A theoretical framework for estimation of AUCs in complete and incomplete sampling designs

Thomas Jaki
Department of Mathematics and Statistics,
Lancaster University, United Kingdom

Martin J. Wolfsegger
Global Biopharm Preclinical Research and Development,
Baxter AG, Vienna, Austria

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

This is a preprint of an article accepted for publication in
Statistics in Biopharmaceutical Research
http://www.amstat.org/publications/sbr/index.cfm?fuseaction=main

Summary

Nonclinical in vivo animal studies have to be completed before starting clinical studies of the pharmacokinetic behavior of a drug in humans. The drug exposure in animal studies is often measured by the area under the concentration versus time curve (AUC). The classic complete data design, where each animal is sampled for analysis once per time point, is usually only applicable for large animals. In the case of rats and mice, where blood sampling is restricted, the batch design or the serial sampling design needs to be considered. In batch designs samples are taken more than once from each animal, but not at all time points. In serial sampling designs only one sample is taken from each animal. In this paper we present an estimator for the AUC from 0 to the last time point that is applicable to all three designs. The variance and asymptotic distribution of the estimator are derived and confidence intervals based upon the asymptotic results are discussed and evaluated in a simulation study. Further, we define an estimator for linear combinations of AUCs and investigate its asymptotic properties mathematically as well as in simulation.

Keywords: Area under the concentration time curve; AUC; batch design; sparse sampling

1 Introduction

Pharmakocinetic studies aim to understand how an organism reacts to a drug and frequently involves measuring the concentration of the drug in the blood. Under the additive heteroscedastic model the observed concentration for subject i at time t is $Y_{it} = \mu_t + \epsilon_{it}$, where the errors, ϵ_{it}, have mean zero and finite variance, and are independent and identically distributed with continuous distribution G_t. Naturally the concentrations have to

be non-negative and μ_t is the mean drug concentration at time t. The area under the concentration versus time curve (AUC) from 0 to the last observed time point, t_{last}, is one of the pharmacokinetic parameters of interest when measuring the exposure to the drug and is defined as

$$\int_0^{t_{last}} \mu_t dt. \tag{1}$$

Techniques to construct confidence intervals for the AUC from 0 to the last observed time point based on the linear trapezoidal rule for the serial sampling design were presented in Bailer (1988), Tang-Liu and Burke (1988) as well as in Nedelman et al. (1995). Procedures based on other rules to approximate the integral can be found in Nedelman and Gibiansky (1996) and in Gagnon and Peterson (1998). Testing for difference of two AUCs is discussed in Heinzl (1996) and in Bailer and Ruberg (1996) whereas testing for equivalence is addressed by Hu et al. (2004), Wolfsegger (2007) and Jaki et al. (2008). Estimation of the AUC from 0 to infinity in the case of serial sampling is addressed in Wolfsegger and Jaki (2005).

Yeh (1990) extended the work of Tang-Liu & Burke for a special case of the batch design assuming that subsets are sampled at mutually exclusive subsets of times. Nedelman and Jia (1998) extended this work to the general batch design with correlated data. A unified approach for confidence intervals for AUCs that is applicable to the complete data, batch and serial sampling designs was presented in Holder et al. (1999). A method for constructing confidence intervals for the AUC in a batch design using a generalized jackknife estimator for the standard error can be found in Singer and Berger (2003).

In this note we build the theoretical framework for the unified approach of Holder et al. which also justifies the generalized jackknife-based confidence interval by Singer & Berger. Throughout this manuscript focus will be given to the batch design, as the serial sampling design and the complete data design are special cases of it. Section 2 discusses an estimator for the AUC based on the linear trapezoidal rule together with its central moments. We then derive the asymptotic distribution and in a small scale simulation study evaluate confidence intervals based on the asymptotic result. Section 3 presents an estimator for linear combinations of AUCs. The asymptotic distribution is given and assessed through simulation. The report concludes with an illustrative example and a brief discussion.

2 Area under the concentration versus time curve

Consider a study in which measurements are taken at J time points, $\{t_1, \ldots, t_J\}$, let B be the number of batches in the study and let $J_b \subseteq \{1, \ldots, J\}$ be the indices of time points investigated in batch $b = 1, \ldots, B$. We consider the case where a specific time point t_j is only used in one batch, i. e. $J_b \cap J_{b'} = \emptyset \; \forall \; b \neq b'$ and $\bigcup_{b=1}^{B} J_b = \{1, \ldots, J\}$. Every subject belongs to exactly one batch and has measurements at all time points of this particular batch. Note that this construction covers both the serial sampling and the complete data design. For the serial sampling situation, only one time point is included in each batch, while for the complete data design only one batch with all time points can be used.

Since we do not observe a continous recording of the concentrations, Y_{it}, over time, the integral in Equation (1) needs to be approximated. In this note we choose to do so by using the linear trapezoidal rule and consequently will define the AUC to be consistent

with this approximation. For other approximations see for example Gagnon and Peterson (1998). The total AUC from the first to the last time point using the linear trapezoidal rule can therefore be defined as

$$AUC = \sum_{b=1}^{B} \sum_{j \in J_b} w_j \mu_{t_j} = \sum_{j=1}^{J} w_j \mu_{t_j} \tag{2}$$

with weights, w_j, equal to

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

In order to estimate the total AUC and the associated standard error, we introduce the partial AUC of batch b as $pAUC_b = \sum_{j \in J_b} w_j \mu_{t_j}$, for convenience. This partial AUC can be thought of as the contribution of a given batch to the total AUC. It is, however, not the total AUC for this batch, as the weights are computed based on all time points and not just the ones observed for this batch. Although it seldomly is of interest to find the AUC of a given batch due to the small number of observed time points, it could be estimated using the total AUC for only this batch. With n_b observations at each time point in batch b, these AUCs can be estimated by

$$\begin{aligned} \widehat{pAUC_b} &= \frac{1}{n_b} \sum_{i=1}^{n_b} \sum_{j \in J_b} w_j Y_{it_j} \\ \widehat{AUC} &= \sum_{b=1}^{B} \widehat{pAUC_b} = \sum_{b=1}^{B} \frac{1}{n_b} \sum_{i=1}^{n_b} \sum_{j \in J_b} w_j Y_{it_j}. \end{aligned} \tag{3}$$

