Journal of Pharmacokinetics and Biopharmaceutics, Vol. 20, No. 3, 1992

Optimum Numerical Integration Methods for Estimation of Area-Under-the-Curve (*AUC*) and Area-Under-the-Moment-Curve (*AUMC*)

Robert D. Purves[1]

Received July 11, 1991—Final March 10, 1992

Eleven numerical methods for estimation of AUC *(including 4 new methods) and 22 methods for* AUMC *(including 8 new methods) were tested on large simulated noisy datasets representing bolus, oral, and infusion concentration-time profiles. Some methods were unacceptable because their mean error was large; these included a commonly recommended form of the linear trapezoidal rule for* AUMC. *Others, notably Lagrange and cubic spline methods, were unacceptable because the variance of their estimates was large. These methods should be abandoned. A simple and easily programmed new method, parabolas-through-the-origin then log-trapezoidal rule, performed especially well.*

KEY WORDS: area-under-the-curve; noncompartmental analysis; numerical integration.

INTRODUCTION

The area-under-the-curve (AUC) and area-under-the-moment-curve ($AUMC$) of a set of concentration-time measurements occupy a central role in noncompartmental pharmacokinetics (1). These quantities may be estimated from experimental data by fitting a function $C(t)$ based on a compartmental model, and then integrating $C(t)$ and $tC(t)$ analytically (1). Alternatively, and more in keeping with the spirit of noncompartmental analysis, they may be estimated by direct numerical integration of the data (1,2). In spite of the importance of AUC and $AUMC$, there has been little work to determine optimum numerical methods for their estimation, other than the early study by Yeh and Kwan (2) which dealt only with AUC, and a recent small-scale comparison by Yeh and Small (3).

[1]Department of Pharmacology, Medical School, University of Otago, P.O. Box 913, Dunedin, New Zealand.

0090-466X/92/0600-0211$06.50/0

The purpose of this paper is provide a comprehensive survey of the integration methods currently in use, and to compare them with some newly devised methods. The comparison is made by using each method on a large number of simulated data sets representing commonly observed types (iv bolus, oral, and infusion). Each data set has superimposed random "noise" to mimic assay and other variability.

INTEGRATION METHODS FOR *AUC*

If concentration values C_i are measured at times t_i ($i = 1 \ldots n$), numerical integration methods may be written as a summation over $n-1$ intervals. The notation in the following is that the first sample point is at time $t_1 = 0$.

$$AUC_0^{t_n} = \int_0^{t_n} C(t)\,dt \approx \sum_{i=1}^{n-1} \int_{t_i}^{t_i+1} g(t, a_i, b_i \ldots)\,dt \tag{1}$$

where g is an integrable interpolating function whose parameters (a_i, $b_i \ldots$) are chosen so that g passes exactly through C_i and C_{i+1} (and, in some methods, additional values of C exterior to the interval of integration). Eleven methods for *AUC* were investigated.

In Method 1, $g = a + bt$. After performing the integration indicated in Eq. (1) we have the familiar (linear) trapezoidal rule

$$AUC_0^{t_n} \approx \sum_{i=1}^{n-1} \tfrac{1}{2}(t_{i+1} - t_i)(C_i + C_{i+1}) \tag{2}$$

In Method 2, $g = a \exp[-bt]$. This choice gives rise to the log trapezoidal rule (2)

$$AUC_0^{t_n} \approx \sum_{i=1}^{n-1} \frac{(t_{i+1} - t_i)(C_{i+1} - C_i)}{\log[C_{i+1}/C_i]} \tag{3}$$

If $C_i = 0$ or $C_{i+1} = 0$ or $C_i = C_{i+1}$ a linear trapezoidal step is taken instead, by replacing summand$_i$ of Eq. (3) by summand$_i$ of Eq. (2). In the course of this study, Eq. (3) was found to generate excessive round-off error if $C_i \approx C_{i+1}$, and so a linear trapezoidal step was taken if $0.999 < C_{i+1}/C_i < 1.001$. This precaution is perhaps unnecessary in the application of Eq. (3) to experimental data rounded off to two or three significant digits, but should always be included in simulation studies. The log trapezoidal rule (with linear step if required) forms part of many hybrid methods described later.

In Method 3, the cubic spline method (2), g is a cubic polynomial whose coefficients are determined by a simultaneous fit to all n points, in such a way as to produce smooth curvilinear interpolation across the whole

data set. Numerical integration is preceded by a call to a spline-fitting routine (4) which returns values of the n second derivatives C_i''. Then

$$AUC_0^{t_n} \approx \sum_{i=1}^{n-1} \left\{ \frac{h}{2}(C_i + C_{i+1}) - \frac{h^3}{24}(C_i'' + C_{i+1}'') \right\} \tag{4}$$

where $h = t_{i+1} - t_i$.

In Method 4, the so-called Lagrange method, g is a cubic polynomial

$$g = a + bt + ct^2 + dt^3 \tag{5}$$

for intervals 2 to $n-2$, and a quadratic for intervals 1 and $n-1$. The coefficients of the cubic polynomial are determined by the requirement that g also pass through C_{i-1} and C_{i+1}. Computational details are given in refs. (2) and (5).

In Method 5, g is a piecewise cubic polynomial with various constraints applied to improve stability. Details of this approach have been given by Yeh and Small (3).

Method 6 is the well-known hybrid in which the trapezoidal rule is applied up to the time t_{max} of the maximum observed concentration C_{max}, and then the log trapezoidal rule is applied for all remaining intervals. If $t_{max} = t_1$, as in most bolus data, the trapezoidal rule is not used at all.

Method 7 is a similar hybrid (2) in which the spline method is used up to t_{max}. If $t_{max} = t_1$, the spline method is not used at all.

