Confidence Intervals for Ratios of
AUC's in the Case of Serial Sampling:
A Comparison of Seven Methods

Thomas Jaki

Department of Mathematics and Statistics,

Lancaster University, United Kingdom

Martin J. Wolfsegger

Global Biopharm Preclinical Research and Development,

Baxter AG, Vienna, Austria

Meinhard Ploner

www.data-ploner.com, Brunico, Italy

To whom correspondence should be addressed: E-mail: jaki.thomas@gmail.com

March 19, 2008

Summary

Pharmacokinetic studies are commonly performed using the two-stage approach. The first stage involves estimation of pharmacokinetic parameters like the area under the concentration versus time curve (AUC) for each analysis subject separately and the second stage uses the individual parameter estimates for statistical inference. This two-stage approach is not applicable in sparse sampling situations where only one sample is available per analysis subject like in non-clinical *in-vivo* studies. In a serial sampling design only one sample is taken from each analysis subject. A simulation study was carried out to assess coverage, power and type I error of seven methods to construct two-sided 90% confidence intervals for ratios of two AUC's assessed in a serial sampling design, which can be used to assess bioequivalence in this parameter.

Keywords: AUC; bioequivalence; bootstrap; serial sampling design; serial sacrifice design; sparse sampling

1   Introduction

A formula to calculate a $1-\alpha$ confidence interval for the area under the concentration versus time curve (AUC) from 0 to the last observed time point assessed in a serial sampling design for normally distributed errors based on the linear trapezoidal rule is presented in [1]. Procedures based on the t-distribution are presented in [2] and in [3]. A simulation study for these approaches can be found in [4]. Wolfsegger & Jaki in [5] derive an estimator and a confidence interval for the AUC from 0 to infinity ($\text{AUC}_{0-\infty}$) for a serial sampling design.

Heinzl [6] presented a test for the null hypothesis of no difference between two AUC's from 0 to the last time point using the critical value from a t-distribution while Bailer & Ruberg [7] propose a permutation test for this null hypothesis using a z-statistic as the test statistic to be resampled.

In this note we will consider the ratio of two AUC's and the corresponding confidence intervals which are often assessed in bioequivalence studies using the confidence interval inclusion approach. One of the main advantages of considering ratios instead of the differences is interpretability. While the same conclusions can be obtained from both approaches it is often easier to discuss ratios. The problem of bioequivalency is discussed in great detail in the FDA guideline 'Statistical Approaches to Establishing Bioequivalence' [8]. The guideline also addresses *in-vitro* and *in-vivo* studies for investigational new drug applications in which sparse sampling may arise. Hu *et al.* [9] present a modelling approach, by ways of a nonlinear model, to assess bioequivalence or PK similarity for parameters estimated in a serial sampling design. Wolfsegger [10] presented three methods for calculation of a $1-\alpha$ confidence interval for the ratio of two AUC's from zero to the last time point without assuming a specific nonlinear model.

We will proceed by giving an overview of the ratio of AUC's in serial sampling designs and its estimation followed by a brief summary of seven methods to construct confidence intervals considered in this simulation study. In Section 4 we will describe the simulation study conducted in detail while Section 5 provides the simulation results. We will conclude with an example and a brief discussion of the findings and future directions.

2 Serial Sampling Design

Consider a study with two treatment groups, k, in which measurements are taken at J time points, t_j $(1 \leq j \leq J)$, and at each time point blood is sampled from n_j analysis subjects. It is assumed that the time points are the same for the two treatment groups and that each subject is only sampled once across all timepoints. This design leads to independent random variables, both per time point as well as between time points. Let X_{ijk} be the measured drug concentration from the

2

i^{th} analysis subject at time t_j receiving the treatment k. Let $E[X_{ijk}] = \mu_{jk}$ and $V[X_{ijk}] = \sigma_{jk}^2$ be the population mean and population variance at time point t_j. The general heteroscedastic model then is defined as

$$X_{ijk} = \mu_{jk} + \epsilon_{ijk} \tag{1}$$

where the errors, ϵ_{ijk}, are identical and independently distributed with continuous distribution G_{jk} and the range of X_{ijk} is effectively positive. The theoretical AUC from 0 to the last time point for a specific treatment to be estimated can be defined as

$$\text{AUC}_k = \sum_{j=1}^{J} w_j \mu_{jk}. \tag{2}$$

Using the linear trapezoidal rule, the weights w_j equal

$$
\begin{aligned}
w_1 &= \tfrac{1}{2}(t_2 - t_1) \\
w_j &= \tfrac{1}{2}(t_{j+1} - t_{j-1}) \quad (2 \le j \le J-1) \\
w_J &= \tfrac{1}{2}(t_J - t_{J-1}).
\end{aligned}
\tag{3}
$$

This AUC can be estimated by

$$\widehat{\text{AUC}_k} = \sum_{j=1}^{J} w_j \bar{X}_{jk} \tag{4}$$

with $V[\widehat{\text{AUC}_k}] = \sum_{j=1}^{J} w_j^2 \sigma_{jk}^2 n_{jk}^{-1}$ where $\bar{X}_{jk} = \frac{1}{n_{jk}} \sum_{i=1}^{n_{jk}} X_{ijk}$ represents the arithmetic mean at time point t_j in the kth treatment group.

In this note the parameter of interest therefore is defined as

$$\Delta = \frac{\sum_{j=1}^{J} w_j \mu_{j1}}{\sum_{j=1}^{J} w_j \mu_{j2}} \tag{5}$$

and will be estimated by

$$\widehat{\Delta} = \frac{\sum\limits_{j=1}^{J} w_j \bar{X}_{j1}}{\sum\limits_{j=1}^{J} w_j \bar{X}_{j2}}. \tag{6}$$

Wolfsegger [10] shows that $\sqrt{n}(\hat{\Delta} - \Delta)$ is asymptotically normal distributed. A similar approach can be used to estimate the ratio of AUCs from 0 to infinity by replacing the presented formulas for the AUC with the respective formulas for the AUC from 0 to infinity in [5]. An approximate $1 - \alpha$ confidence interval for the ratio can be obtained by application of Fieller's theorem [11] based on the asymptotic normal distribution and corresponding standard errors.

3   Methods Compared

In this section we will give a brief overview of the seven different methods to construct (approximate) $1 - \alpha$ confidence intervals that are considered in the simulation study of Section 4. We will denote the parameter of interest by θ and a standard estimator for it (e. g. maximum likelihood estimator) by $\hat{\theta}$. Further, in accordance to Davison & Hinkley [12] and others, we will use the 'star' notation to indicate bootstrap based estimators. Therefore, $\hat{\theta}^*(b)$ is the bth bootstrap replication of the estimator $\hat{\theta}$. The α-percentile of the set of bootstrap replicates $\{\hat{\theta}^*(b), \ b = 1, \ldots, B\}$ will be denoted by $\hat{\theta}^*_\alpha$.