2.1 Moments

It is easy to see that $E\left[\widehat{AUC}\right] = \sum_{b=1}^{B} \sum_{j \in J_b} w_j \mu_{t_j}$ and hence the estimator of the total AUC is unbiased. To find the variance of the AUC, consider first the variance of a partial

AUC for one subject and denote $Cov(Y_{it_j}, Y_{it_k})$, the covariance of the drug concentration of a subject between time points, as σ_{t_j,t_k}. Then the variance of the partial AUC for the i^{th} subject is

$$V\left[\sum_{j\in J_b} w_j Y_{it_j}\right] = \sum_{j\in J_b}\sum_{k\in J_b} w_j w_k \sigma_{t_j,t_k}$$

and because the subjects within each batch are independent the variance of the partial AUC is

$$V\left[\widehat{pAUC_b}\right] = \frac{1}{n_b}\sum_{j\in J_b}\sum_{k\in J_b} w_j w_k \sigma_{t_j,t_k}.$$

As batches are independent the variance of the total AUC is

$$V\left[\widehat{AUC}\right] = \sum_{b=1}^{B}\frac{1}{n_b}\sum_{j\in J_b}\sum_{k\in J_b} w_j w_k \sigma_{t_j,t_k}. \tag{4}$$

In practice the variance of the AUC needs to be estimated because the covariance is usually unknown.

Lemma 1 *Let σ_{t_j,t_k} be estimated by the sample covariance between times t_j and t_k, then*

$$\widehat{V}(\widehat{AUC}) = \hat{\theta}^2 = \sum_{b=1}^{B}\frac{s_b^2}{n_b} \tag{5}$$

where $s_b^2 = \dfrac{1}{n_b-1}\displaystyle\sum_{i=1}^{n_b}\Big(\sum_{j\in J_b} w_j Y_{it_j} - \frac{1}{n_b}\sum_{k=1}^{n_b}\sum_{j\in J_b} w_j Y_{kt_j}\Big)^2$.

This result implies that the variance of the total AUC can be estimated based on the sample variance of the individual partial AUCs for each batch which is in fact the estimator proposed in Holder et al. (1999). The proof of this lemma is given in the appendix.

2.2 Asymptotic distribution

Lemma 2 *Let $J_b \subseteq \{1,\ldots,J\}$ with $J_b \cap J_{b'} = \emptyset\ \forall\ b \neq b'$ and $\bigcup_{b=1}^{B} J_b = \{1,\ldots,J\}$, and let each batch b be measured on a strictly monotone increasing sequence of fixed time points for*

which the Lindeberg-Levy condition (e.g. Resnick, 1999) holds at every time point. Then as $n_b \to \infty$,

$$\sqrt{n_b}\frac{\widehat{pAUC}_b - pAUC_b}{\vartheta_b} \xrightarrow{d} N(0,1)$$

where $\vartheta_b^2 = \sum_{j\in J_b} \sum_{k\in J_b} w_j w_k \sigma_{t_j,t_k}$. *Let* $n_b = n,\ 1 \le b \le B$, *then as* $n \to \infty$,

$$\frac{\widehat{AUC} - AUC}{\theta} \xrightarrow{d} N(0,1)$$

where

$$\theta^2 = \frac{1}{n}\sum_{b=1}^{B}\vartheta_b^2 = \frac{1}{n}\sum_{b=1}^{B}\sum_{j\in J_b}\sum_{k\in J_b} w_j w_k \sigma_{t_j,t_k}. \tag{6}$$

The proof of the lemma is given in the appendix.

2.3 Confidence Intervals

Three different confidence intervals can be constructed based on the asymptotic result presented above. As defined in Lemma 1, s_b^2 and $\hat{\theta}^2$ denote the estimated asymptotic variances of the partial and total AUC, respectively. The asymptotic confidence interval uses the asymptotic normal distribution to construct a $1-\alpha$ confidence interval as

$$\left[\widehat{AUC} + z_{\frac{\alpha}{2}}\hat{\theta}\,;\widehat{AUC} + z_{1-\frac{\alpha}{2}}\hat{\theta}\right]$$

where z_α is the lower α percentile of the standard normal distribution. In addition to relying on the convergence of the distribution of the estimated AUC to the normal distribution, this interval in theory requires the standard error, θ, to be known. As this is usually not the case, this interval will only have nominal coverage for large sample sizes. A more adequate interval which explicitly takes into account that the standard error is in fact estimated while assuming that the distribution of the AUC is normal can be found by

using the t-distribution as the reference distribution. The interval has the form

$$\left[\widehat{AUC} + t_{\nu,\frac{\alpha}{2}}\hat{\theta}\,;\widehat{AUC} + t_{\nu,1-\frac{\alpha}{2}}\hat{\theta}\right]$$

where $t_{\nu,\alpha}$ denotes the lower α percentile of a t-distribution with ν degrees of freedom. The degrees of freedom can be estimated using Satterthwaite's formula (Satterthwaite, 1946) as

$$\nu = \frac{\left(n\hat{\theta}^2\right)^2}{\sum_{b=1}^{B}\frac{\left(s_b^2\right)^2}{n-1}}.$$

