Coverage and Precision of Confidence Intervals for Area Under the Curve Using Parametric and Non-parametric Methods in a Toxicokinetic Experimental Design

Peter L. Bonate[1,2]

Received September 18, 1997; accepted December 15, 1997

Purpose. The coverage and precision of parametric Bailer-type confidence intervals (CIs) for area under the curve (AUC) was compared to nonparametric bootstrap confidence intervals.

Methods. Concentration-time data was simulated using Monte Carlo simulation under a toxicokinetic paradigm with sparse (SSC) and dense sampling (DSC) conditions. AUC was calculated using the trapezoidal rule and 95% CIs were computed using various parametric and nonparametric methods.

Results. Under SSC, the various parametric CIs contained the true population AUC with coverage probabilities ranging from 0.77 to 0.95 with low inter-subject variation (coefficient of variation (CV) = 15%) and from 0.82 to 0.95 with high inter-subject variation (CV = 50%). The nomimal value should be close to 0.95. DSC tended to increase coverage by about 0.05. Bailer's method always produced the lowest coverage of all parametric CIs examined. Under SSC, bootstrap CIs had coverage probabilities ranging from 0.62 (CV = 15%) to 0.68 (CV = 50%). DSC increased coverage to 0.77. Parametric CIs were wider than their nonparametric counterparts, often giving lower CI estimates less than zero. Bailer's method and Bailer's method using the jackknife estimate of the standard error were the worst in this respect. Bootstrap CIs never had lower CI estimates less than zero. However, SSC tends to produce bootstrap distributions that are not continuous which, if used, may produce biased CI estimates.

Conclusions. Bootstrap CI estimates were judged to be the "best". However, the limitations of the bootstrap should be clearly recognized and it should not be used indiscriminately. Examination of the bootstrap distribution for its degree of discrete-ness must be part of the statistical process.

KEY WORDS: bailer method; jackknife; bootstrap; destructive sampling.

INTRODUCTION

The role of toxicokinetics in drug development is becoming increasingly important. In a typical toxicokinetic experiment, individual animals are sampled at discrete time points and the concentration of drug in blood and or tissues determined. The number of time points and animals used is dependent on the required degree of precision in estimation of the pharmacokinetic parameters of interest. Of primary importance is characterization of the area under the plasma concentration-time curve (AUC) because this estimate may then be used to predict exposure in humans or correlate drug exposure to some toxicologic finding.

In addition to finding a point estimate for AUC, called $\widehat{AUC}$, it is often desirable to have some assessment of the precision of $\widehat{AUC}$. To this end, the variance of $\widehat{AUC}$, called Var ($\widehat{AUC}$), is calculated and a (1-α)% confidence interval reported. When the researcher assumes that $\widehat{AUC}$ has some defined statistical probability distribution, usually the normal distribution, a parametric $(1 - \alpha)$% confidence interval of the type $\widehat{AUC} \pm c \cdot \sqrt{Var\ (\widehat{AUC})}$ may be formed, where c is some constant defined to give coverage with probability 1-α. This type of method for confidence interval generation is sometimes called the pivot method. Alternatively, non-parametric confidence intervals, which make no assumptions regarding the underlying distribution of the point estimate, may be generated using computer-intensive techniques.

Recently, so-called "sparse sampling" designs, in which few animals are used at a minimum number of time points, have gained in popularity in the literature. These studies have shown that using appropriate sampling times, sparse sampling gives as accurate an estimate of $\widehat{AUC}$ as does their "intensive-sampling" counterparts. The purpose of this paper was to compare the myriad of methods used to generate (1-α)% confidence intervals for $\widehat{AUC}$ under the auspices of a "sparse sampling" toxicokinetic experimental design in an attempt to find those methods which provide both good coverage and precise intervals.