The simplest and probably most often used resampling method to construct confidence intervals among practitioners is the percentile method [13, pp. 170-177]. This method assumes that for an unknown monotone increasing transformation $h(\theta)$, a statement of the type

$$h(\hat{\theta}) - h(\theta) \sim N(0, \sigma^2_{h(\hat{\theta})})$$

4

holds. A simple approximate $1 - \alpha$ confidence interval can be found as

$$[\hat{\theta}^*_{\frac{\alpha}{2}}; \hat{\theta}^*_{1-\frac{\alpha}{2}}]. \tag{7}$$

An improvement upon the percentile method was introduced by Efron [14] in the bias-corrected and accelerated percentile method, BCa. As before this method has the assumption of the existence of an unknown monotone increasing transformation, $h(\theta)$, for which the statement

$$h(\hat{\theta}) - h(\theta) \sim N(-z_0 \sigma_{h(\theta)}, \sigma^2_{h(\theta)})$$

holds. z_0 denotes a constant bias correction factor and the relationship $\sigma_{h(\theta)} = 1 - ah(\theta)$ holds for some a, the skewness correction factor (also referred to as the acceleration constant). An approximate $1 - \alpha$ confidence interval therefore can be estimated as

$$\left[\hat{\theta}^*_{\Phi(z^{corr}_{\frac{\alpha}{2}})}; \hat{\theta}^*_{\Phi(z^{corr}_{1-\frac{\alpha}{2}})} \right] \tag{8}$$

with $z^{corr}_\alpha = z_0 + \frac{z_0 + z_\alpha}{1 - a(z_0 + z_\alpha)}$, $\Phi(x)$ the CDF of the standard normal distribution and z_α the corresponding α-percentile. While it is not necessary for both methods described above to know the transformation $h(.)$, they will fail if the transformation to a normal distribution is not possible.

A different type of bootstrap intervals can be constructed based on pivotal quantities. The hybrid method [15] is the most intuitive of these approaches. The idea is to estimate the distribution of the pivot $\hat{\theta} - \theta$ by the bootstrap distribution constructed upon $\hat{\theta}^*(b) - \hat{\theta}$, the bootstrap equivalent of the pivot. The confidence limits then can be found to be

$$[2\hat{\theta} - \hat{\theta}^*_{1-\frac{\alpha}{2}}; 2\hat{\theta} - \hat{\theta}^*_{\frac{\alpha}{2}}]. \tag{9}$$

Although this method has a totally different justification than the percentile method, it is easy to see the connection between both methods. The hybrid method, however, in this note is particularly interesting since the next method, called the

ratio method is based upon its idea.

For the ratio method introduced in [16], we consider a different pivot, $\frac{\hat{\theta}}{\theta}$, whose distribution is to be approximated by the bootstrap equivalent quantity, $\frac{\hat{\theta}^*(b)}{\hat{\theta}}$ which yields

$$\left[\frac{\hat{\theta}^2}{\hat{\theta}^*_{1-\frac{\alpha}{2}}}; \frac{\hat{\theta}^2}{\hat{\theta}^*_{\frac{\alpha}{2}}}\right] \tag{10}$$

as the confidence bounds. A very particular feature of this method is that it should only be used if the values of the statistic are strictly positive, a feature clearly prominent in the estimation of the AUC and ratio of two AUC's. It does, however, also imply that the propagated error, which arises by every further iteration, emerges in an approximately multiplicative way. In other words, the method will work poorly if the distribution to be estimated deviates strongly from the estimate. This, furthermore, means that the sample sizes required tend to be large.

The assumptions of the bootstrap-t-interval [17] are less restrictive than those of the hybrid method, allowing the pivot to be of a more general form $\frac{\hat{\theta}-\theta}{\hat{\sigma}}$ whose distribution is to be estimated by the distribution of $t^* = \frac{\hat{\theta}^*(b)-\hat{\theta}}{\hat{\sigma}^*}$. The resulting confidence interval then becomes

$$[\hat{\theta} - t^*_{1-\frac{\alpha}{2}}\hat{\sigma}; \hat{\theta} - t^*_{\frac{\alpha}{2}}\hat{\sigma}]. \tag{11}$$

Notice that, in order to obtain good results for this type of interval, in addition to the bootstrap statistic t^* a separate estimator of the standard deviation, $\hat{\sigma}^*$ (usually the jackknife estimator for σ) is needed. To obtain this estimator generally one additional 'layer' of resampling is necessary, making this method often more computational intensive than the intervals presented before. Furthermore, this additional bootstrap layer may lead to poor estimates in the case of small sample sizes per time point which is frequently the case in non-clinical *in-vivo* studies. In this study, instead of a resampled estimator for σ, we used the asymptotic standard deviation

derived in [10] to avoid these problems.

The last two methods considered in the manuscript differ from the previous methods significantly as they are not resampling based methods. The asymptotic confidence interval simply uses normal theory to obtain a z-type interval as

$$[\hat{\theta} + z_{\frac{\alpha}{2}}\hat{\sigma}_{\hat{\theta}}; \hat{\theta} + z_{1-\frac{\alpha}{2}}\hat{\sigma}_{\hat{\theta}}].$$

This technique has the implicit assumption that some form of a limiting theory holds for the statistic such that $\frac{\hat{\theta}-\theta}{\hat{\sigma}_{\hat{\theta}}} \sim N(0,1)$. This approach will result in a symmetric confidence interval around the observed effect which is often inappropriate in the case of ratios since the parameter space for ratios ranges from 0 to infinity. The Fieller-type procedure, which is described in great detail in [10] for the ratio of AUC's, on the other hand uses a t-distribution to model the distribution of the corresponding pivot. The degrees of freedom are approximated using Satterthwaite's method [18], yielding a complicated looking, but computationally straightforward, confidence interval.

4 Simulations

The following one-compartmental model with first order absorption and elimination after extravascular administration (e. g. oral, intramuscular, rectal, etc.) was used for data generation

$$X_{ij} = f(t_j) + \epsilon_{ij} = \frac{k_a F D}{V(k_a - \lambda)}\left(e^{-\lambda t_j} - e^{-k_a t_j}\right) + \epsilon_{ij} \tag{12}$$

with the parameterization $\lambda = 0.0693$, $k_a = 0.231$, $V = 10$, $F = 1$ and the dose $D = 500$. To eliminate the bias created by the linear trapezoidal rule to approximate the integral, the true AUC for treatment k was defined as in Equation (2) using $\mu_{jk} = f(t_j) = \frac{k_a F D}{V(k_a-\lambda)}\left(e^{-\lambda t_j} - e^{-k_a t_j}\right)$ specified at baseline and ten time points (1h, 2h, 3h, 4h, 6h, 8h, 12h, 18h, 24h and 36h) post study drug administration.