A resampling-based confidence interval is the bootstrap-t-interval (e.g. Davison and Hinkley, 1997). This interval usually works well for location parameters according to Efron and Tibshirani (1993, page 360). We will use the 'star' notation to indicate bootstrap-based estimators. Therefore, $\widehat{AUC}^*(r)$ is the r^{th} bootstrap replication of the estimator $\widehat{AUC}$. The bootstrap-t-interval estimates the distribution of $\frac{\widehat{AUC}-AUC}{\hat{\theta}}$ by the bootstrap equivalent $t^*(r) = \frac{\widehat{AUC}^*(r)-\widehat{AUC}}{\hat{\theta}^*}$. The resulting confidence interval then becomes

$$\left[\widehat{AUC} - t^*_{1-\frac{\alpha}{2}}\hat{\theta}\,;\widehat{AUC} - t^*_{\frac{\alpha}{2}}\hat{\theta}\right]$$

where t^*_α is the lower α percentile of the bootstrap estimates $t^*(r),\ r = 1,\ldots,R$. Notice that this interval requires two separate estimates for the standard error of $\widehat{AUC}$. In addition to the estimate, $\hat{\theta}$, used in the previous two intervals, a second estimate, $\hat{\theta}^*$, is necessary to find $t^*(r)$. To obtain, $\hat{\theta}^*$, usually a jackknife estimator on the bootstrap sample is used which would require a computer intensive additional 'layer' of resampling. Furthermore, this additional resampling 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. We hope to avoid this problem and obtain the desired interval faster by using the asymptotic variance from Equation (6) on the bootstrap sample instead.

2.4 Simulations

A small simulation study was performed using R (R Development Core Team, 2007) to evaluate the intervals described above. In addition to the methods described, the generalized jackknife approach presented in Singer and Berger (2003), was included in the study. 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

$$Y_{it} = \mu_t + \epsilon_{it} = \frac{k_a F X_0}{V\left(k_a - \lambda\right)}\left(e^{-\lambda t} - e^{-k_a t}\right) + \epsilon_{it} \tag{7}$$

with the parameterization $\lambda = 0.0693$, $k_a = 0.231$, $V = 10$, $X_0 = 500$ and $F = 1$ taken from Gibaldi and Perrier (1982, page 440) specified at baseline and ten time points (1h, 2h, 3h, 4h, 6h, 8h, 12h, 18h, 24h and 36h) after drug administration. We define three batches with time points {1,4,12,36}, {2,6,18} and {3,8,24} hours and generated data for each batch with correlations of 0, 0.3, 0.6 and 0.9 to incorporate the dependence of measurements within subjects. For sample sizes of $n_b = n = 3, 5, 10$ and 100 the data were generated such that the coefficient of variation (CV) of Y_{it} is 20% for all time points with mean level μ_t as defined above and follow either a normal or a log-normal distribution. Although in theory a normal distribution could yield negative concentrations, the probability of this happening was negligible (<0.0001) for a CV of 20% under the above model for all time points and consequently was never observed in the simulations.

10000 simulation runs were carried out for each parameter setting with preselected sample sizes. 1000 bootstrap replications were used for the bootstrap-t-confidence interval. For the Singer & Berger generalized jackknife approach, one observation was left out for $n = 3$

and 100 while a leave-2-out approach was used for $n = 5$ and 10. The regular jackknife estimator was used for $n = 3$ as leaving out more than one observations would omit over half of the observations and therefore make the estimator very unstable, while for $n = 100$ the procedure was computationally too expensive.

Tables 1 and 2 give the empirical coverage of the four intervals with nominal level of 90% and the average relative length of the intervals, i.e. the average length of the interval scaled by the true AUC. The t-interval, bootstrap-t and generalize jackknife interval show coverage at nominal level for sample sizes of 5 or larger. Throughout the different settings the generalized jackknife approach, however, appears to have the smallest average interval length among these intervals. In addition it also maintains nominal coverage for $n = 3$ while the t-interval and bootstrap-t-interval are conservative. The purely asymptotic procedure expectedly undercovers for smaller sample sizes and only reaches nominal level for $n = 100$.

Overall the generalized jackknife procedure appears to be superior or on par with the other methods presented. The t-distribution-based interval, however, is a strong alternative for sample sizes of 5 or larger as it is computationally much more efficient than the method by Singer & Berger, which can be excrutiating slow for larger sample sizes.

Table 1: Empirical coverage for normally distributed concentrations with 3 and 4 time points per batch using a nominal coverage of 90%.