New Methods

Method 8 is a trivial modification of Method 6. The summand of Eq. (2) (linear trapezoidal) is used if $C_{i+1} \geq C_i$, otherwise the summand of Eq. (3) (log trapezoidal) is used. This avoids the need to precompute t_{max}.

The interpolating function in the trapezoidal method has zero curvature, and that in the log trapezoidal method has positive curvature; neither seems well adapted for regions of $C(t)$ with negative curvature, such as the approach to C_{max} after an oral dose or during an infusion. Of the limitless range of two-parameter functions I have chosen three for investigation, all of which have the special feature of passing through the origin (Fig. 1). None of these methods should be applied directly to data (such as bolus) for which $C_1 > 0$; instead a log trapezoidal step is taken for the first interval, and (in Methods 9 and 11) all subsequent intervals. In forming the summations, the first interval requires separate treatment, because g is already constrained to pass through the origin so that the value of $C_1 = 0$ at t_1 cannot determine parameters a_1 and b_1. Methods 9 and 11 are used only up to t_p (upper limit of summation i_p) where

$$\left.\begin{array}{l} t_p = t_{max} \text{ and } i_p \text{ is the index to } t_{max} \text{ and } C_{max} \\ \text{unless } t_{max} = t_2, \text{ in which case } t_p = t_3,\ i_p = 3. \end{array}\right\} \tag{6}$$

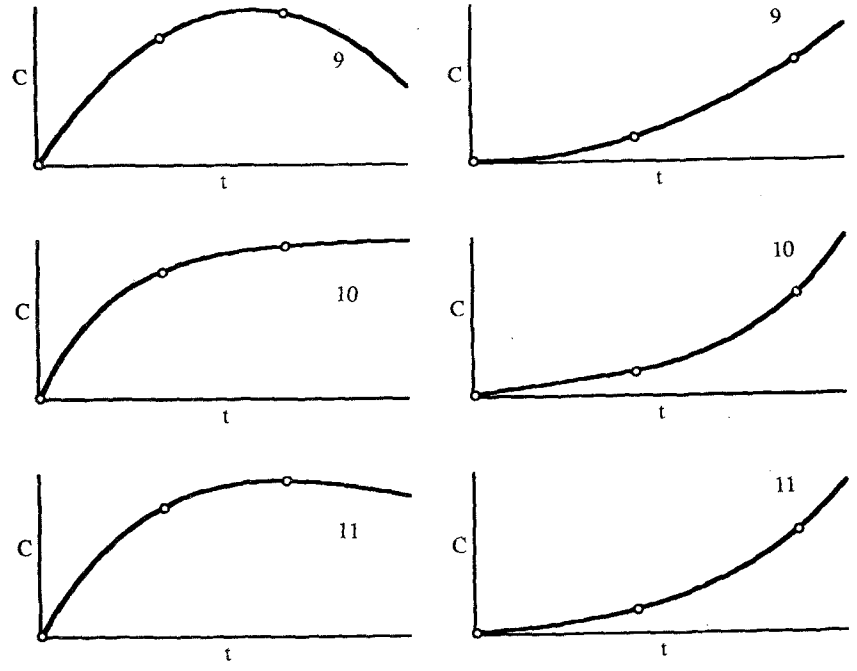

Fig. 1. Interpolation functions used in Methods 9, 10, and 11. Left-hand graphs show interpolation with negative curvature. Right-hand graphs show interpolation with positive curvature. Note that all functions pass through the origin. Numbers against curves indicate the method. For interpolation with positive curvature by Method 10, the first interval is interpolated linearly (as in Method 1), and the second interval is interpolated as in the log trapezoidal method (Method 2).

This definition of t_p is required because the case $i_p = 2$ is forbidden: Methods 9 and 11 use C_2 and C_3 to determine a_2 and b_2 for area estimates in the first two intervals. The definition allows the methods to "go over the top" of the curve if $C_{max} = C_2$ (as could occur, for example, with exceptionally poorly sampled oral data).

In Method 9 the interpolating function is a parabola-through-the-origin (PTTO): $g = at^2 + bt$.

$$AUC_0^{t_p} \approx \int_0^{t_2} (a_2 t^2 + b_2 t)\, dt + \sum_{i=2}^{i_p-1} \int_{t_i}^{t_{i+1}} (a_i t^2 + b_i t)\, dt \approx \frac{a_2 t_2^3}{3} + \frac{b_2 t_2^2}{2} + \sum_{i=2}^{i_p-1} \left\{ \frac{a_i}{3}(t_{i+1}^3 - t_i^3) + \frac{b_i}{2}(t_{i+1}^2 - t_i^2) \right\} \quad (7)$$

where

$$a_i = \frac{1}{t_{i+1} - t_i} \left(\frac{C_{i+1}}{t_{i+1}} - \frac{C_i}{t_i} \right) \quad \text{and} \quad b_i = \frac{C_i}{t_i} - a_i t_i.$$

Equation (7) is applied up to t_p and the log trapezoidal method [summand of Eq. (3)] up to t_n. A computer listing is given in the Appendix.