Both, model and time points, are taken from Gibaldi & Perrier, page 440 [19]. For each subject only one observation is generated across all time points to reflect the sparse sampling situation of the serial sampling design. As a result all generated observations are independent of each other. Consequently missing data effectively reduces the sample size at specific time points, but need no special consideration in the estimation of the AUC's in serial sampling designs as the presented procedures allow for unequal sample sizes per time point.

Normal, log-normal and double exponential distributed errors were used for various combinations of sample size (N=3, 5 and 10) and time point variabilities. Table 1 shows the three variability scenarios that were studied which are inspired by real data. The coefficient of variations at later time points tend to increase which might be due to the inaccuracy of an assay to determine the drug concentration in the blood when dealing with values close to the limit of detection. 10000 simulation runs (yielding an estimation error of 0.003 for nominal coverage of 90%) were carried out for each parameter setting with preselected sample sizes. Within each simulation run, N random samples were generated for each of the specified time points. 1000 bootstrap replications were used for bootstrap based confidence intervals.

Table 1: Coefficient of Variations (%) Used for Simulations

		Time point		
Scenario	Baseline	1 to 18 hours	24 hours	36 hours
1	0%	20%	20%	20%
2	0%	40%	40%	40%
3	0%	40%	80%	80%

Empirical coverage estimates are reported for a nominal coverage probability of $1 - \alpha = 0.90$. Empirical lower and upper tail probabilities presented additionally were defined as the probability that the true ratio is below/above the calculated two-sided 90% confidence interval for the ratio. Empirical type I errors, γ, in the sense of bioequivalency are reported under non-equivalence for a nominal coverage

probability of $1 - \alpha = 0.90$.

Empirical power and type I error estimates are presented using the conventional limits of bioequivalence for ratio of averages ranging from 0.8 to 1.25. Expected values at different time points used to study $\Delta = 0.799$ and $\Delta = 0.90$ were determined by varying parameters λ, k_a, V, D and F accordingly.

All simulations were performed with R version 2.4.1 [20]. Empirical coverage, power and type I error for the methods compared for a given scenario and sample size per time point were calculated on basis of the same simulation runs.

4.1 Generation of Error Terms

Our starting point to model the uncertainty in the AUC will be normally distributed errors, since some of the theoretical results are based upon this assumption [10]. Additionally, log-normal distributed errors will be evaluated as drug levels cannot be negative, while the upper end is open which may lead to a non-symmetrical distribution. The last distribution considered, a double exponential distribution, is used to represent frequent extreme observations. For all distributions we use the parameterization in Casella & Berger, pages 623-625 [21].

The variation in the drug levels at a given time point has been fixed in terms of the dimensionless coefficient of variation, $c_v = \sigma/\mu$, where μ is the mean drug level at a given time point and σ the corresponding standard deviation. While the mean level, μ, is given by the true AUC the challenge is to find the parameters of the underlying distribution that yield a pre-specified coefficient of variation. For the normal distribution, where the parameters of the distribution correspond to the moments of interest, the solution is simple and given by data generated from a $N\left(\mu, \mu^2 c_v^2\right)$ -distribution.

To find the parameters for the other two distributions, however, is more difficult.

9

Lets first consider the log-normal case for which we let α and β be the mean and standard deviation of the underlying normal distribution, respectively. The mean of the distribution then is $\exp\left(\alpha + \beta^2/2\right)$ and the coefficient of variation can be found to be $\sqrt{\exp\left(\beta^2\right) - 1}$. To determine the proper values for α and β for given values of μ and c_v we need to solve the system of equations

$$
\begin{aligned}
\mu &= \exp\left(\alpha + \beta^2/2\right) \\
c_v &= \sqrt{\exp\left(\beta^2\right) - 1},
\end{aligned}
\tag{13}
$$

which yields $\alpha = \ln\left(\mu\right) - \frac{1}{2}\ln\left(1 + c_v^2\right)$ and $\beta^2 = \ln\left(1 + c_v^2\right)$. The data for this model therefore can be obtained by generating data from a log-normal distribution with the parameters from above.

Using the equivalent approach for the double exponential distribution parameterized by α and β gives the set of equations

$$
\begin{aligned}
\mu &= \alpha \\
c_v &= \frac{1}{\alpha}\sqrt{2\beta^2},
\end{aligned}
\tag{14}
$$

which are solved by $\alpha = \mu$ and $\beta^2 = \frac{1}{2}\mu^2 c_v^2$. Thus, we can generate the heavy tailed distribution by using a double exponential distribution with the respective parameters.

In the rare case that we generated a negative value using the normal and double exponential distribution, we chose to simply replace the value by a newly generated nonnegative observation. While we are aware that this will change the true underlying structure of the model we found that the difference in μ and c_v was negligible as only a few observations had to be replaced. Further, this approach was applied for both samples for which the ratio of AUC's was calculated and additional simulation studies indicated that this replacement method has asymptotically no effect on the

true ratios studied.

4.2 Contaminated Data

Aside from the 'clean' data situations described before, we also study contaminated data. The first set of contaminations is thought to describe either erroneous data due to measurement errors or extreme drug-levels for a few subjects at some time points. In our model we use the same normal and log-normal distributions as described in Section 4.1 with $c_v = 0.2$ but have 10% of the data generated come from the same distribution with the mean shifted by 3 standard deviations. For the normal distribution this yields a true c_v of 0.254 while the coefficient of variation becomes 0.305 for log-normal data.

The second type of contamination studied is a change in the error distribution over time. This is motivated by potentially different decomposition of the drug among subjects. We will model this by having the initial error distribution be normal and change it to double exponential for the last three time points. The coefficient of variation will be 0.2 at all time points. Note that this approach is different to just changing c_v for the last 3 time points since this still has the coefficient of variation fixed while increasing the number of extreme values.

5 Results

Tables 2 - 5 show some of the results of the simulation study for the 'clean' data for the scenarios in Table 1. The omitted results are available from the authors upon request. For all combinations of sample size, distribution and Δ, the empirical overall coverage and the empirical power/type-I-error is presented. Additionally left and right tail coverages are included as they are important for one sided hypothesis such as tests for non-inferiority. The power in this context is defined as the probability that the confidence interval estimated is within 0.8 and 1.25, the conventional range for bioequivalency.