		Confidence Interval			
n	ρ	asymptotic	t-interval	bootstrap-t	gen. jackknife
3	0	0.849 (0.1398)	0.923 (0.1822)	0.928 (0.2112)	0.894 (0.1614)
	0.3	0.849 (0.1644)	0.925 (0.2149)	0.925 (0.2484)	0.895 (0.1898)
	0.6	0.849 (0.1855)	0.923 (0.2420)	0.923 (0.2792)	0.894 (0.2142)
	0.9	0.847 (0.2053)	0.921 (0.2681)	0.923 (0.3097)	0.891 (0.2371)
5	0	0.874 (0.1109)	0.906 (0.1232)	0.910 (0.1269)	0.899 (0.1208)
	0.3	0.874 (0.1304)	0.906 (0.1450)	0.907 (0.1492)	0.900 (0.1421)
	0.6	0.876 (0.1468)	0.907 (0.1633)	0.908 (0.1683)	0.902 (0.1599)
	0.9	0.874 (0.1621)	0.908 (0.1803)	0.907 (0.1855)	0.901 (0.1765)
10	0	0.891 (0.0791)	0.903 (0.0824)	0.903 (0.0827)	0.902 (0.0821)
	0.3	0.891 (0.0931)	0.905 (0.0969)	0.902 (0.0973)	0.904 (0.0966)
	0.6	0.884 (0.1052)	0.896 (0.1096)	0.895 (0.1101)	0.896 (0.1092)
	0.9	0.889 (0.1163)	0.901 (0.1211)	0.903 (0.1217)	0.900 (0.1207)
100	0	0.898 (0.0253)	0.899 (0.0254)	0.898 (0.0254)	0.899 (0.0254)
	0.3	0.895 (0.0297)	0.897 (0.0298)	0.895 (0.0298)	0.897 (0.0298)
	0.6	0.897 (0.0335)	0.899 (0.0336)	0.897 (0.0336)	0.899 (0.0336)
	0.9	0.896 (0.0370)	0.897 (0.0371)	0.895 (0.0371)	0.897 (0.0371)

Values in parenthesis are the average interval length relative to the total AUC

n ... sample size per time point

ρ ... correlation between measurements for a subject

Table 2: Empirical coverage for log-normal-distributed concentrations with 3 and 4 time points per batch using a nominal coverage of 90%.

		Confidence Interval			
n	ρ	asymptotic	t-interval	bootstrap-t	gen. jackknife
3	0	0.847 (0.1375)	0.920 (0.1802)	0.925 (0.2101)	0.890 (0.1588)
	0.3	0.852 (0.1617)	0.922 (0.2124)	0.924 (0.2475)	0.895 (0.1867)
	0.6	0.843 (0.1823)	0.918 (0.2391)	0.920 (0.2780)	0.888 (0.2106)
	0.9	0.840 (0.2017)	0.919 (0.2651)	0.920 (0.3107)	0.888 (0.2329)
5	0	0.867 (0.1090)	0.902 (0.1214)	0.906 (0.1258)	0.896 (0.1187)
	0.3	0.868 (0.1285)	0.902 (0.1431)	0.899 (0.1485)	0.896 (0.1400)
	0.6	0.868 (0.1446)	0.903 (0.1612)	0.902 (0.1677)	0.897 (0.1575)
	0.9	0.875 (0.1600)	0.906 (0.1785)	0.905 (0.1859)	0.899 (0.1743)
10	0	0.881 (0.0783)	0.894 (0.0816)	0.895 (0.0824)	0.893 (0.0813)
	0.3	0.889 (0.0918)	0.902 (0.0956)	0.900 (0.0965)	0.900 (0.0952)
	0.6	0.884 (0.1038)	0.897 (0.1081)	0.895 (0.1092)	0.896 (0.1077)
	0.9	0.889 (0.1148)	0.902 (0.1197)	0.901 (0.1211)	0.901 (0.1192)
100	0	0.900 (0.0250)	0.901 (0.0251)	0.899 (0.0251)	0.901 (0.0251)
	0.3	0.893 (0.0293)	0.894 (0.0294)	0.893 (0.0294)	0.894 (0.0294)
	0.6	0.894 (0.0331)	0.895 (0.0332)	0.894 (0.0333)	0.895 (0.0332)
	0.9	0.896 (0.0366)	0.898 (0.0367)	0.897 (0.0367)	0.898 (0.0367)

Values in parenthesis are the average interval length relative to the total AUC

n ... sample size per time point

ρ ... correlation between measurements for a subject

3 Linear combination of AUCs

Evaluating differences between pharmacokinetic parameters of different drugs or doses of one drug is a common problem in pharmacokinetic studies. To do so, this section considers linear combinations of K independent area under the curves. Denote the k^{th} estimated area under the curve as $\widehat{AUC}_k,\ k = 1, \dots K$, with mean AUC_k and asymptotic variance θ_k^2. We allow the sample sizes to be different between each AUC, but they are assumed to be constant within each AUC.

Lemma 3 *Let $c_1, \dots, c_K$ be constants such that $\sum_{k=1}^K c_j = c < \infty$. If the conditions of Lemma 2 are satisfied for $AUC_1, \dots, AUC_K$, then*

$$\frac{\sum_{k=1}^K c_k \widehat{AUC}_k - \sum_{k=1}^K c_k AUC_k}{\tau} \quad \xrightarrow{d} \quad N(0,1) \tag{8}$$

where

$$\tau^2 = \sum_{k=1}^K c_k^2 \theta_k^2. \tag{9}$$

The proof of the lemma is given in the appendix.

3.1 Confidence Intervals

In a similar fashion as described in Section 2.3 the asymptotic result can be used to form $1-\alpha$ confidence intervals for linear contrasts of AUCs. The normal theory asymptotic interval can be defined as

$$\left[\sum_{k=1}^K c_k \widehat{AUC}_k + z_{\frac{\alpha}{2}}\hat{\tau}; \sum_{k=1}^K c_k \widehat{AUC}_k + z_{1-\frac{\alpha}{2}}\hat{\tau}\right]$$

while the t-distribution based interval has the form

$$\left[\sum_{k=1}^K c_k \widehat{AUC}_k + t_{\eta,\frac{\alpha}{2}}\hat{\tau}; \sum_{k=1}^K c_k \widehat{AUC}_k + t_{\eta,1-\frac{\alpha}{2}}\hat{\tau}\right]$$

with degrees of freedom $\eta = \sum_{k=1}^{K} \nu_k$.