In Method 10 the interpolating function is the "plateau function," $g = a(1-\exp[-bt])$.

$$AUC_0^{t_n} \approx \int_0^{t_2} a_2(1-\exp[-b_2 t])\,dt + \sum_{i=2}^{n-1} \int_{t_i}^{t_{i+1}} a_i(1-\exp[-b_i t])\,dt$$

$$\approx a_2 t_2 + \frac{a_2}{b_2}\exp[-b_2 t_2] - \frac{a_2}{b_2} + \sum_{i=2}^{n-1}\left\{a_i(t_{i+1}-t_i) + \frac{a_i}{b_i}(\exp[-b_i t_{i+1}] - \exp[-b_i t_i])\right\} \tag{8}$$

Parameter b_i is found numerically as the root of the nonlinear equation

$$f(b_i) = C_i \frac{1-\exp[-b_i t_{i+1}]}{1-\exp[-b_i t_i]} - C_{i+1} \tag{9}$$

The secant method (4) is suitable, with starting values $1/t_i$, $1/t_{i+1}$ and a convergence criterion of $|f(b_i)| < 10^{-5} C_{i+1}$. Once b_i is known, $a_i = C_i/(1-\exp[-b_i t_i])$. If $C_{i+1} < C_i$ or $C_i t_{i+1} < 1.01 C_{i+1} t_i$, iteration of Eq. (9) does not converge; the summand of Eq. (8) cannot be calculated and is replaced by the summand of Eq. (3) (a log trapezoidal step).

In Method 11 (α function then log trapezoidal) the interpolating function is $g = at\exp[-bt]$. The name "α function" derives from the form $\alpha^2 t \exp[-\alpha t]$ used, in a different context, by Jack *et al.* (6).

$$AUC_0^{t_p} \approx \int_0^{t_2} a_2 t \exp[-b_2 t]\,dt + \sum_{i=2}^{i_p-1} \int_{t_i}^{t_{i+1}} a_i t \exp[-b_i t]\,dt$$

$$\approx \frac{C_2}{b_2}\left(\frac{\exp[b_2 t_2]-1}{t_2 b_2} - 1\right) + \sum_{i=2}^{i_p-1} \frac{C_i}{b_i}\left\{1 + \frac{1}{b_i t_i} - \frac{1}{z_i}\left(\frac{t_{i+1}}{t_i} + \frac{1}{b_i t_i}\right)\right\} \tag{10}$$

where $z_i = C_i t_{i+1}/C_{i+1} t_i$ and $b_i = \ln z_i/(t_{i+1}-t_i)$. If $z_i = 0$ or $C_{i+1} = 0$ a linear trapezoidal step [summand of Eq. (2)] is taken. A linear trapezoidal step is also taken if $0.99 < z_i < 1.01$, to avoid serious round-off error if the rate constant b_i is near zero. Equation (10) is applied up to t_p [Eq. (6)] and the log trapezoidal method [summand of Eq. (3)] up to t_n.

INTEGRATION METHODS FOR *AUMC*

It seems not to be widely known that there are two classes of methods for estimating *AUMC*. First, one may write

$$AUMC_0^{t_n} = \int_0^{t_n} tC(t)\,dt \approx \sum_{i=1}^{n-1} \int_{t_i}^{t_{i+1}} tg(t, a_i, b_i \ldots)\,dt \tag{11}$$

where g interpolates between C_i and C_{i+1}, and the parameters $(a_i, b_i \ldots)$ are to be found exactly as in the *AUC* methods, but in contrast to Eq. (1) the integrand is tg instead of g. Equation (11) gives rise to *AUMC* Methods 1A–11A.

Method 1A is a trapezoidal method for *AUMC* (e.g., 7).

$$AUMC_0^{t_n} \approx \sum_{i=1}^{n-1} \frac{(t_{i+1}-t_i)}{6}\{t_{i+1}(C_i+2C_{i+1})+t_i(2C_i+C_{i+1})\} \tag{12}$$

Method 2A is a log trapezoidal method for *AUMC* (8).

$$AUMC_0^{t_n} \approx \sum_{i=1}^{n-1} \{z_i(C_{i+1}t_{i+1}-C_it_i)-z_i^2(C_{i+1}-C_i)\} \tag{13}$$

where $z_i = (t_{i+1}-t_i)/\log[C_{i+1}/C_i]$, and the various numeric precautions discussed under Method 2 for *AUC* are taken.
In Method 3A

$$AUMC_0^{t_n} \approx \sum_{i=1}^{n-1} \left\{ t_i\left(\frac{h}{2}(C_i+C_{i+1})-\frac{h^3}{24}(C_i''+C_{i+1}'')\right) + h^2\left(\frac{C_i}{6}+\frac{C_{i+1}}{3}\right)-\frac{h^4}{6}\left(\frac{7}{60}C_i''+\frac{2}{15}C_{i+1}''\right)\right\} \tag{14}$$

where $h = t_{i+1}-t_i$ and the C_i'' are second derivatives returned from a cubic spline fit to all n values of C as in Method 3 for *AUC*.

Method 4A (Lagrange) is easily derived by integration of tg where g is defined by Eq. (5). Parameters a–d are found as in *AUC* Method 4.

Method 5A is derived by a straightforward integration of ty where y is the cubic interpolating function given in Eq. (1) of Yeh and Small (3). Yeh and Small did not derive or recommend this approach, preferring to use Method 5B (see below).

Method 6A is a hybrid employing the trapezoidal rule up to t_{max} and the log trapezoidal method [summand of Eq. (13)] thereafter. Because of the poor behavior of Method 1A (see Discussion), Method 1B [summand of Eq. (18)] is used for the trapezoidal steps.

Method 7A uses a spline [summand of Eq. (14)] up to t_{max} and then the log trapezoidal method [summand of Eq. (13)].

New Methods

Method 8A uses a trapezoidal step [summand of Eq. (18)] if $C_{i+1} \geq C_i$, otherwise the log trapezoidal method [summand of Eq. (13)] is applied.