THEORY

Consider a typical toxicokinetic experiment where destructive samples are made at each time point of interest. Given that r_j replicates ($j > 1$) are made at each of t_i time points, $i = 1, 2, \ldots, m$, Bailer (1) showed that a point estimate for AUC, $\widehat{AUC}$, from time t_1 to t_m can be approximated by

$$\widehat{AUC} = \sum_{i=1}^{m} w_i \cdot \bar{C}_i \qquad (1)$$

where

$$w_i = \begin{cases} \frac{1}{2}(t_2 - t_1) & \text{for } i = 1 \\ \frac{1}{2}(t_{i+1} - t_{i-1}) & \text{for } i = 2, \ldots m - 1 \\ \frac{1}{2}(t_m - t_{m-1}) & \text{for } i = m \end{cases}$$

and $\bar{C}_i$ is the ith mean concentration. A generic equation for the variance of the sum of a linear combination of independent random variables, $X_1, X_2, \ldots, X_m$, is

$$Var(aX_1 + bX_2 + \ldots zX_m) = a^2Var(X_1) + b^2Var(X_2) + \ldots z^2Var(X_m).$$

Similarly, the variance of $\widehat{AUC}$ can be written as

$$Var(\widehat{AUC}) = \sum_{i=1}^{m} w_i^2 \left[\frac{\sigma_i^2}{r_j}\right],$$

which may be estimated using the maximum likelihood estimate for the sample variance

$$Var(\widehat{AUC}) = \sum_{i=1}^{m} w_i^2 \left[\frac{S_i^2}{r_j}\right], \qquad (2)$$

[1] Hoechst Marion Roussel, North American Clinical Pharmacokinetics, P.O. Box 9627 (F4-M3112), Kansas City, MO 64134.

[2] To whom correspondence, should be addressed. (e-mail: Peter.bonate@hmrag.com)

0724-8741/98/0300-0405$15.00/0

where

$$S_i^2 = \frac{\sum_{j=1}^{r_i} [C_{ij} - \bar{C}_i]^2}{r_j - 1,}$$

Bailer proposed that a $(1 - \alpha)\%$ confidence interval for $\widehat{AUC}$ can be constructed as

$$\widehat{AUC} \pm z_{crit} \sqrt{Var\,(\widehat{AUC})} \quad (3)$$

where z_{crit} is the critical value from a standard normal distribution. Eq. (3) has come to be known as the Bailer's method for confidence interval construction.

Nedelman (2) argued that the assumption that $\widehat{AUC}$ has a standard normal distribution is presumptive because it assumes that the variance of the $\widehat{AUC}$ estimate is known, when in fact it is not. They therefore suggested that t_{crit} be used instead of z_{crit} in eq. (3), where t_{crit} is the critical value associated with a 2-tailed students t-distribution. Nedelman et al. (2) argued that "substituting sample variances for population variances is safe when sample sizes are large enough, for then t_{crit} approximates z_{crit}." The problem with using t_{crit} instead of z_{crit} is that the degrees of freedom, v, must be estimated using Satterwaite's approximation:

$$v = \frac{\left[\sum_{i=1}^{m} \frac{w_i^2 \cdot s_i^2}{r_j}\right]^6}{\sum_{i=1}^{m} \left[\frac{w_i^4 \cdot s_i^4}{r_j^2(r_j - 1)}\right]} \quad (4)$$

This modification of the Bailer method has come to be known as the Bailer-Satterwaite method.

Pai et al. (3) has indicated "that indiscriminately applying the Bailer-Satterwaite method to sparse sampling can yield confidence intervals that are so wide to be of no practical utility." This is because Satterwaite's correction can lead to degrees of freedom that are near unity. Assuming $\alpha = 0.05$, when Satterwaite's approximation leads to degrees of freedom estimates of three, two, or one the resulting t_{crit} and confidence interval range is 1.6, 2.2, and 6.5 times larger, respectively, than the corresponding z_{crit} value. In the report by Nedelman et al. (2) demonstrating the utility of the Bailer-Satterwaite method, four of nine confidence intervals presented in Table III had a lower confidence limit less than zero, which is physically impossible. Similarly, six of nine estimates of the degrees of freedom using Satterwaite's approximation were less than three. It may be inferred that when the degrees of freedom is small, the corresponding confidence interval estimate using the Bailer-Satterwaite method will be much larger than the Bailer-method.