The bootstrap confidence intervals with the exception of the bootstrap-t-interval show a clear pattern in terms of coverage for all parameter settings considered. For a sample size of 3 per time point these intervals undercover severely, yielding only an empirical coverage of approximately 0.8 while they approach nominal coverage as sample size increase. The bootstrap-t-interval as well as the asymptotic and Fieller type interval are superior in coverage than the other procedures investigated for all sample sizes and scenarios studied. Empirical coverage of the asymptotic procedure is slightly below the nominal coverage level for a sample size of 3 whereas the bootstrap-t and Fieller approach yield nominal coverage. In addition, the asymptotic interval indicates imbalance in tail probabilities. On average, the Fieller approach is marginally more conservative in terms of coverage than the bootstrap-t approach with double exponential errors and a sample size of 3 per time point. However, differences in power and type-I-error between the Fieller interval and bootstrap-t-interval are small across all sample size and error distributions considered.

Another interesting point is that there appears to be no influence of the error distribution on the coverage of the different types of confidence intervals while the power clearly depends on the distributional shape and the values of the coefficient of variation across time points.

A surprising result can be found for the hybrid intervals for which the empirical coverage of the tails differ strongly such that the lower bound covers much better and, in fact almost always, yields the desired coverage on the lower tail, while badly undercovering on the upper tail. Interestingly the same behavior can not be found for the related ratio method. While the poor coverage suggests a poor approximation of the distribution of the pivot by the bootstrap distribution, the resulting intervals appear to be more symmetric in probability.

In the simulation for equivalent true AUC's ($\Delta = 1$), presented in Tables 2 and

12

5, the power quickly reaches one for a coefficient of variation of 0.2 and 0.4 as sample size increases. An interesting side note here is that the double exponential errors yield a higher power than the other two error distributions considered. This may be due to the effectively smaller value of c_v obtained by replacing negative values by a new, positive, random error.

For the situation of different AUC's that are still considered to be bioequivalent ($\Delta = 0.9$) hardly any difference in the performance of the intervals dependent on the error distribution can be found once again. Only for the high variation scenario a slight difference in power between the symmetric and the non-symmetric error distribution can be seen (Table 5). The magnitude of this difference is at most 7%. It is notable, however, that the power of all intervals is markedly lower than for $\Delta = 1$ across all sample sizes with up to 40% reduced power for scenario 3.

The most encouraging results for all the methods considered can be found for AUC's that are not considered bioequivalent. Even when the true parameter is only 0.001 outside of the range of bioequivalency, all the confidence intervals reflect this yielding only a type-I-error, γ, that is the probability to decide on equivalency when you should not, of approximately 5%. Even more astonishing is that the value of γ is stable across all choices of distribution, sample size and variations.

5.1   Contamination

We will now look at the performance of the intervals for contaminated data. In terms of coverage the results mimic the pattern discussed for the 'clean' data. The bootstrap-t-interval, asymptotic and Fieller interval are superior in coverage to the other procedures investigated across all sample size and error distributions considered.

The power of the contaminated errors show the expected pattern as the power

Table 2: Empirical Coverage and Power for $\Delta = 1$ Using a Nominal Coverage of 90% for Scenario 1

		Error distribution					
		Normal		Log-normal		Double exponential	
N	Method	Coverage	Power	Coverage	Power	Coverage	Power
3	Percentile	0.8063 (0.9048;0.9015)	0.9459	0.8013 (0.9012;0.9001)	0.9505	0.7986 (0.9004;0.8982)	0.9373
	Hybrid	0.8072 (0.9240;0.8832)	0.9450	0.7948 (0.9165;0.8783)	0.9459	0.8109 (0.9259;0.8850)	0.9378
	Ratio	0.8067 (0.9049;0.9018)	0.9490	0.7977 (0.8987;0.8990)	0.9492	0.8091 (0.9055;0.9036)	0.9436
	BCa	0.8046 (0.9032;0.9014)	0.9464	0.8001 (0.9003;0.8998)	0.9493	0.7985 (0.9000;0.8985)	0.9338
	Boot-t	0.9110 (0.9579;0.9531)	0.8438	0.9110 (0.9579;0.9531)	0.8388	0.9118 (0.9602;0.9516)	0.7951
	Asymptotic	0.8835 (0.9514;0.9321)	0.8985	0.8796 (0.9474;0.9322)	0.9002	0.8857 (0.9543;0.9314)	0.8847
	Fieller	0.9072 (0.9527;0.9545)	0.8699	0.9078 (0.9541;0.9537)	0.8718	0.9314 (0.9684;0.9630)	0.8245
5	Percentile	0.8513 (0.9259;0.9254)	0.9934	0.8483 (0.9247;0.9236)	0.9952	0.8415 (0.9219;0.9196)	0.9913
	Hybrid	0.8526 (0.9397;0.9129)	0.9937	0.8443 (0.9374;0.9069)	0.9945	0.8524 (0.9428;0.9096)	0.9921
	Ratio	0.8525 (0.9261;0.9264)	0.9935	0.8469 (0.9251;0.9218)	0.9944	0.8526 (0.9273;0.9253)	0.9926
	BCa	0.8522 (0.9252;0.9270)	0.9930	0.8469 (0.9229;0.9240)	0.9948	0.8388 (0.9193;0.9195)	0.9898
	Boot-t	0.9040 (0.9534;0.9506)	0.9871	0.8999 (0.9524;0.9475)	0.9878	0.8897 (0.9471;0.9426)	0.9746
	Asymptotic	0.8925 (0.9526;0.9399)	0.9892	0.8899 (0.9531;0.9368)	0.9910	0.8907 (0.9533;0.9374)	0.9860
	Fieller	0.9038 (0.9510;0.9528)	0.9873	0.9039 (0.9535;0.9504)	0.9896	0.9025 (0.9523;0.9502)	0.9840
10	Percentile	0.8790 (0.9426;0.9364)	1.0000	0.8698 (0.9364;0.9334)	1.0000	0.8765 (0.9415;0.9350)	0.9999
	Hybrid	0.8763 (0.9520;0.9243)	1.0000	0.8687 (0.9448;0.9239)	1.0000	0.8803 (0.9536;0.9267)	0.9999
	Ratio	0.8802 (0.9430;0.9372)	1.0000	0.8689 (0.9351;0.9338)	1.0000	0.8807 (0.9442;0.9365)	0.9999
	BCa	0.8768 (0.9404;0.9364)	1.0000	0.8683 (0.9325;0.9358)	1.0000	0.8723 (0.9379;0.9344)	0.9999
	Boot-t	0.9019 (0.9538;0.9481)	1.0000	0.8936 (0.9480;0.9456)	1.0000	0.8945 (0.9504;0.9441)	0.9999
	Asymptotic	0.8985 (0.9560;0.9425)	1.0000	0.8882 (0.9496;0.9386)	1.0000	0.8988 (0.9567;0.9421)	0.9999
	Fieller	0.9023 (0.9529;0.9494)	1.0000	0.8949 (0.9475;0.9474)	1.0000	0.9053 (0.9547;0.9506)	0.9999