Method 9A (parabolas-through-the-origin then log trapezoidal)

$$AUMC_0^{t_p} \approx \frac{a_2 t_2^4}{4} + \frac{b_2 t_2^3}{3} + \sum_{i=2}^{i_p-1} \left\{ \frac{a_i}{4}(t_{i+1}^4 - t_i^4) + \frac{b_i}{3}(t_{i+1}^3 - t_i^3) \right\} \tag{15}$$

Equation (15) is applied up to t_p [Eq. (6)] and the log trapezoidal method [summand of Eq. (13)] up to t_n. a_i and b_i are defined after Eq. (7). See the Appendix.

Method 10A (plateau function and log trapezoidal)

$$\begin{aligned} AUMC_0^{t_n} \approx{}& \frac{a_2 t_2^2}{2} + a_2 \exp[-b_2 t_2]\left(\frac{t_2}{b_2} + \frac{1}{b_2^2}\right) - \frac{a_2}{b_2^2} \\ &+ \sum_{i=2}^{n-1} a_i \left\{ \left[\frac{t_{i+1}^2}{2} + \exp[-b_i t_{i+1}]\left(\frac{t_{i+1}}{b_i} + \frac{1}{b_i^2}\right) \right] \right. \\ &\left. - \left[\frac{t_i^2}{2} + \exp[-b_i t_i]\left(\frac{t_i}{b_i} + \frac{1}{b_i^2}\right) \right] \right\} \end{aligned} \tag{16}$$

Parameter b_i is found by iteration as described in the discussion of Eq. (9). If $C_{i+1} < C_i$ or $C_i t_{i+1} / C_{i+1} t_i < 1.01$ the summand of Eq. (16) is replaced by the summand of Eq. (13) (a log trapezoidal step).

Method 11A (α function then log trapezoidal)

$$\begin{aligned} AUMC_0^{t_p} \approx{}& \frac{C_2}{b_2}\left\{ \frac{2\exp[b_2 t_2] - 2}{t_2 b_2^2} - t_2 - \frac{2}{b_2} \right\} \\ &+ \sum_{i=2}^{i_p-1} \frac{C_i}{b_i}\left\{ t_i + \frac{2}{b_i} + \frac{2}{b_i^2 t_i} - \frac{1}{z_i}\left(\frac{t_{i+1}^2}{t_i} + \frac{2t_{i+1}}{b_i t_i} + \frac{2}{b_i^2 t_i} \right) \right\} \end{aligned} \tag{17}$$

where z_i and b_i are as shown after Eq. (10) above. A linear trapezoidal step is taken if $z_i = 0$ or $0.99 < z_i < 1.01$. The summand of Eq. (17) is used up to t_p (Eq. 6) and then the log trapezoidal rule [summand of Eq. (13)] up to t_n.

In the second class of methods for *AUMC*, one may use the various functions g discussed earlier to interpolate between $t_i C_i$ and $t_{i+1} C_{i+1}$ instead of between C_i and C_{i+1}. If an algorithm F for *AUC* is written in the form $AUC \approx F(t_1, C_1, t_2, C_2 \ldots t_n, C_n)$ then a viable method is immediately obtained as $AUMC \approx F(t_1, t_1 C_1, t_2, t_2 C_2 \ldots t_n, t_n C_n)$. This gives AUMC

Methods 1B–11B directly from *AUC* Methods 1–11 above. Method 1B is the "classical" linear trapezoidal rule for *AUMC*, often written in the form.

$$AUMC_0^{t_n} \approx \sum_{i=1}^{n-1} \tfrac{1}{2}(t_{i+1} - t_i)(t_i C_i + t_{i+1} C_{i+1}) \tag{18}$$

The algorithm for Method 5B differs slightly from that of Method 5, in that the slope at t_1 is set to the value of C_1, and the assumption of monoexponential decay in the last interval is not made (3).

COMPARISON OF INTEGRATION METHODS

Simulated data sets of nine basic types were used (Fig. 2). In each type the theoretical concentration value was calculated at 15 different sample times.

1. Intravenous bolus 1 compartment with $k_e = 1$, at times $t = 0$, 0.1, 0.2, 0.3, 0.4, 0.6, 0.8, 1, 1.25, 1.5, 2, 2.5, 3, 5, 7.5.
2. Intravenous bolus 2 compartment with $\alpha = 1.5$, $\beta = 0.2$, $k_{21} = 0.4$, at times 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.75, 1, 1.5, 2, 3, 5, 7.5, 10, 15.

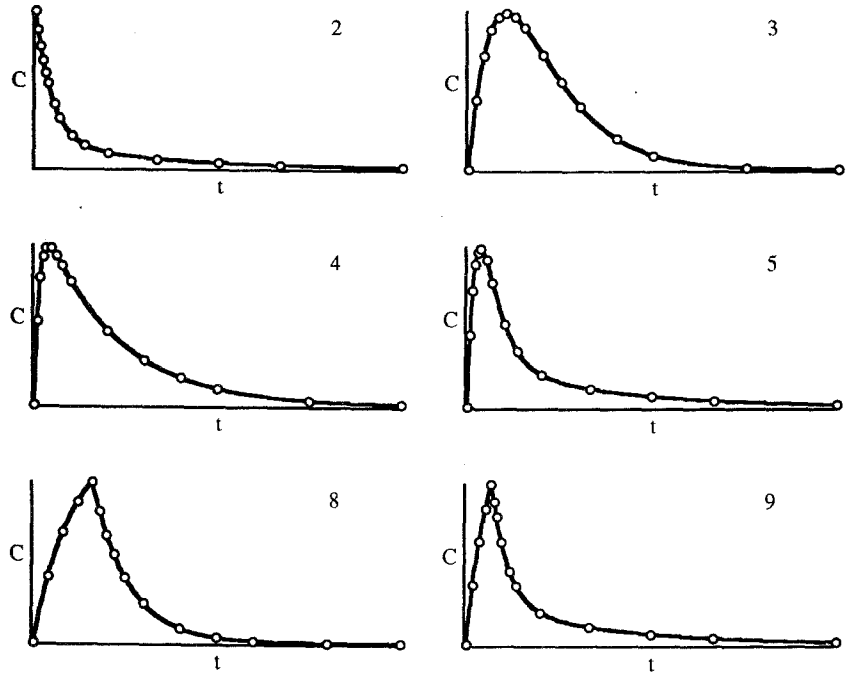

Fig. 2. Error-free data sets 2, 3, 4, 5, 8, and 9. Sample times are marked by dots.