The bootstrap-t-interval can in a similar fashion be found as

$$\left[\sum_{k=1}^{K} c_k \widehat{AUC}_k - t^*_{1-\frac{\alpha}{2}}\hat{\tau}; \sum_{k=1}^{K} c_k \widehat{AUC}_k - t^*_{\frac{\alpha}{2}}\hat{\tau}\right]$$

where $t^*(r) = \dfrac{\left(\sum_{k=1}^{K} c_k \widehat{AUC}_k\right)^*(r) - \sum_{k=1}^{K} c_k \widehat{AUC}_k}{\tau^*}$.

Note that p-values can also easily be obtained using the asymptotic variance of the contrasts, which are usually more useful when correcting for multiple comparisons. We can find the p-value for the 2-sides hypothesis as $2P(T_\eta > |t|)$ because the distribution of the statistic $t = \frac{\sum_{k=1}^{K} c_k \widehat{AUC}_k - Q}{\hat{\tau}}$, where Q is the value under the null hypothesis, can be approximated by a t-distribution with η degrees of freedom.

3.2 Simulations

We present simulation results for the empirical coverage and power for the difference of two AUCs in this section in order to evaluate the confidence intervals for the linear combination of AUCs. Once more 10000 Monte Carlo samples were obtained and 1000 bootstrap resamples used for the bootstrap-t-interval. Both AUCs are generated under the model in equation (7) with one having a 10% higher mean than the other at all time points with the same within-subject correlations varying between 0 and 0.9. Table 3 gives the results for normal distributed concentrations when the coefficient of variation is constant at 20% for both AUCs while in Table 4 it is 20% for the lower and 40% for the higher AUC. The same setup was also used with log-normal distributed concentrations. This situation, however, yielded almost identical results and hence these are not presented in this note.

Table 3: Empirical coverage and power for normally distributed concentrations with 3 and 4 time points per batch using a nominal coverage of 90% for comparing two AUCs with a 10% mean difference and identical coefficients of variation.

		Confidence Interval		
n	ρ	asymptotic	t-interval	bootstrap-t
3	0	0.872 (0.471)	0.913 (0.393)	0.920 (0.363)
	0.3	0.874 (0.392)	0.909 (0.321)	0.920 (0.298)
	0.6	0.869 (0.333)	0.907 (0.265)	0.917 (0.244)
	0.9	0.876 (0.294)	0.914 (0.232)	0.920 (0.212)
5	0	0.885 (0.622)	0.905 (0.587)	0.905 (0.583)
	0.3	0.886 (0.519)	0.903 (0.486)	0.904 (0.481)
	0.6	0.887 (0.457)	0.903 (0.425)	0.905 (0.416)
	0.9	0.884 (0.389)	0.900 (0.355)	0.902 (0.353)
10	0	0.892 (0.864)	0.898 (0.856)	0.897 (0.855)
	0.3	0.887 (0.761)	0.894 (0.751)	0.893 (0.751)
	0.6	0.896 (0.669)	0.902 (0.657)	0.901 (0.655)
	0.9	0.896 (0.594)	0.903 (0.580)	0.903 (0.583)
100	0	0.901 (1.000)	0.902 (1.000)	0.901 (1.000)
	0.3	0.899 (1.000)	0.899 (1.000)	0.898 (1.000)
	0.6	0.903 (1.000)	0.904 (1.000)	0.905 (1.000)
	0.9	0.903 (1.000)	0.904 (1.000)	0.902 (1.000)

Values in parenthesis are the empirical power
n ... sample size per time point
ρ ... correlation between measurements for a subject

For the situation of equal coefficient of variation, both the t-distribution-based interval and the bootstrap-t-interval are conservative for three subjects per time point but reach nominal coverage for larger sample sizes. The purely asymptotic interval, however, once more only reaches nominal coverage for $n = 100$ per time point. The estimated power also shows the expected behavior as it increases with sample size. Additionally the power decreases as the correlation within batches increases, dropping about 40% between independence and strong correlation within subjects.

The results for unequal coefficients of variation again show that for sample sizes of

Table 4: Empirical coverage and power for normally distributed concentrations with 3 and 4 time points per batch using a nominal coverage of 90% for comparing two AUCs with a 10% mean difference and coefficients of variation of 20% and 40%.

		Confidence Interval		
n	ρ	asymptotic	t-interval	bootstrap-t
3	0	0.860 (0.275)	0.901 (0.218)	0.915 (0.189)
	0.3	0.864 (0.237)	0.899 (0.184)	0.916 (0.155)
	0.6	0.857 (0.212)	0.895 (0.165)	0.914 (0.137)
	0.9	0.861 (0.195)	0.900 (0.153)	0.917 (0.125)
5	0	0.882 (0.339)	0.899 (0.310)	0.904 (0.298)
	0.3	0.878 (0.282)	0.895 (0.254)	0.902 (0.239)
	0.6	0.882 (0.246)	0.898 (0.222)	0.905 (0.209)
	0.9	0.870 (0.214)	0.887 (0.189)	0.894 (0.180)
10	0	0.894 (0.530)	0.901 (0.516)	0.900 (0.511)
	0.3	0.891 (0.404)	0.897 (0.392)	0.897 (0.383)
	0.6	0.895 (0.346)	0.901 (0.334)	0.904 (0.327)
	0.9	0.896 (0.300)	0.901 (0.289)	0.905 (0.281)
100	0	0.900 (1.000)	0.901 (1.000)	0.899 (1.000)
	0.3	0.898 (0.998)	0.899 (0.998)	0.897 (0.998)
	0.6	0.903 (0.988)	0.904 (0.988)	0.901 (0.988)
	0.9	0.902 (0.971)	0.902 (0.971)	0.902 (0.970)

