a]	CV%	Skewness	Kurtosis	Lowest	Highest
V_d	50	16	0.0143	−0.8386	31.4	68.1
k_a	0.05	16	0.0280	−0.7413	0.032	0.068
k_e	0.005	16	−0.1192	−0.4545	0.0028	0.0072
V_d	50	24	0.0143	−0.8386	22.2	77.2
k_a	0.05	24	0.0280	−0.7413	0.023	0.077
k_e	0.005	24	−0.1192	−0.4545	0.0017	0.0082

[a] Values are in arbitrary units.

[1] Department of Drug Metabolism and Pharmacokinetics, Schering-Plough Research Institute, Kenilworth, New Jersey.
[2] Department of Drug Metabolism and Pharmacokinetics, Sandoz Pharmaceuticals, East Hanover, New Jersey.
[3] Department of Biostatistics, Schering-Plough Research Institute, Kenilworth, New Jersey.
[4] To whom correspondence should be addressed.

0724-8741/96/0900-1280$09.50/0

Table 2. Point Estimates of Mean AUCs, and Estimation of Their Precision (Confidence Intervals) from Serial and Destructive Sampling

Parameter CV (%)[a]	n[b]	Sample	df[c]	AUC[d]		CI Width		Coverage
				Mean	CV(%)	Mean	CV(%)	
16	2	Destructive	1.8	11,409	11.5	14,916	72.2	97.9
		Serial	1	11,398	17.2	37,802	77.4	94.3
	4	Destructive	6.4	11,458	8.6	4,516	33.7	94.8
		Serial	3	11,379	12.2	8,087	43.8	95.1
24	2	Destructive	1.7	11,754	17.2	23,529	83.7	97.3
		Serial	1	11,749	26.2	57,616	81.5	93.8
	4	Destructive	6.2	11,848	13.0	7,099	44.1	95.0
		Serial	3	11,721	18.5	12,471	48.4	95.0

[a] CV added to V_d, k_a, and k_e; a normally distributed measurement error with 10% CV was added to each concentration.
[b] Number of animals per time-point.
[c] df for destructive sampling was computed by the Satterthwaite approximation.
[d] Values are for 1000 simulations, and are in arbitrary units.

Table 3. Probability of Over/Under Estimation of s from Normal-Distribution Theory

Probability	df					
	1	2	4	8	16	32
$P(s < \sigma/2)$	0.38	0.22	0.090	0.019	1.1×10^{-3}	4.9×10^{-6}
$P(s > \sigma/2)$	0.046	0.018	3.0×10^{-3}	9.3×10^{-5}	1.1×10^{-7}	2.0×10^{-13}

demonstrating that the AUCs are estimated with acceptable precision. The CIs showed large widths but did exhibit approximate 95% coverage.When Cp-t were generated under realistic assumptions about random variation, the confidence intervals retained their validity (coverage) even under sparse sampling situations. However, the intervals from sparse sampling were so wide as to be practically useless. Interestingly, whereas the average CI widths approximately tripled when the number of animals per time-point decreased from 4 to 2, the means of the point estimates were essentially identical, and the CVs of the AUC point estimates increased only slightly. Therefore, whereas the CIs from sparse sampling had little utility, the AUCs had acceptable precision.

Why is this so? This is because the standard errors (SE) are underestimated in sparse sampling situations which leads to wide CIs due to large $t_{0.975}$. This is explained as follows. Let σ be the true SE of the AUC based on infinite degrees of freedom, and let s be its estimate either based on the Bailer method for destructive sampling or computed using standard methods with serial sampling. If σ were known, then a 95% CI of the mean AUC could be computed using $t_{0.975} = 1.96$ (infinite df):

$$\text{AUC} \pm 1.96\sigma$$

However, if σ is not known, and there are, say, only 2 df, $t_{0.975}$ becomes 4.3, and the 95% CI will be

$$\text{AUC} \pm 4.3s$$

Thus, the critical value has approximately doubled compared to when the df was infinite. Why must it be this way? This is because s too often underestimates σ (due to lower df), and the critical value needs to be large to compensate. When normal distribution theory is valid, probabilities involving s can be computed exactly. Table 3 shows probabilities that s is less than half of σ or more than twice σ for various degrees of freedom. Clearly the tendency for s to underestimate σ is greater than the tendency to overestimate. Now, confidence intervals are designed to achieve a specified level of coverage. In order to attain the appropriate level of coverage, a larger critical value must be used to compensate for when s is underestimated.

Thus, with sparse sampling, the AUC may be determined with acceptable precision, but its confidence interval may be misleading; the CIs tend to underestimate precision. To avoid this situation sparse sampling designs must be implemented to increase the df in appropriate regions of the Cp-t profile. This was one of the intended messages in the original paper of Nedelman et al (2). In that paper the 1-2-3-3-1 design gave narrower intervals for the compound considered there. Thus, to implement the Bailer-Satterthwaite method, prior information about Cp-t profiles is required. From such information, sampling designs can be selected to assure that both AUC point estimates and its CI are reliably determined.

ACKNOWLEDGMENTS

The authors thank Dr. Mitchell N. Cayen, Senior Director, Department of Drug Metabolism and Pharmacokinetics, Scher-

ing-Plough Research Institute, for many helpful discussions during the preparation of this commentary.

REFERENCES

1. J. A. Bailer. Testing for the equality of area under the curves when using destructive measurement techniques. *J. Pharmacokinetic and Biopharm.* **16**:303–309 (1988).
2. 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).
3. J. S. Ramberg, E. J. Dudewicz, P. R. Tadikamalla, and E. F. Mykytka. A probability distribution and its uses in fitting data. *Technometrics.* **21**:201–214 (1979).
4. L. Ott. An Introduction to Statistical Methods and Data Analysis. 3rd edition, PWS-Kent, Boston, MA, 1988, pp. 127–129.
5. M. Gibaldi and D. Perrier. Pharmacokinetics, 2nd edition, Marcel Dekker, New York, NY, 1982, pp 445–449.