3. Oral dose 1 compartment with $k_a = k_e = 1$, at times 0, 0.2, 0.4, 0.6, 0.8, 1, 1.25, 1.5, 2, 2.5, 3, 4, 5, 7.5, 10.
4. Oral dose 1 compartment with $k_a = 4$, $k_e = 0.25$, at times 0, 0.15, 0.3, 0.45, 0.6, 0.9, 1.2, 1.5, 2, 4, 6, 8, 10, 15, 20.
5. Oral dose 2 compartment with $k_a = 3$, $\alpha = 1.5$, $\beta = 0.2$, $k_{21} = 0.4$, at times 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.75, 1, 1.5, 2, 3, 5, 7.5, 10, 15.
6. As 3 but with a lag time $t_{lag} = 0.15$ added to each value of $t > 0$, to mimic a delay before the onset of absorption.
7. As 4 but with a lag time $t_{lag} = 0.1$ added to each value of $t > 0$, to mimic a delay before the onset of absorption.
8. Intravenous infusion 1 compartment with $T_{inf} = 1.6$ and $k_e = 1$, at times 0, 0.4, 0.8, 1.2, 1.6, 1.8, 2, 2.2, 2.5, 3, 4, 5, 6, 8, 10.
9. Intravenous infusion 2 compartment with $T_{inf} = 1$, $\alpha = 1.5$, $\beta = 0.2$, $k_{21} = 0.4$, at times 0, 0.25, 0.5, 0.75, 1, 1.1, 1.2, 1.4, 1.75, 2, 3, 5, 7.5, 10, 15.

The theoretical concentration at each sample time was perturbed by adding to it a pseudorandom normal deviate with a mean of zero and standard deviation equal to 10% of the theoretical value. For each of the nine types, $n = 800$ data sets with different pseudorandom perturbations were constructed, and every method for *AUC* and *AUMC* was tested on every data set. In Data Sets 8 and 9, the first derivative of the theoretical concentration is discontinuous at T_{inf}. Methods 3-5, 3A-5A and 3B-5B presuppose continuity of the derivatives, and were therefore applied in two steps: $t_1 - T_{inf}$ and $T_{inf} - t_n$.

For the purposes of displaying the errors due to each method, data sets were grouped into bolus (sets 1-2; $n_g = 2$), oral (sets 3-7; $n_g = 5$), and infusion (sets 8-9; $n_g = 2$). The percentage bias (average error) and percentage root-mean-squared error were calculated for each group

$$\%\text{bias} = \frac{100}{n_g n} \sum^{n_g} \sum^{n} \frac{A - E(A)}{E(A)} \tag{19}$$

$$\%\text{RMSE} = 100 \left\{ \frac{1}{n_g n} \sum^{n_g} \sum^{n} \left(\frac{A - E(A)}{E(A)} \right)^2 \right\}^{1/2} \tag{20}$$

where A is an estimate of *AUC* or *AUMC* and $E(A)$ is the expected theoretical value for the noise-free data set, due allowance being made for the "tail" corrections from t_n to ∞. The value $n = 800$ was chosen so that the standard errors of the mean estimates [and hence the standard errors of Eq. (19)] were all less than 0.1%. These results are shown in Tables I and II.

For each of the above data types, an additional, poorly sampled, type (1′-9′) was used in which the second, fourth, sixth, and ninth samples were omitted. The preceding analysis was repeated, but with n increased to 1600

Table I. Errors in *AUC* Estimates for Well-Sampled Data Sets

	Bolus 1-2		Oral 3-7		Infusion 8-9	
Method	%bias	%RMSE	%bias	%RMSE	%bias	%RMSE
1	2.0	3.5[a]	2.0	3.9[a]	1.5	3.5[a]
2	0.0	2.8[c]	−0.5	3.3[c]	−0.5	3.1
3	−0.2	2.9	0.0	3.8[b]	−0.2	3.1
4	−1.2	3.1[b]	−0.6	3.7	−0.9	3.3[b]
5	0.1	2.9	0.3	3.4	0.1	3.0[c]
6	0.0	2.8[c]	−0.3	3.3[c]	−0.2	3.0[c]
7	0.0	2.8[c]	−0.3	3.3[c]	0.1	3.0[c]
8	0.0	2.8[c]	−0.3	3.3[c]	−0.2	3.0[c]
9	0.0	2.8[c]	−0.2	3.3[c]	0.4	3.0[c]
10	0.1	2.8[c]	−0.3	3.3[c]	0.2	3.0[c]
11	0.0	2.8[c]	−0.2	3.3[c]	0.3	3.0[c]

[a] Largest error.
[b] Second largest error.
[c] Smallest error.

to ensure again that the standard errors of the mean estimates were all less than 0.1%. These results are shown in Tables III and IV. Figure 3 shows the distribution of estimates of *AUC* in Data Set 4′, by Method 4 (Lagrange) and Method 9 (PTTO). A large excess variance due to the Lagrange algorithm is evident.