Values in parenthesis are the empirical power
n ... sample size per time point
ρ ... correlation between measurements for a subject

5 or above the coverage is on par for the t-interval as well as the bootstrap-t-interval. For $n = 3$, however, a surprising difference can be seen as the bootstrap-t-interval once more is conservative, while the t-interval maintains nominal coverage. The power in this scenario shows the same overall patterns as before. The power increases with sample size and decreases with increased correlation. The added variability in this situation is also reflected in the power which generally is drastically lower than in the equal CV case.

4 Example

In this example we use the dataset in Holder et al. (1999) which investigates the plasma levels of a single-oral-dose toxicokinetic study at 6 different dose levels (100, 300, 450, 600,

750 and 1000 mg/kg). Each dose level is measured using 3 batches with time points {1, 6}, {2, 10} and {4, 24} hours for 3 female rats in each batch. Holder et al. show in detail how the AUCs and corresponding standard errors for the dose groups can be obtained.

Our focus will be on identifying the highest dose for which we cannot reject the null hypothesis of dose proportionality, which can be formulated as a sequence of hypothesis testing problems. On the assumption of dose proportionality and using the power law model (Wixley, 1997), $\frac{\text{AUC}}{\text{dose}} = \mu$ holds for every dose irrespective of equally or non-equally spaced doses. Let $\mu_1, \ldots, \mu_k$ be the dose-normalized AUCs of the $k = 6$ increasing dose levels investigated. The alternative hypothesis of a saturable absorption leads to the following sequence of hypothesis

$$H_{0i} : \mu_1 = \ldots = \mu_i \quad vs. \quad H_{1i} : \mu_1 = \ldots = \mu_{i-1} < \mu_i \quad (2 \leq i \leq k)$$

whereas the alternative hypothesis of a saturable metabolism leads to

$$H_{0i} : \mu_1 = \ldots = \mu_i \quad vs. \quad H_{1i} : \mu_1 = \ldots = \mu_{i-1} > \mu_i \quad (2 \leq i \leq k)\,.$$

Testing such sequences of hypothesis with control of the type I error in a strong sense can be archieved by application of the closure principle (Marcus et al., 1976). Ruberg (1989) and Tamhane et al. (1996) address the closure principle based on contrasts in great detail. We used a step-down approach based on reverse helmert contrasts as this contrast type has good power properties for rejection of H_{0i} $(2 \leq i \leq k)$ for low dose levels. We think that the risk of not detecting a low non-proportional dose is more crucial than the risk of not detecting a high non-proportional dose. The coefficients of these contrasts are given in Table 5.

Table 5: Coefficients for reverse Helmert contrasts.

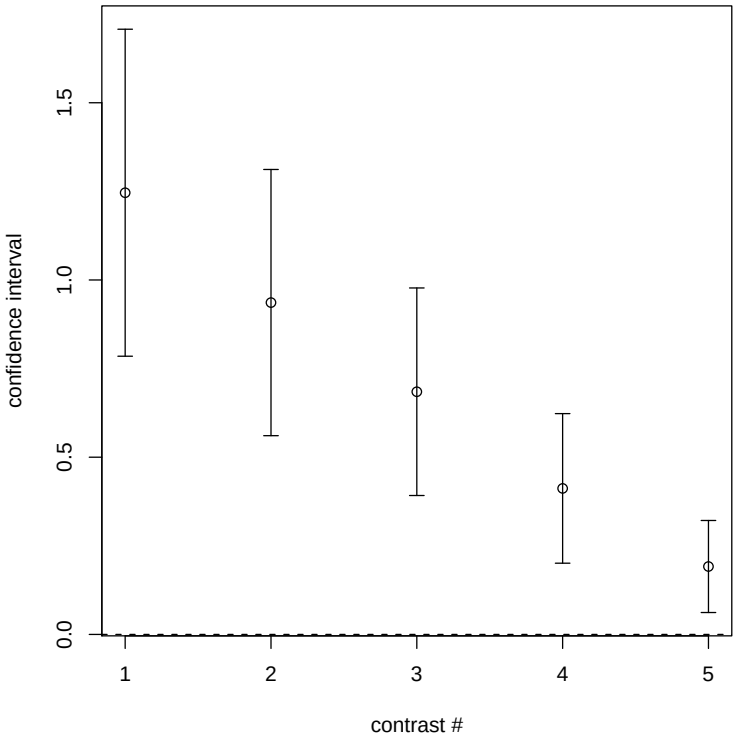
Contrast	Dose					
	100mg/kg	300mg/kg	450mg/kg	600mg/kg	750mg/kg	1000mg/kg
	c_1	c_2	c_3	c_4	c_5	c_6
1	5	-1	-1	-1	-1	-1
2	4	-1	-1	-1	-1	0
3	3	-1	-1	-1	0	0
4	2	-1	-1	0	0	0
5	1	-1	0	0	0	0

Using the step-down approach, the hypothesis are tested sequentially from H_{0k} to H_{02} at a pre-specified α-level until we fail to reject for the first time. We used two-sided p-values to account for both alternatives at H_{0k}. For this reason the procedure also has to be stopped in the case of rejection of H_{0k} and subsequent rejection of H_{0i} $(i < k)$ when the corresponding t-statistic has a different direction than the t-statistic of hypothesis H_{0k}. Table 6 provides the estimates, standard error and degrees of freedom for each contrast as well as the corresponding p-values. Two-sided 95% confidence intervals for the contrasts are depicted in Figure 1. The null hypothesis of dose proportionality is rejected for the dose range investigated against the alternative of a saturable absorption as can be seen from the graph as well as the p-values.