An alternative method of confidence interval construction can be made using Chebychev's inequality (4) which states that if X is a random variable with mean μ and variance σ^2 then the probability that X should deviate from its mean by more than k times its standard deviation is less than or equal to $1/k^2$, or

$$p\left(\frac{|X - \mu|}{\sigma} > k\right) \leq \frac{1}{k^2}. \quad (5)$$

If X is unimodal and monotonically decreasing on both sides of the mode, then Chebychev's inequality can be modified to state the probability that X should deviate from its mean by more than k times its standard deviation is less than or equal to $1/2.25k^2$ (5) or

$$p\left(\frac{|X - \mu|}{\sigma} > k\right) \leq \frac{1}{2.25 \cdot k^2}. \quad (6)$$

Eq. (6) can be used to find the probability that a value X is k standard deviations from its mean. For example, suppose that $\mu = 100$ and $\sigma^2 = 100$, what is the probability that X is greater than 2 standard deviations from its mean? Using eq. (6) the probability is less than or equal to,

$$p\left(\frac{|X - 100|}{10} > 2\right) \leq \frac{1}{2.25 \cdot 2^2} \leq 0.1\bar{1}.$$

Alternatively eq. (6) can be rearranged to find a constant k such that the true value lies within k standard deviation units from the mean with probability less than or equal to α:

$$k = \sqrt{\frac{1}{2.25 \cdot p\left(\frac{|X - \mu|}{\sigma}\right)}}$$

Setting $p = \alpha = 0.05$, then $k = \sqrt{\frac{1}{2.25 \cdot 0.05}} = 2.98$. Thus for any unimodal random variable the probability that X is greater than 2.98 standard deviation units from its mean is less than or equal to 0.05. Similarly an approximate $(1 - \alpha)\%$ confidence interval may be constructed as:

$$\widehat{AUC} \pm k \cdot \sqrt{Var\,(\widehat{AUC})}. \quad (7)$$

This method will be called the Bailer-Chebychev method. The Bailer-Chebychev method should provide confidence intervals that are less precise than Bailer's method, but more precise than the Bailer-Satterwaite method because usually $Z_{crit} < k < t_{crit}$.

Pai et al. (6) presented a non-parametric method to estimate the sampling distribution of AUC. The authors showed that bootstrapping the concentration-time profile in a sparse sample experimental protocol resulted in AUC estimates that were precise and accurate. In order for bootstrap confidence intervals to be valid the number of bootstraps drawn must be large, typically 1000 or more (7). The resulting bootstrap distribution of $\widehat{AUC}$ can then be used to determine the $(1 - \alpha)\%$ confidence interval for AUC by finding the $(1 - \alpha/2)\%$ tails of the sorted bootstrap sampling distribution.

One problem not addressed in Pai et al. (6) is what is the minimum number of blood sample replicates that must be collected at each time interval for bootstrap estimation to be valid. That question will not be addressed here, but it should be apparent that the number of distinct concentration-time combinations available for bootstrapping should be larger than the number of bootstrap samples actually drawn. Let q be the number of distinct combinations at any time i, r be the number of observations at any time i, and m be the number of time points such that $n = rm$. Assume no data are missing. Thus the data may be visualized as a matrix with q rows and m columns. Let

$$h_i = q \cdot h_{i-1},\ i = 2,3, \ldots, m,$$

where

$$h_1 = q^2.$$

The number of distinct combinations, h, available for bootstrapping while maintaining the independence between time columns is

$$h = \sum_{i=2}^{m} h_i. \tag{8}$$

Thus in the case of a 5 × 2 design where there are 5 time points and 2 animals per time point there are 3 distinct combinations at each time period {AA, BB, AB}. The number of combinations available for bootstrapping is only

$$h = \sum_{i=2}^{5} 3 \cdot h_{i-1} = \Sigma\ 9 + 27 + 81 + 243 = 360.$$

Thus when h is much less than the number of bootstrap samples drawn there will be considerable redundancy in the bootstrap distribution. The result being that the distribution takes on the characteristics of a discrete distribution as opposed to a continuous distribution. As the bootstrap distribution becomes more and more discrete, the resultant confidence intervals will become more and more biased.