Values in parentheses refer to left and right tail probabilities
Power is based on conventional limits for bioequivalency
N is sample size per time point

increases with sample size for $\Delta = 1$ and $\Delta = 0.9$ while the type-I-error is rather stable at 5% for non-equivalent AUC's. The biggest and most surprising difference can be seen between the error distributions. The contaminated log-normal distribution (Table 6) has markedly better power for the values of the parameter that are considered bioequivalent than the contaminated normal distribution (Table available upon request). In fact the power of the contaminated log-normal distribution is up to 25% higher than for normally distributed errors and is almost identical to the power for the 'clean' normal errors. This result suggests that the power of the intervals is not so much influenced by the skewness of the error distribution than by the frequency of outliers. Keeping this in mind it is striking that the same feature was not seen in the initial simulations using 'clean' errors. This further indicates that the extreme tail needs to be very heavy in order to yield a reduction in power as seen here.

The results for changing error distribution once again show the familiar patterns and the details therefore have been omitted, but are available upon request. The coverage of the bootstrap based methods converges in sample size toward nominal level for all these methods besides the bootstrap-t-interval whos coverage is on target

Table 3: Empirical Coverage and Type I Error for $\Delta = 0.799$ Using a Nominal Coverage of 90% for Scenario 1

		Error distribution					
		Normal		Log-normal		Double exponential	
N	Method	Coverage	γ	Coverage	γ	Coverage	γ
3	Percentile	0.8046 (0.9028;0.9018)	0.0949	0.8047 (0.9020;0.9027)	0.0946	0.7984 (0.8985;0.8999)	0.0961
	Hybrid	0.8082 (0.9235;0.8847)	0.1107	0.8036 (0.9194;0.8842)	0.1123	0.8090 (0.9222;0.8868)	0.1090
	Ratio	0.8084 (0.9036;0.9048)	0.0921	0.8013 (0.9007;0.9006)	0.0965	0.8071 (0.9033;0.9038)	0.0931
	BCa	0.8046 (0.9015;0.9031)	0.0930	0.8055 (0.9012;0.9043)	0.0923	0.7922 (0.8948;0.8974)	0.0985
	Boot-t	0.9142 (0.9572;0.9570)	0.0414	0.9149 (0.9588;0.9561)	0.0422	0.9100 (0.9580;0.9520)	0.0453
	Asymptotic	0.8853 (0.9505;0.9348)	0.0636	0.8821 (0.9496;0.9325)	0.0645	0.8826 (0.9512;0.9314)	0.0655
	Fieller	0.9258 (0.9627;0.9631)	0.0357	0.9187 (0.9591;0.9596)	0.0384	0.9047 (0.9533;0.9514)	0.0466
5	Percentile	0.8462 (0.9246;0.9216)	0.0751	0.8451 (0.9246;0.9205)	0.0754	0.8429 (0.9197;0.9232)	0.0727
	Hybrid	0.8481 (0.9407;0.9074)	0.0897	0.8439 (0.9378;0.9061)	0.0905	0.8512 (0.9392;0.9120)	0.0841
	Ratio	0.8465 (0.9251;0.9214)	0.0761	0.8453 (0.9244;0.9209)	0.0761	0.8506 (0.9238;0.9268)	0.0703
	BCa	0.8449 (0.9232;0.9217)	0.0753	0.8436 (0.9221;0.9215)	0.0757	0.8384 (0.9161;0.9223)	0.0751
	Boot-t	0.8984 (0.9511;0.9473)	0.0503	0.8976 (0.9511;0.9465)	0.0506	0.8894 (0.9442;0.9452)	0.0524
	Asymptotic	0.8861 (0.9529;0.9332)	0.0635	0.8853 (0.9517;0.9336)	0.0654	0.8892 (0.9509;0.9383)	0.0584
	Fieller	0.9005 (0.9518;0.9487)	0.0492	0.8992 (0.9520;0.9472)	0.0500	0.9055 (0.9518;0.9537)	0.0436
10	Percentile	0.8760 (0.9396;0.9364)	0.0595	0.8763 (0.9401;0.9362)	0.0606	0.8762 (0.9371;0.9391)	0.0572
	Hybrid	0.8751 (0.9488;0.9263)	0.0679	0.8749 (0.9489;0.9260)	0.0698	0.8815 (0.9501;0.9314)	0.0632
	Ratio	0.8776 (0.9398;0.9378)	0.0581	0.8760 (0.9395;0.9365)	0.0587	0.8817 (0.9401;0.9416)	0.0552
	BCa	0.8757 (0.9379;0.9378)	0.0584	0.8740 (0.9375;0.9365)	0.0594	0.8731 (0.9347;0.9384)	0.0578
	Boot-t	0.8976 (0.9501;0.9475)	0.0483	0.8973 (0.9501;0.9472)	0.0484	0.8941 (0.9478;0.9463)	0.0511
	Asymptotic	0.8953 (0.9531;0.9422)	0.0534	0.8937 (0.9531;0.9406)	0.0556	0.8977 (0.9537;0.9440)	0.0526
	Fieller	0.8994 (0.9506;0.9488)	0.0486	0.9007 (0.9516;0.9491)	0.0471	0.9014 (0.9509;0.9505)	0.0464

Values in parentheses refer to left and right tail probabilities
N is sample size per time point

for all sample sizes. Additionally we see again the increasing power in sample size for equivalent AUC's while the type-I-error remains stable at about 5% if the areas are in fact different. This suggests that the constructed intervals are rather insensitive to distributional assumptions.

6 Example

To illustrate the seven methods the rats data [3] were used. We consider testing for dose proportionality using the plasma concentrations based on doses of 30 mg/kg and 100 mg/kg. The observed concentrations are scaled by the administered dose and the corresponding AUC's tested for equivalence. Figure 1 shows the two sets of data while Table 7 displays the seven confidence intervals discussed based on 10000 bootstrap resamples.