Table 6: Summary of reverse Helmert contrasts with two-sided p-values.

Hypothesis	Contrast	Estimate	SE	df	p-value
H_{06}	1	1.2462	0.2149	13.8799	0.0000
H_{05}	2	0.9362	0.1721	11.8656	0.0002
H_{04}	3	0.6847	0.1307	9.6044	0.0004
H_{03}	4	0.4120	0.0901	7.3462	0.0023
H_{02}	5	0.1917	0.0504	4.9721	0.0127

Figure 1: Two-sided 95% confidence intervals for the reverse Helmert contrasts.



5 Discussion

In this note we provide the theoretical framework of the estimator for the area under the concentration curve in batch designs proposed by Holder et al. (1999). The asymptotic distribution of the estimator is derived and used to form asymptotic confidence intervals. These intervals are evaluated in a simulation study against the generalized jackknife approach, a method that also relies on the asymptotic result, which is presented in Singer and Berger (2003). The simulations show that the generalized jackknife is superior or equivalent to the presented methods for all situations considered. The t-distribution-based interval, however, is a viable and fast alternative for sample sizes of 5 or larger per time point.

In the second part of this manuscript the results are extended to linear combinations of area under the concentration curves in batch designs. Confidence intervals based on asymptotic theory are derived and evaluated in a simulation study in terms of coverage and power. The t-distribution-based interval and the bootstrap-t-interval showed very similar properties throughout the sitations considered. The only exception was found for three subjects per time point, where, in the case of different coefficients of variation, the t-interval had nominal coverage while the bootstrap-based interval was conservative. We therefore recommend the use of the t-interval for linear combinations of AUCs in the batch design.

We have shown that the sample size between batches can be different, while it was necessary for the derivation of the asymptotic distribution for the sample sizes to be equal within batches. Future work will try to relax this assumption and investigate methods for handling missing data in batch designs. Another, related, point of interest are observations that fall below the detection limit. Common practice in this situation is to either set those values to half the detection limit or to zero. A different approach to these ad-hoc methods is to model non-detected data as censored data. Lambert et al. (1991) suggest a method in the context of environmental data that should be explored further for medical data in general and for the estimation of pharmacokinetic parameters in particular.

A Appendix

Proof of Lemma 1: The estimated asymptotic variance can be written as

$$\widehat{V}\left[\widehat{AUC}\right] = \sum_{b=1}^{B}\frac{1}{n_b}\sum_{j\in J_b}\sum_{k\in J_b} w_j w_k \hat{\sigma}_{t_j,t_k} = \sum_{b=1}^{B}\frac{\hat{\xi}_b^2}{n_b}$$

and it therefore suffices to show that $\hat{\xi}_b^2 = s_b^2$. Using the sample covariance estimate,

$$\hat{\sigma}_{t_j,t_k} = \frac{1}{n_b-1}\sum_{i=1}^{n_b} Y_{it_j}Y_{it_k} - \frac{1}{n_b(n_b-1)}\sum_{i=1}^{n_b} Y_{it_j}\sum_{l=1}^{n_b} Y_{lt_k}$$

we get

$$\begin{aligned}
\hat{\xi}_b^2 &= \sum_{j\in J_b}\sum_{k\in J_b} w_{t_j}w_{t_k}\Big(\frac{1}{n_b-1}\sum_{i=1}^{n_b} Y_{it_j}Y_{it_k} - \frac{1}{n_b(n_b-1)}\sum_{i=1}^{n_b} Y_{it_j}\sum_{l=1}^{n_b} Y_{lt_k}\Big) \\
&= \frac{1}{n_b-1}\sum_{i=1}^{n_b}\Big(\sum_{j\in J_b} w_j Y_{it_j}\Big)^2 - \frac{1}{n_b(n_b-1)}\Big(\sum_{i=1}^{n_b}\sum_{j\in J_b} w_j Y_{it_j}\Big)^2 \\
&= s_b^2.
\end{aligned}$$

⊠

Proof of Lemma 2: Because the partial area under the curves are independent and identically distributed, the central limit theorem (e.g. Resnick, 1999) gives that

$$\sqrt{n_b}\frac{\widehat{pAUC}_b - pAUC_b}{\vartheta_b} \xrightarrow{d} N(0,1)$$

where

$$\vartheta_b^2 = \sum_{j\in J_b}\sum_{k\in J_b} w_j w_k \sigma_{t_j,t_k}$$

if $\vartheta_b^2 < \infty$. Since we assumed that $\sigma_{t_j}^2 < \infty$ the covariances are also finite and therefore $\vartheta_b^2 < \infty$ in fact holds.