However, bootstrapping is not the only nonparametric, computer intensive method that can be used to find the standard error of a random variable. The jackknife (8,9) may also be used and the algorithm for its use is as follows:

1. compute the sample statistic using all the available data and call this value $\hat{\sigma}$. Let n be the total sample size used to compute $\hat{\sigma}$.
2. delete the ith observation and recompute the sample statistic. Call this value θ_i.
3. compute the bias corrected pseudovalue estimate as $V_i = n\ \hat{\theta} - (n - 1)\ \theta_i$.
4. repeat steps 2 and 3 until i = n.
5. compute the standard error of the sample statistic as

$$SE(\hat{\theta}) = \sqrt{\frac{\sum_{i=1}^{n} (V_i - \overline{V})^2}{n(n-1)}},$$

where $\overline{V}$ is the average of all pseudovalues

$$\overline{V} = \frac{\sum_{i=1}^{n} V_i}{n}.$$

The jackknife estimate of the standard error of the sampling statistic is a non-parametric estimate that may used in place of $\sqrt{\widehat{\text{Var}\ (\text{AUC})}}$ in eq. (3) or (7). The jackknife cannot be used using the Bailer-Satterwaite method because an estimate of the weights, w_i, is not available. Computationally, the jackknife is much faster than the bootstrap and the minimum sample size requirement seen with the bootstrap is not an issue.

SIMULATION

A Monte Carlo simulation was undertaken to examine the precision and coverage of the following methods in estimation of the AUC in a sparse sampling setting: the Bailer-method, the Bailer-Satterwaite method, the Bailer-Chebychev method, the bootstrap, the Bailer-method using the jackknife estimate of the standard error, and the Bailer-Chebychev method using the jackknife estimate of the standard error. Plasma concentration-time samples were simulated using a one-compartment model with absorption. Pharmacokinetic parameters were generated from a log-normal distribution using the method of Johnson (10). Arbitrary pharmacokinetic parameters were defined to have a population mean of 50000, 1, and 0.1 for volume of distribution (V_d), absorption rate constant (K_a), and elimination rate constant (K_{el}), respectively. The corresponding pharmacokinetic parameters were set to have a between-animal coefficient of variation of either 15% or 50%, respectively. It was assumed that V_d was independent of K_a and K_{el}, but that k_a and K_{el} were correlated (r = 0.7071, r^2 = 0.50). Ten (10) animals were used in the experiment with two animals sampled at five arbitrary time points. The time points were {0, 2, 4, 8, and 24}. Once the concentration-time profile was generated, 10% random assay error was added to each measurement. AUC was calculated using the trapezoidal rule with the mean concentration at each time point. The experiments were simulated 1000 times with low and high between-animal variation. These conditions were very similar to the ones chosen by Pai et al. (3).

The coverage of each method was calculated by determining the proportion of simulations in which the corresponding 95% confidence intervals contained the population AUC. The population AUC was calculated by using an explicit equation for AUC using the one-compartment model:

$$AUC = \int_0^{t_1} \frac{D}{V_d}\left(\frac{K_a}{K_a - K_{el}}\right)[e^{-K_{el} \cdot t} - e^{-K_a \cdot t}] = \frac{D}{V_d}\left(\frac{K_a}{K_a - K_{el}}\right) \left[\frac{1}{K_a} e^{-K_a \cdot t_1} - \frac{1}{K_{el}} e^{-K_{el} \cdot t_1} - \frac{1}{K_a} + \frac{1}{K_{el}}\right] \tag{9}$$

where t_1 is the time of the last measurable concentration. Taking the limit as $t \to \infty$, eq. (9) simplifies to

$$AUC = \frac{D}{V_d}\left(\frac{K_a}{K_a - K_{el}}\right)\left[\frac{1}{K_{el}} - \frac{1}{K_a}\right]. \tag{10}$$

Substituting the population means and a dose of 1.5E6 into eq. (10), the population AUC was determined to be 300. The "best" method for computing the 95% confidence interval was determined within each simulation in the sense that "best" meant finding the method that had the smallest confidence interval range, yet still contained the population AUC. The proportion of simulations in which the methods gave lower confidence limits less than or equal to zero was determined. Also, the proportion of simulations in which Satterwaite's approximate degrees of freedom was less than two was determined.