As a test of the robustness of the methods when presented with very poorly sampled data, oral Data Sets 3″-7″ were prepared, consisting of the 1st, 5th, 7th, 9th, 12th, 14th, and 15th samples from Types 3-7. Results of *AUC* estimation with $n = 3200$ are given in Table V.

Finally, Data Sets 3P′-7P′ were prepared, identical to set 3′-7′ except that a constant concentration equal to the theoretical C_{max} was added at each sample time. The concentration in these data sets thus rises from, and declines to, a constant plateau level; this contrasts with all other data sets, whose declining phases are exponential or polyexponential decays to zero. Results of *AUC* estimation with $n = 800$ are given in Table V.

DISCUSSION

The %bias for any method in Tables I–V shows how close *on average* the estimates of *AUC* or *AUMC* produced by that method are to the theoretical value. Any entry outside the range ±1% indicates a bias highly significant in the statistical sense, representing a deviation of more than 10 standard errors. However, it can be argued that a bias of 1-2% is of little

Table II. Errors in *AUMC* Estimates for Well-Sampled Data Sets

	Bolus 1-2		Oral 3-7		Infusion 8-9	
Method	%bias	%RMSE	%bias	%RMSE	%bias	%RMSE
1A	6.1	7.4[a]	5.0	6.3[a]	3.7	5.1[a]
2A	0.2	3.5[c]	−0.2	3.7[c]	0.2	3.3[c]
3A	0.7	4.0	0.6	4.2	0.4	4.0
4A	−4.2	6.7[b]	−2.1	4.6[b]	−1.7	4.2[b]
5A	0.3	3.7	0.3	3.8	0.3	3.5
6A	0.2	3.5[c]	−0.2	3.7[c]	0.6	3.4
7A	0.2	3.5[c]	−0.2	3.7[c]	0.3	3.4
8A	0.2	3.5[c]	−0.2	3.7[c]	0.6	3.4
9A	0.2	3.5[c]	−0.2	3.7[c]	0.3	3.3[c]
10A	0.2	3.5[c]	−0.2	3.7[c]	0.3	3.3[c]
11A	0.2	3.5[c]	−0.2	3.7[c]	0.3	3.3[c]
1B	1.8	4.3	1.2	4.0	1.4	3.7
2B	−0.9	3.5[c]	−1.3	3.8	−0.7	3.3[c]
3B	0.2	3.9	0.1	4.0	0.1	3.6
4B	−0.6	4.1	−0.2	3.9	−0.2	3.6
5B	0.4	3.7	0.1	3.8	0.4	3.5
6B	−0.8	3.5[c]	−1.2	3.8	−0.1	3.3[c]
7B	−0.8	3.5[c]	−1.0	3.9	−0.3	3.3[c]
8B	−0.8	3.5[c]	−1.2	3.8	−0.1	3.3[c]
9B	−0.6	3.5[c]	−0.7	3.7[c]	−0.5	3.3[c]
10B	−0.7	3.5[c]	−1.1	3.8	−0.6	3.3[c]
11B	−0.7	3.5[c]	−0.9	3.7[c]	−0.5	3.3[c]

[a] Largest error.
[b] Second largest error.
[c] Smallest error.

practical pharmacokinetic consequence. Furthermore, the bias is sensitive to the precise values of the sample times t_i in each data set. In several trials (not shown) with different t_i values, the %bias for some methods (notably the trapezoidal Methods 1, 1A, and 1B) could differ by as much as ±1 from the values given in the tables. For this reason, as well as common sense, the occasional appearance of a bias outside the range ±1% should not be taken to condemn a method outright.

As well as producing estimates that are reasonably close to the true value on average, a good numerical integration method should produce estimates with little scatter, i.e., the variance of its estimates should be small. The %RMSE values shown in Tables I–V represent a convenient measure of error that combines both bias and variance errors, indicating the overall risk of obtaining discrepant estimates. These root-mean-squared (rms) errors thus provide a rational basis for comparison: Methods whose %RMSE values are consistently among the smallest (as indicated by Footnote *c* in the Tables) may be considered optimal, at least for the range of data types

Table III. Errors in *AUC* Estimates for Poorly Sampled Data Sets

	Bolus 1′-2′		Oral 3′-7′		Infusion 8′-9′	
Method	%bias	%RMSE	%bias	%RMSE	%bias	%RMSE
1	3.1	4.7[a]	1.9	4.4	0.7	4.2
2	0.4	3.5[c]	−1.1	3.9	−1.9	4.5[a]
3	0.0	3.8	0.4	6.0[a]	−0.5	4.1[c]
4	−1.1	3.9[b]	0.0	5.9[b]	−1.1	4.3
5	0.4	3.5[c]	0.7	4.1	−0.3	4.1[c]
6	0.4	3.5[c]	−1.0	3.9	−1.5	4.4[b]
7	0.4	3.5[c]	−0.4	3.8[c]	−0.2	4.1[c]
8	0.4	3.5[c]	−1.0	3.9	−1.5	4.4[b]
9	0.4	3.5[c]	0.0	3.8[c]	0.3	4.1[c]
10	0.4	3.5[c]	−0.2	3.8[c]	0.4	4.2
11	0.4	3.5[c]	−0.1	3.8[c]	0.3	4.1[c]

[a] Largest error.
[b] Second largest error.
[c] Smallest error.

considered here, whereas those with the largest %RMSE values (indicated by Footnotes *a* or *b* in the Tables) may be rejected as unsatisfactory.

The only method for *AUC* with consistently smallest %RMSE values in Tables I, III, and V is Method 9. However Methods 7 and 11 are inferior to it only in Data Sets 3P′-7P′ of Table V, and then by only a negligible amount. The performance of Method 10 is also scarcely distinguishable from Method 9. These four methods may therefore be considered optimal. From a practical standpoint, Method 9 has the advantage of being the easiest to program (see Appendix).