The graph of the data shows that plasma concentrations at a dose of 30 mg/kg are slightly lower at one and two hours and shows much higher variability at 4 hours while otherwise matching the scaled results for a dose of 100 mg/kg well. All seven

Table 4: Empirical Coverage for $\Delta = 0.9$ Using a Nominal Coverage of 90% for Scenario 3

		Error distribution		
		Normal	Log-normal	Double exponential
N	Method	Coverage	Coverage	Coverage
3	Percentile	0.8100 (0.9052;0.9048)	0.7931 (0.8965;0.8966)	0.7998 (0.9002;0.8996)
	Hybrid	0.8180 (0.9463;0.8717)	0.7894 (0.9356;0.8538)	0.8049 (0.9398;0.8651)
	Ratio	0.8201 (0.9100;0.9101)	0.7835 (0.8911;0.8924)	0.8039 (0.9024;0.9015)
	BCa	0.8103 (0.9035;0.9068)	0.7922 (0.8939;0.8983)	0.7970 (0.8968;0.9002)
	Boot-t	0.9165 (0.9634;0.9531)	0.9091 (0.9617;0.9474)	0.9077 (0.9582;0.9495)
	Asymptotic	0.8921 (0.9646;0.9275)	0.8743 (0.9579;0.9164)	0.8834 (0.9591;0.9243)
	Fieller	0.9220 (0.9634;0.9586)	0.9098 (0.9572;0.9526)	0.9240 (0.9644;0.9596)
5	Percentile	0.8522 (0.9265;0.9257)	0.8454 (0.9238;0.9216)	0.8446 (0.9207;0.9239)
	Hybrid	0.8587 (0.9615;0.8972)	0.8378 (0.9520;0.8858)	0.8497 (0.9522;0.8975)
	Ratio	0.8588 (0.9301;0.9287)	0.8412 (0.9221;0.9191)	0.8498 (0.9225;0.9273)
	BCa	0.8519 (0.9241;0.9278)	0.8428 (0.9189;0.9239)	0.8403 (0.9151;0.9252)
	Boot-t	0.9059 (0.9556;0.9503)	0.8947 (0.9508;0.9439)	0.8951 (0.9480;0.9471)
	Asymptotic	0.8975 (0.9652;0.9323)	0.8825 (0.9569;0.9256)	0.8895 (0.9572;0.9323)
	Fieller	0.9135 (0.9565;0.9570)	0.9129 (0.9582;0.9547)	0.9062 (0.9511;0.9551)
10	Percentile	0.8846 (0.9435;0.9411)	0.8647 (0.9328;0.9319)	0.8740 (0.9359;0.9381)
	Hybrid	0.8848 (0.9649;0.9199)	0.8646 (0.9555;0.9091)	0.8778 (0.9577;0.9201)
	Ratio	0.8868 (0.9447;0.9421)	0.8635 (0.9324;0.9311)	0.8743 (0.9361;0.9382)
	BCa	0.8830 (0.9397;0.9433)	0.8611 (0.9274;0.9337)	0.8709 (0.9306;0.9403)
	Boot-t	0.9072 (0.9559;0.9513)	0.8878 (0.9454;0.9424)	0.8923 (0.9461;0.9462)
	Asymptotic	0.9055 (0.9641;0.9414)	0.8839 (0.9544;0.9295)	0.8950 (0.9561;0.9389)
	Fieller	0.9068 (0.9552;0.9516)	0.8968 (0.9489;0.9479)	0.8993 (0.9486;0.9507)

Values in parentheses refer to left and right tail probabilities

N is sample size per time point

lower bounds of the two-sided 90% confidence intervals fall below the conventional margins of bioequivalence for ratios of averages ranging from 0.8 to 1.25, indicating that dose proportionality can not be established. As expected the bootstrap-t and the Fieller interval are widest, which is reflected in the higher coverage in the simulation studies above.

7 Discussion

In this note the performance of seven different types of confidence intervals for the ratio of two area under the concentration versus time curves in a serial sampling design, where the otherwise comonly used two-stage approach is not applicable due to sparsity, are evaluated. For all combinations of variation, sample size and distribution the asymptotic, Fieller and bootstrap-t-interval are clearly superior to the other approaches considered. Among these three intervals only small difference in coverage or power/type-I-error can be seen which is particularly surprising for the asymptotic interval for contaminated data as it implies that the asymptotic normality shown in [10] is reached quickly. This result strengthened the argument to use the asymptotic

Table 5: Power and Type I error for different Δ Using a Nominal Coverage of 90% for Scenario 3

		Error distribution								
		Normal			Log-normal			Double exponential		
		$\Delta = 1$	$\Delta = 0.9$	$\Delta = 0.799$	$\Delta = 1$	$\Delta = 0.9$	$\Delta = 0.799$	$\Delta = 1$	$\Delta = 0.9$	$\Delta = 0.799$
N	Method	Power	Power	γ	Power	Power	γ	Power	Power	γ
3	Percentile	0.2447	0.1823	0.0644	0.2249	0.1685	0.0680	0.3418	0.2554	0.0819
	Hybrid	0.2164	0.1845	0.0753	0.1875	0.1700	0.0776	0.3153	0.2667	0.1010
	Ratio	0.2489	0.1852	0.0632	0.2232	0.1662	0.0702	0.3441	0.2520	0.0803
	BCa	0.2465	0.1789	0.0643	0.2227	0.1656	0.0669	0.3376	0.2514	0.0818
	Boot-t	0.0506	0.0366	0.0118	0.0469	0.0308	0.0143	0.1051	0.0769	0.0212
	Asymptotic	0.0879	0.0665	0.0240	0.0771	0.0634	0.0268	0.1651	0.1333	0.0393
	Fieller	0.0387	0.0298	0.0089	0.0492	0.0363	0.0159	0.0890	0.0645	0.0166
5	Percentile	0.4805	0.3170	0.0696	0.4306	0.2832	0.0702	0.5714	0.3715	0.0730
	Hybrid	0.4525	0.3589	0.0923	0.3903	0.3133	0.0952	0.5526	0.4157	0.0974
	Ratio	0.4898	0.3184	0.0665	0.4284	0.2835	0.0729	0.5747	0.3693	0.0703
	BCa	0.4756	0.3106	0.0668	0.4261	0.2788	0.0679	0.5672	0.3618	0.0714
	Boot-t	0.3367	0.2138	0.0413	0.2590	0.1638	0.0407	0.4162	0.2644	0.0472
	Asymptotic	0.3603	0.2569	0.0559	0.3021	0.2135	0.0575	0.4634	0.3153	0.0626
	Fieller	0.3119	0.1883	0.0341	0.2366	0.1453	0.0335	0.4231	0.2569	0.0414
10	Percentile	0.8550	0.5196	0.0574	0.7809	0.4685	0.0657	0.8973	0.5656	0.0605
	Hybrid	0.8428	0.5825	0.0767	0.7656	0.5307	0.0878	0.8917	0.6234	0.0764
	Ratio	0.8562	0.5225	0.0561	0.7798	0.4667	0.0663	0.8994	0.5661	0.0596
	BCa	0.8497	0.5088	0.0551	0.7727	0.4578	0.0638	0.8945	0.5533	0.0580
	Boot-t	0.8239	0.4854	0.0470	0.7233	0.4182	0.0551	0.8687	0.5255	0.0512
	Asymptotic	0.8262	0.5226	0.0565	0.7431	0.4675	0.0686	0.8780	0.5647	0.0590
	Fieller	0.8215	0.4823	0.0464	0.7349	0.4167	0.0501	0.8742	0.5219	0.0475