A second simulation was done using the same conditions as above with the exception that the experimental design used was an "intensive sampling design" with time points {0, 0.5, 1, 1.5, 2, 3, 4, 6, 8, 10, 12, and 24} and four replicates per point point. The inter-animal coefficient of variation was set at 25%. This simulation was designed to compare the "sparse sampling design" with the "intensive sampling design." All simulations were done in Gauss, version 3.2, on a personal computer (11).

Table 1. Proportion of Simulations in Which the Corresponding 95% Confidence Intervals Contained the Population AUC

Method	Sparse Sampling Design		Intensive Sampling Design
	Low CV (15%)	High CV (50%)	Moderate CV (25%)
Bailer	0.77	0.82	0.82
Bailer-Satterwaite	0.95	0.95	0.86
Bailer-Chebychev	0.88	0.91	0.94
Bootstrap	0.62	0.68	0.77
Bailer with jackknife estimate of Var(AUC)	0.84	0.89	0.87
Bailer-Chebychev with jackknife estimate of Var(AUC)	0.95	0.95	0.97

RESULTS

The proportion of simulations in which the corresponding 95% confidence interval contained the population AUC is shown in Table 1. For both the sparse and intensive sampling paradigm, the bootstrap confidence interval had poor coverage in the sense that the proportion of confidence intervals which contained the true population AUC was not near its nominal expected value of 0.95. The inter-subject coefficient of variation had little impact on the coverage rate under the sparse sampling design. The parametric confidence intervals had greater coverage near the nomimal value, with coverage rates ranging from 0.77 (Bailer's method) to 0.95 (Bailer's Method with Satterwaite correction and Bailer-Chebychev method with jackknife estimate of the variance). There was very little change in the coverage rates of the parametric methods when switching from a sparse sampling design to an intensive sampling design. The greatest improvement in coverage was seen in switching from a sparse sampling design to an intensive sampling design with the bootstrap; coverage rates increased from 0.62 to 0.77.

The distribution of the simulations in which a particular method had the smallest confidence intervals and yet still contained the population AUC is shown in Table 2. Although the bootstrap had poor coverage, it had the smallest confidence interval range and contained the population AUC the majority of the time. This was observed regardless of the sampling design or inter-subject variation. An inverse relationship was observed between coverage rate and confidence interval length which occurred regardless of the inter-subject variation.

The proportion of simulations in which the lower bound of the resulting confidence interval was less than zero is shown in Table 3. Bailer's method with Satterwaite correction gave confidence intervals in which the lower bound was less than zero 15% of the time under low inter-subject variation and 58% of the time with high inter-subject variation. One-hundred (100%) of the simulations had Satterwaite estimates of the degrees of freedom less than two. None of the other methods had confidence intervals had lower limits whose value was less than or equal to zero when inter-subject variation was low. When inter-subject variation was high, all the parametric methods gave improbable lower estimates of the confidence interval at some point during the simulation. However, when an intensive sampling design was used none of the methods gave negative confidence limits.

Figure 1 shows the bootstrap distribution of the sparse sampling design and the intensive sampling design randomly drawn from one of the simulations. Clearly the sparse sampling design resulted in a more discrete resampling distribution compared to the smooth, continuous resampling distribution seen with the intensive sampling design. The intensive sampling bootstrap distribution appears to have a normal distribution, whereas the distribution of the sparse sampling design is not so apparent. A shift to the left in the mode of the bootstrap

Table 2. Distribution of Simulations in Which the Corresponding 95% Confidence Interval had the Smallest Range and Yet Still Contained the Population AUC