Selection of a list of optimal methods for *AUMC* is more difficult, since no method has consistently smallest %RMSE values in Tables II and IV. Methods 2A, 9A, 10A, 11A, and 9B appear to be the best, with Methods 7A and 11A scarcely distinguishable from them.

Bias errors in the determination of *AUC* were surprisingly small (Tables I and III) even for the trapezoidal rule (Method 1) whose theoretical deficiencies are well known (2,13). Much larger errors occurred in the estimation of *AUMC* (Tables II and IV), most of them due to Method 1A. This method has appeared in the pharmacokinetic literature in many algebraic forms (7,9,10,11,12), not always correctly. It is remarkable chiefly for having the worst bias and the worst rms error in every data set examined here. The method is therefore uniformly bad and should be abandoned.

Of particular concern are the large rms errors shown by some methods, notably Lagrange (Method 3) and cubic spline (Methods 4, 4A, and 4B). In most instances the large rms errors were *not* associated with large bias

Table IV. Errors in *AUMC* Estimates for Poorly Sampled Data Sets

	Bolus 1′–2′		Oral 3′–7′		Infusion 8′–9′	
Method	%bias	%RMSE	%bias	%RMSE	%bias	%RMSE
1A	6.6	7.9[a]	5.1	6.5[a]	3.9	5.7[a]
2A	0.3	3.7[c]	−0.4	3.9[c]	0.9	4.0
3A	0.5	4.2	0.9	4.8	0.4	4.4
4A	−4.4	6.8[b]	−1.8	5.4[b]	−1.6	4.8
5A	0.3	3.9	0.4	4.1	0.3	4.1
6A	0.3	3.7[c]	−0.3	3.9[c]	1.6	4.3
7A	0.3	3.7[c]	−0.4	3.9[c]	0.4	4.0
8A	0.3	3.7[c]	−0.2	3.9[c]	1.6	4.3
9A	0.3	3.7[c]	−0.3	3.9[c]	0.5	4.0
10A	0.3	3.7[c]	−0.4	3.9[c]	0.4	4.0
11A	0.3	3.7[c]	−0.4	3.9[c]	0.4	4.0
1B	1.3	4.3	0.6	4.0	2.4	4.9[b]
2B	−1.4	3.9	−2.2	4.4	−0.5	3.9[c]
3B	0.1	4.1	0.3	4.4	0.5	4.2
4B	−0.5	4.3	0.0	4.3	0.0	4.2
5B	0.2	3.9	−0.3	3.9[c]	0.8	4.1
6B	−1.3	3.9	−1.8	4.2	0.8	4.1
7B	−1.0	3.8	−1.2	4.4	0.1	3.9[c]
8B	−1.3	3.9	−1.8	4.2	0.8	4.1
9B	−0.6	3.8	−0.8	3.9[c]	−0.3	3.9[c]
10B	−0.9	3.8	−1.5	4.1	−0.4	3.9[c]
11B	−0.8	3.8	−1.2	4.0	−0.3	3.9[c]

[a] Largest error.
[b] Second largest error.
[c] Smallest error.

errors, indicating that these methods have a large excess variance. As illustrated in Fig. 3 the use of these methods carries a completely unnecessary risk of obtaining estimates far from the true value. The reason for the poor behavior of the spline methods is shown graphically by Yeh and Kwan (2); the widely used Lagrange methods suffer the same disadvantage. These

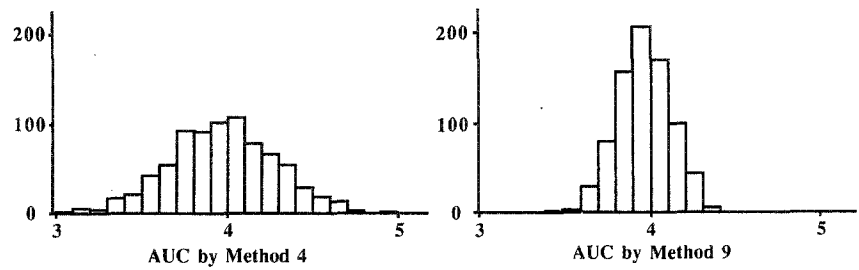

Fig. 3. Distribution of 800 estimates of *AUC* for Data Set 4′ by Method 4 (Lagrange) and Method 9 (PTTO). Theoretical value of *AUC* is 3.97.

Table V. Errors in *AUC* Estimates for Oral Data (Very Poorly Sampled and Poorly Sampled on Plateau)

	Oral 3″–7″		Oral 3P′–7P′	
Method	%bias	%RMSE	%bias	%RMSE
1	8.7	11.0	0.3	3.6[c]
2	−2.3	6.3	0.0	3.6[c]
3	0.1	11.5[b]	0.0	4.2[a]
4	−3.3	13.1[a]	0.0	4.2[a]
5	1.0	6.9	0.3	3.7
6	−2.3	6.3	0.1	3.6[c]
7	−1.5	6.1	0.1	3.6[c]
8	−2.3	6.3	0.1	3.6[c]
9	0.4	5.9[c]	0.0	3.6[c]
10	−0.7	6.1	0.3	3.6[c]
11	1.1	6.0	0.0	3.6[c]

[a] Largest error.
[b] Second largest error.
[c] Smallest error.

methods should be abandoned. The variance of the spline methods becomes tolerable when they are applied over only a small part of the total area, as in Methods 7 and 7A. The modified cubic polynomial (Methods 5 and 5B) introduced by Yeh and Small (3) is much better than the Lagrange and spline methods, but in most trials its rms error exceeded that of Method 9 (Tables I–V). The evidence of this study is that for area estimation a well-chosen two-point interpolation formula is preferable to even a carefully crafted four-point formula.