Power is based on conventional limits for bioequivalency
N is sample size per time point

standard deviation as an estimator for σ for the bootstrap-t and asymptotic intervals. We recommend the use of the Fieller interval if the number of time points is moderate to large since it does not rely on heavy computation like the bootstrap-t-interval and it does not give symmetric confidence intervals as the asymptotic approach. For a small number of time points where computation time is not an issue, the bootstrap-t-interval should be used since it has fewer underlying assumptions and yields almost identical results. The Fieller, asymptotic and bootstrap-t-interval presented in this paper will be implemented in the R-package PK [22] in the near future.

In vast contrast to prior expectation, the ratio method specifically designed for problems of this type, showed poor coverage for small sample sizes and was ultimately indistinguishable from the related hybrid interval in terms of overall coverage. A big difference between these two, however, was to be found when looking at the tail-behavior. The ratio method lead to more balanced tail coverages while the hybrid method severely undercovered on the upper bound. This insight suggests that it might be possible to refine the ratio method further by including a small sample bias correction.

Table 6: Empirical Coverage and Power for a contaminated log-normal distribution using a Nominal Coverage of 90% for varying delta

| | | Delta | | | | | |
| | | 1 | | 0.9 | | 0.799 | |
N	Method	Coverage	Power	Coverage	Power	Coverage	γ
3	Percentile	0.7911 (0.8933;0.8978)	0.9516	0.7916 (0.8941;0.8975)	0.6753	0.7915 (0.8949;0.8966)	0.1028
	Hybrid	0.7925 (0.9117;0.8808)	0.9487	0.7930 (0.9125;0.8805)	0.7152	0.7927 (0.9135;0.8792)	0.1202
	Ratio	0.7888 (0.8935;0.8953)	0.9510	0.7888 (0.8940;0.8948)	0.6729	0.7890 (0.8954;0.8936)	0.1059
	BCa	0.7922 (0.8931;0.8991)	0.9518	0.7916 (0.8933;0.8983)	0.6715	0.7921 (0.8944;0.8977)	0.1021
	Boot-t	0.9066 (0.9524;0.9542)	0.8406	0.9063 (0.9525;0.9538)	0.5040	0.9060 (0.9528;0.9532)	0.0465
	Asymptotic	0.8725 (0.9442;0.9283)	0.8992	0.8724 (0.9446;0.9278)	0.5970	0.8720 (0.9448;0.9272)	0.0728
	Fieller	0.8951 (0.9454;0.9497)	0.8842	0.8947 (0.9458;0.9489)	0.5290	0.8948 (0.9464;0.9484)	0.0513
5	Percentile	0.8430 (0.9192;0.9238)	0.9945	0.8429 (0.9200;0.9229)	0.7935	0.8428 (0.9210;0.9218)	0.0779
	Hybrid	0.8415 (0.9347;0.9068)	0.9938	0.8412 (0.9353;0.9059)	0.8217	0.8412 (0.9364;0.9048)	0.0950
	Ratio	0.8415 (0.9192;0.9223)	0.9945	0.8415 (0.9201;0.9214)	0.7922	0.8412 (0.9208;0.9204)	0.0791
	BCa	0.8415 (0.9163;0.9252)	0.9940	0.8410 (0.9172;0.9238)	0.7892	0.8398 (0.9179;0.9219)	0.0777
	Boot-t	0.8955 (0.9485;0.9470)	0.9862	0.8954 (0.9489;0.9465)	0.7217	0.8957 (0.9497;0.9460)	0.0539
	Asymptotic	0.8851 (0.9481;0.9370)	0.9909	0.8852 (0.9486;0.9366)	0.7637	0.8854 (0.9495;0.9359)	0.0639
	Fieller	0.8957 (0.9470;0.9487)	0.9890	0.8953 (0.9475;0.9478)	0.7280	0.8952 (0.9478;0.9474)	0.0525
10	Percentile	0.8830 (0.9414;0.9416)	1.0000	0.8830 (0.9424;0.9406)	0.9558	0.8832 (0.9436;0.9396)	0.0599
	Hybrid	0.8825 (0.9503;0.9322)	1.0000	0.8826 (0.9510;0.9316)	0.9653	0.8819 (0.9518;0.9301)	0.0693
	Ratio	0.8814 (0.9409;0.9405)	1.0000	0.8817 (0.9419;0.9398)	0.9569	0.8817 (0.9430;0.9387)	0.0609
	BCa	0.8831 (0.9394;0.9437)	1.0000	0.8830 (0.9403;0.9427)	0.9527	0.8824 (0.9412;0.9412)	0.0584
	Boot-t	0.9046 (0.9530;0.9516)	1.0000	0.9043 (0.9535;0.9508)	0.9419	0.9041 (0.9542;0.9499)	0.0499
	Asymptotic	0.9011 (0.9558;0.9453)	1.0000	0.9009 (0.9568;0.9441)	0.9525	0.9011 (0.9576;0.9435)	0.0563
	Fieller	0.9058 (0.9544;0.9514)	1.0000	0.9061 (0.9554;0.9507)	0.9440	0.9058 (0.9561;0.9497)	0.0500

Values in parentheses refer to left and right tail probabilities
Power is based on conventional limits for bioequivalency
N is sample size per time point

Table 7: Two-sided 90% confidence intervals for the ratio of AUCs of dose scaled plasma concentrations of CPI 975 in rats

Method	Lower limit	Upper limit
Percentile	0.7258	1.2081
Hybrid	0.6681	1.1504
Ratio	0.7285	1.2125
BCa	0.7322	1.2215
Boot-t	0.6741	1.2778
Asymptotic	0.6652	1.2110
Fieller	0.6760	1.2839

The most encouraging finding in this note is the great distinction between bioequivalent and non-equivalent results for all intervals. While the intervals show good power to assess bioequivalence when it is present ($\Delta = 1$ and $\Delta = 0.9$), the type-I-error drops to about 5% as soon as the true ratio is less than 0.8.