Method	Sparse Sampling Design		Intensive Sampling Design
	Low CV (15%)	High CV (50%)	Moderate CV (25%)
Bailer	0.14	0.15	0.06
Bailer-Satterwaite	0.01	0.01	0.02
Bailer-Chebychev	0.00	0.00	0.07
Bootstrap	0.62	0.68	0.77
Bailer with jackknife estimate of Var(AUC)	0.18	0.13	0.03
Bailer-Chebychev with jackknife estimate of Var(AUC)	0.00	0.00	0.03

Table 3. Proportion of Simulations in Which the Lower 95% Confidence Value was Less than or Equal to Zero

Method	Sparse Sampling Design		Intensive Sampling Design
	Low CV (15%)	High CV (50%)	Moderate CV (25%)
Bailer	0.00	0.02	0.00
Bailer-Satterwaite	0.15	0.58	0.00
Bailer-Chebychev	0.00	0.18	0.00
Bootstrap	0.00	0.00	0.00
Bailer with jackknife estimate of Var(AUC)	0.00	0.11	0.00
Bailer-Chebychev with jackknife estimate of Var(AUC)	0.00	0.40	0.00

distributions is seen in the intensive sampling design (302) compared to the sparse sampling design (330) primarily as a result of a difference in point estimates for AUC between the two designs.

DISCUSSION

The ideal method used to generate a (1-α)% confidence interval for AUC should be precise, give physically possible values, and yet still contain the population AUC with probability

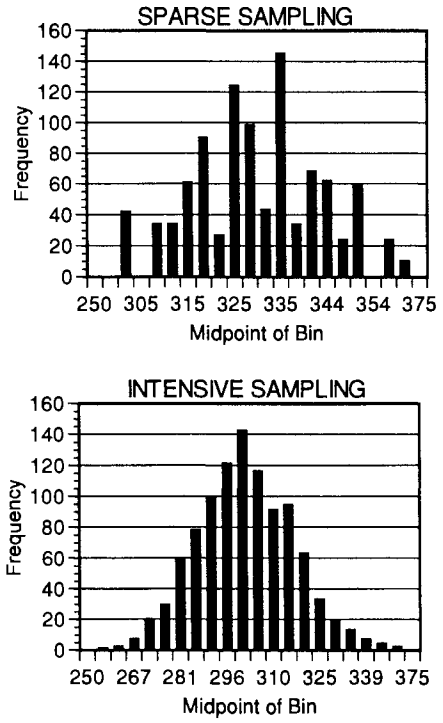

Fig. 1. One realization of the bootstrap distribution of the sparse sampling design (top) and the intensive sampling design (bottom). Samples were simulated to be collected at {0, 2, 4, 8, 24 h} with 2 replicates per time point in the sparse design and at {0, 0.5, 1, 1.5, 2, 3, 4, 6, 8, 10, 12, 24 h} with 4 replicates per time point in the intensive sampling design. One-thousand (1000) bootstrap simulations were simulated for each sampling design. The point estimate for AUC under the sparse design was 330, whereas for the intensive sampling design the point estimate was 302. The population AUC was 300.

coverage near the nominal (1-α) level. None of the methods studied meet all three of these criteria. The bootstrap gave precise confidence intervals with physically possible lower bounds 100% of the time. The bootstrap is, however, sensitive to the number of sampling points and number of replicates drawn at each time point. Sparse sampling designs run the risk of having too few possible combinations for the bootstrap to be valid. Bootstrapping sampling designs that have few random combinations changes the bootstrap distribution from a continuous one to a distribution that takes on the qualities of a discrete distribution. Other disadvantages of the boostrap are that it had the worst coverage rates of any method studied, is computational intensive, must be programmed using a language such as SAS or GAUSS, and is difficult to explain to a lay person.