Let $n_b = n$ and $\mathbf{Z}_n = (\widehat{pAUC}_1, \widehat{pAUC}_2, \ldots, \widehat{pAUC}_B)$ which are pairwise independent because time points are only used once. Let $\boldsymbol{\mu} = E[\mathbf{Z}_n]$ and $V[\mathbf{Z}_n] = \frac{1}{n}\boldsymbol{\Sigma} =$

$\frac{1}{n}\begin{pmatrix} \vartheta_1^2 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \vartheta_B^2 \end{pmatrix}$, by the first result and independence, $\sqrt{n}\left(\mathbf{Z}_n - \boldsymbol{\mu}\right) \xrightarrow{d} MVN(\mathbf{0}, \boldsymbol{\Sigma})$. The total AUC, $\sum_{b=1}^{B} \widehat{pAUC}_b$, then is a linear statistic in $\mathbf{Z}_n$ and therefore also converges in distribution (e.g. Serfling, 1980), that is

$$\frac{\widehat{AUC} - AUC}{\theta} \xrightarrow{d} N(0,1)$$

where

$$\theta^2 = \frac{1}{n}\sum_{b=1}^{B}\sum_{j\in J_b}\sum_{k\in J_b} w_j w_k \sigma_{t_j,t_k}.$$

⊠

Proof of Lemma 3: From Lemma 2 we know that each total AUC converges in distribution and because the AUCs are independent from each other we also get

$$\frac{\begin{pmatrix} \widehat{AUC}_1 \\ \vdots \\ \widehat{AUC}_K \end{pmatrix} - \begin{pmatrix} AUC_1 \\ \vdots \\ AUC_K \end{pmatrix}}{\begin{pmatrix} \theta_1^2 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \theta_K^2 \end{pmatrix}} \xrightarrow{d} MVN(\mathbf{0}, \boldsymbol{I})$$

if $n_k \to \infty$. $\sum_{k=1}^{K} c_k AUC_k$, however, is again a linear statistic and therefore,

$$\frac{\sum_{k=1}^{K} c_k \widehat{AUC}_k - \sum_{k=1}^{K} c_k AUC_k)}{\tau} \xrightarrow{d} N(0,1)$$

with

$$\tau^2 = \sum_{k=1}^{K} c_k^2 \theta_k^2.$$

⊠

References

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

Bailer, A. J. and Ruberg, S. J. (1996), "Randomization tests for assessing the equality of area under curves for studies using destructive sampling," *Journal of applied Toxicology*, 16, 391–395.

Davison, A. C. and Hinkley, D. V. (1997), *Bootstrap Methods and their Applications*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press.

Efron, B. and Tibshirani, R. J. (1993), *An Introduction to the Bootstrap*, Chapman and Hall, London.

Gagnon, R. C. and Peterson, J. J. (1998), "Estimation of confidence intervals for area under the curve from destructively obtained pharmacokinetic data," *Journal of Pharmacokinetics and Biopharmaceutics*, 26, 87–102.

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

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

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

Hu, C., Moore, K. H. P., Kim, Y. H., and Sale, M. E. (2004), "Statistical issues in a modeling approach to assessing bioequivalence or PK similarity with presence of sparsely sampled subjects," *Journal of Pharmacokinetics and Pharmacodynamics*, 31, 321–339.

Jaki, T., Wolfsegger, M. J., and Ploner, M. (2008), "Confidence intervals for ratios of AUCs in the case of serial sampling: A comparison of seven methods," *Pharmaceutical Statistics*, published online April 14 2008, DOI: 10.1002/pst.321.

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

Marcus, R., Peritz, E., and Gabriel, K. R. (1976), "On closed testing procedures with special reference to ordered analysis of variance," *Biometrika*, 63, 655–660.

Nedelman, J. R. and Gibiansky, E. (1996), "The variance of a better AUC estimator for sparse, destructive sampling in toxicokinetics," *Journal of Pharmaceutical Sciences*, 85, 884–886.

Nedelman, J. R., Gibiansky, E., and Lau, D. T. W. (1995), "Applying Bailer's method for AUC confidence intervals to sparse sampling," *Pharmaceutical Research*, 12, 124–128.

Nedelman, J. R. and Jia, X. (1998), "An Extension of Satterthawaite's approximation applied to pharmacokinetics," *Journal of Biopharmaceutical Statistics*, 8, 317–328.

R Development Core Team (2007), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0.

Resnick, S. I. (1999), *A Probability Path*, Birkhäuser.

Ruberg, S. J. (1989), "Contrasts for Identifying the Minimum Effective Dose," *Journal of the American Statistical Association*, 84, 816–822.

Satterthwaite, F. E. (1946), "An approximate distribution of estimates of variance components," *Biometrics Bulletin*, 2, 110–114.

Serfling, R. J. (1980), *Approximation Theorems of Mathematical Statistics*, Wiley Series in Probability and Statistics, John Wiley & Sons, Inc.

Singer, J. and Berger, M. (2003), "The Jackknife applied to incomplete blood sampling models," *Pharmaceutical Statistics*, 2, 161–166.

Tamhane, A. C., Hochberg, Y., and Dunnett, C. W. (1996), "Multiple Test Procedures for Dose Finding," *Biometrics*, 52, 21–37.

Tang-Liu, D. D.-S. and Burke, P. J. (1988), "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, 238–241.

Wixley, R. A. J. (1997), "Testing AUC dose-proportionality in destructive toxicokinetics," *Drug Information Journal*, 31, 529–549.

Wolfsegger, M. J. (2007), "Establishing bioequivalence in serial sacrifice designs," *Journal of Pharmacokinetics and Pharmacodynamics*, 34, 103–113.

Wolfsegger, M. J. and Jaki, T. (2005), "Estimation of AUC from 0 to infinity in serial sacrifice designs," *Journal of Pharmacokinetics and Pharmacodynamics*, 32, 757–766.

Yeh, C. (1990), "Estimation and significant tests of area under the curve derived from

incomplete blood sampling," *American Statistical Association Proceedings of the Biopharmaceutical Section*, 74–81.