A natural extension of the work presented here is the use of batch designs. In batch designs each analysis subject is sampled at more than one time point but not at all time points. While a variety of authors such as Holder *et al.* [23] and Yeh

Figure 1: Dose scaled plasma concentrations of CPI 975 at different time points in rats.
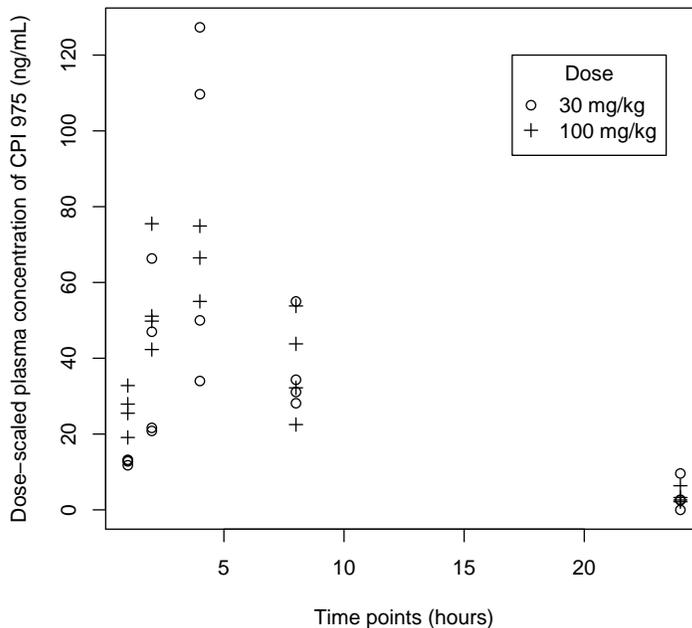
[24] have proposed estimators for the AUC for this sampling design, its theoretical properties are not yet well established. Furthermore, tests for bioequivalence have to be developed and evaluated in this context.

Another point of interest regards observations that fall below the detection limit. Common practice in this situation is to either set those values to half the detection limit or zero. Setting all values below the detection limit to zero may be sufficient for calculation of AUC's but is not an option for other PK parameters as for example when estimating terminal elimination rate where values of zero cannot be used. A different approach to the ad-hoc methods mentioned before is to model non-detected data as censored data. Lambert *et al*. [25] suggest a method in the context of environmental data that should be explored further for medical data in general and the estimation of pharmacokinetic parameters in particular.

References

1 Bailer JA. Testing for the equality of area under the curves when using destructive measurement techniques. *Journal of Pharmacokinetics and Biopharmaceutics*, **16(3)**:303–309, 1988.

2 Tang-Liu DDS, Burke PJ. The effect of azone on ocular levobunolol absorption: Calculating the area under the curve and its standard error using tissue sampling compartments. *Pharmaceutical Research*, **5(4)**:238–241, 1988.

3 Nedelman JR, Gibiansky E, Lau DTW. Applying Bailer's method for AUC confidence intervals to sparse sampling. *Pharmaceutical Research*, **12(1)**:124–128, 1995.

4 Pai SM, Nedelman JR, Hajian G, Gibiansky E, Batra VK. Performance of Bailer's method for AUC confidence intervals from sparse non-normally distributed drug concentrations in toxicokinetic studies. *Pharmaceutical Research*, **13(9)**:1280–1282, 1996.

5 Wolfsegger MJ, Jaki T. Estimation of AUC from 0 to infinity in serial sacrifice designs. *Journal of Pharmacokinetics and Pharamcodynamics*, **32(5-6)**:757–766, 2005.

6 Heinzl H. A note on testing areas under the curve when using destructive measurement techniques. *Journal of Pharmacokinetics and Biopharmaceutics*, **24(6)**:651–655, 1996.

7 Bailer JA, Ruberg SJ. Randomization tests for assessing the equality of area under curves for studies using destructive sampling. *Journal of Applied Toxicology*, **16(5)**:391–395, 1996.

8 Guidance for Industry. *Statistical Approaches to Establishing Bioequivalence*. U.S. Department of Health and Human Services Food and Drug Administration. Center for Drug Evaluation and Research (CDER), 2001.

URL: `http://www.fda.gov/Cder/guidance/3616fnl.pdf`; last visited at 2007-09-25.

9 Hu C, Moore KHP, Kim YH, Sale ME. Statistical issues in a modeling approach to assessing bioequivalence or PK similarity with presence of sparsely sampled subjects. *Journal of Pharmacokinetics and Pharmacodynamics*, **31(4)**:321–339, 2004.

10 Wolfsegger MJ. Establishing bioequivalence in serial sacrifice designs. *Journal of Pharmacokinetics and Pharmacodynamics*, **34(1)**:103–113, 2007.

11 Fieller EC. Some problems in interval estimation. *Journal of the Royal Statistical Society Series B*, **16(2)**:175-185, 1954.

12 Davison AC, Hinkley DV. *Bootstrap Methods and their Applications*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 1997.

13 Efron B, Tibshirani RJ. *An Introduction to the Bootstrap*. Chapman and Hall, London, 1993.

14 Efron B. Better bootstrap confidence intervals (with discussion). *Journal of the American Statistical Association*, **82(397)**:171–200, 1987.

15 Hall P. *The Bootstrap and Edgeworth Expansion*. Springer-Verlag, New York, 1992.

16 Ploner M. *Bootstrap Confidence Intervals, an overview and comparisons by means of selected applications*. Ph. D. thesis, University of Vienna, 2002.

17 Efron B. Nonparametric standard errors and confidence intervals (with discussion). *The Canadian Journal of Statistics*, **9(2)**:139–172, 1981.

18 Satterthwaite FE. An approximate distribution of estimates of variance components. *Biometrics Bulletin*, **2(6)**:110–114, 1946.

19 Gibaldi M, Perrier D. *Pharmacokinetics*. Marcel Dekker, New York and Basel, 1982.

20 R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2006. ISBN 3-900051-07-0.

21 Casella G, Berger RL. *Statistical Inference*. Duxbury Press, 2nd edition edition, 2002.

22 Wolfsegger MJ, Jaki T. *PK: Basic Pharmacokinetics*, 2006. R package Version 0.03.

23 Holder DJ, Hsuan F, Dixit R, Soper K. A method for estimating and testing area under the curve in serial, batch and complete data designs. *Journal of Biopharmaceutical Statistics*, **9(3)**:451–464, 1999.

24 Yeh C. Estimation and significance tests of area under the curve derived from incomplete blood sampling. In *Biopharmaceutical Section of the Proceedings of the Annual Meeting of the American Statistical Association*, 1999.

25 Lambert D, Peterson B, Terpenning I. Nondetects, detection limits, and the probability of detection. *Journal of the American Statistical Association*, **86(414)**:266–277, 1991.