In contrast, the symmetrical methods give coverage probabilities near their nomimal expected level, but at the expense of precision. The symmetrical methods sometimes resulted in confidence intervals whose lower limit was physically impossible, i.e., less than zero, because there is no guarantee with these methods that $c \cdot \sqrt{\text{Var}\,(\widehat{AUC})}$ will be less than $\widehat{AUC}$. The high percent of simulations under the sparse sampling design that had lower confidence interval estimates less than zero may have been the result of only two samples collected at each time period. Recall that the standard error of AUC is a linear function of the number of replicates at each time point. Increasing the number of replicates will decrease the standard error of AUC estimates. This is evidenced in the intensive sampling experimental design — none of the symmetrical methods had lower confidence interval estimates less than zero. Also, under an intensive sampling design both parametric and non-parametric methods give approximately equal results. Thus as the number of replicates increases the non-parametric method converges to the parametric methods. The advantage of the parametric confidence interval methods is that using the appropriate multiplier factor they give coverage rates near their nominal level, they are easy to compute, and are easy to explain to other investigators. The disadvantage is that they may result in unusually large confidence interval ranges that could have non-positive lower bounds.

One bias in this study was the choice of sampling times, which were chosen to provide good point estimates for the population AUC, a known quantity. Obviously coverage depends on how close the point estimate is to the true value. An inaccurate point estimate for AUC will probably have poor

coverage unless the precision is so poor as to be useless. Thus the utility of these methods will depend on how close the estimate of AUC is to its true value. For sparse sampling designs, sampling times are critical in obtaining good point estimates for AUC and indeed most sparse sampling designs are chosen based on some a priori pharmacokinetic information. If poor sampling times are selected (such as not accurately capturing the maximal plasma concentration) then it follows that AUC may be inaccurate. For this reason, an advantage of intensive sampling times (simply by nature of having more samples) will be better estimates of AUC, smaller standard errors, and more precise confidence intervals, but at the expense of more money.

In summary, choosing a method in which to calculate the confidence interval for AUC depends on which quality the researcher deems most important — high probability that the confidence interval contains the true population value or precise confidence intervals which may not contain the true population value. Parametric symmetrical confidence intervals of the type $\hat{y} \pm c \cdot SE(\hat{y})$ have greater coverage than nonparametric bootstrap confidence intervals, but tend to have wider ranges making them less precise. Overall, the bootstrap was the best confidence interval method: it had coverage rates near its nominal value and yet contained the true population value the majority of the time. However, the limitations of the method should be clearly recognized and it should not be used indiscriminately. Examination of the bootstrap distribution for its degree of discrete-ness must be part of the statistical process.

REFERENCES

1. A. J. Bailer. Testing for equality of area under the curves when using destructive measurement techniques. *J. Pharmacokin. Biopharm.* **16**:303–309 (1988).
2. J. R. Nedelman, E. Gibiansky, and D. T. Lau. Applying Bailer's methods for AUC confidence intervals to sparse sampling. *Pharm. Res.* **12**:124–128 (1995).
3. S. M. Pai, J. R. Nedelman, G. Hajian, R. Gibiansky, and V. K. Batra. Performance of Bailer's method for AUC confidence intervals from sparse non-normally distributed drug concentrations in toxicokinetic studies. *Pharm. Res.* **13**:1280–1282 (1996).
4. J. A. Rice. *Mathematical Statistics and Data Analysis,* Wadworth & Brooks/Cole Advanced Books and Software, Pacifac Grove CA, 1988.
5. A. J. Duncan. *Quality Control and Industrial Statistics,* Irwin, Homewood Il, 1986.
6. S. M. Pai, S. H. Fettner, G. Hajian, M. N. Cayen, and V. K. Batra. Characterization of AUCs from sparsely sampled populations in toxicology studies. *Pharm. Res.* **13**:1283–1290 (1996).
7. B. Efron. *The Jackknife, the Bootstrap, and Other Resampling Plans,* Society for Industrial and Applied Mathematics, Philadelphia, PA, 1982.
8. C. F. J. Wu. Jackknife, bootstrap, and other resampling methods in regression analysis. *Ann. Stats.* **14**:1261– 1295 (1986).
9. R. G. Miller. The jackknife — a review. *Biometrika* **61**:1–15 (1974).
10. M. E. Johnson. *Multivariate Statistical Simulation,* John Wiley and Sons, Inc., New York, 1987.
11. Aptech Systems. *GAUSS System, Version 3.2,* Aptech Systems, Inc., Kent, WA, 1995.