Many of the methods make use of the log trapezoidal rule for the declining phase. Because this rule is on theoretical grounds appropriate for data that decays exponentially to zero, it was of interest to include data sets (3P′–7P′) with very different decay characteristics. Comparison of Table V with Table III (Data Sets 3′–7′) suggests that the accuracy of the log trapezoidal rule is not appreciably disturbed. Indeed data sets 3P′–7P′ appear less demanding than Sets 3′–7′, since the rms errors are smaller. The log trapezoidal rule therefore seems to play a satisfactory role in the hybrid methods (Methods 6–11).

One cannot always tell which method for *AUMC* is used in published work. The term "trapezoidal method" in general appears to refer to Method 1B, and "log trapezoidal method" means Method 2A. But "cubic spline" without further qualification could mean either Method 3A or 3B. Similarly "Lagrange" could mean either Method 4A or 4B. Since the type A methods described here (other than 1A and 2A) seem not to be well known, in most

cases a type B method may be assumed. The ambiguity should in future be avoided by precise specification of the method used.

In this study I did not examine the accuracy of the "tail" corrections (extrapolations to ∞), formulas for which are well known (8,12), or the extrapolation to zero time for bolus data (12), although in practice the errors due to these extrapolations may be of similar magnitude to integration error. Nor have I explored strategies for optimum sample times. It would be interesting to extend the work of D'Argenio and Katz (14), who investigated optimum sample times for the linear trapezoidal rule, to the improved integration methods discussed here.

APPENDIX

The listing of a Microsoft QuickBasic program for AUC by Method 9 and AUMC by Method 9A is given below, with example data and results. Note that in this language, array arguments are indicated by empty parentheses following the variable name, e.g., C( ).

```
DEFINT i
DIM t(25), C(25)
CALL GetData (t(),C(),n)                    ' user-supplied routine
CALL PTTO (t(),C(),n, AUC,AUMC)
CALL ShowResults (t(),C( ),n,AUC,AUMC)      ' user-supplied routine
END

SUB PTTO (t(),C(),n,AUC,AUMC)
' Methods 9 & 9A     parabolas-through-the-origin then log trapezoidal.
' t() is array of sample times, C() is array of concentrations,
' n is number of samples.
ipeak = 1
IF C(1) = 0.0 AND n>2 THEN
   FOR i = 2 TO n
      IF C(i) > = C(ipeak) THEN ipeak = i
   NEXT i
   IF ipeak = 2 THEN ipeak = 3 ' go over top
END IF
AUC = 0.0
AUMC = 0.0
FOR i = 2 TO ipeak - 1      ' skip this if ipeak = 1
   ai = (C(i+1)/t(i+1) - C(i)/t(i))/(t(i+1) - t(i))
   bi = C(i)/t(i) - ai*t(i)
   IF i = 2 THEN
      AUC = ai/3.0*t(2)^3 + 0.5*bi*t(2)^2
      AUMC = 0.25*ai*t(2)^4 + bi/3.0*t(2)^3
   END IF
   AUC = AUC + ai/3.0*(t(i+1)^3 - t(i)^3) + 0.5*bi*(t(i+1)^2 - t(i)^2)
   AUMC = AUMC + 0.25*ai*(t(i+1)^4 - t(i)^4) + bi/3.0*(t(i+1)^3 - t(i)^3)
```

```
NEXT i
FOR i = ipeak TO n - 1    ' log trapezoidal (Methods 2 & 2A)
   IF C(i)*C(i+1) > 0.0 AND (C(i+1) < 0.999*C(i) OR C(i+1) > 1.001*C(i)) THEN
      zi = (t(i+1) - t(i))/LOG(C(i+1)/C(i))
      AUC = AUC + (C(i+1) - C(i))*zi
      AUMC = AUMC + (t(i+1)*C(i+1) - t(i)*C(i))*zi - (C(i+1) - C(i))*zi*zi
   ELSE                ' linear trapezoidal (Methods 1 & 1B)
      AUC = AUC + 0.5*(t(i+1) - t(i))*(C(i+1) + C(i))
      AUMC = AUMC + 0.5*(t(i+1) - t(i))*(t(i+1)*C(i+1) + t(i)*C(i))
   END IF
NEXT i
END SUB

t =   0.0  1.0  2.0  5.0  10.0  20.0  50.0
C =   0.0  1.5  2.0  1.0   0.5   0.25  0.1

AUC(0,50) = 19.12    AUMC(0,50) = 256.58
```

REFERENCES

1. K. Yamaoka, T. Nakagawa, and T. Uno. Statistical moments in pharmacokinetics. *J. Pharmacokin. Biopharm.* **6**:547-558 (1978).
2. K. C. Yeh and K. C. Kwan. A comparison of numerical integrating algorithms by trapezoidal, Lagrange, and spline approximation. *J. Pharmacokin. Biopharm.* **6**:79-98 (1978).
3. K. C. Yeh and R. D. Small. Pharmacokinetic analysis of stable piecewise cubic polynomials as numerical integration functions. *J. Pharmacokin. Biopharm.* **17**:721-740 (1989).
4. W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. *Numerical Recipes: The Art of Scientific Computing*, Cambridge University Press, Cambridge, 1985.
5. M. L. Rocci and W. J. Jusko. LAGRAN program for area and moments in pharmacokinetic analysis. *Comput. Prog. Biomed.* **16**:203-216 (1983).
6. J. J. B. Jack, D. Noble, and R. W. Tsien. *Electric Current Flow in Excitable Cells*, Clarendon Press, Oxford, 1975.
7. Erratum. *J. Pharmacokin. Biopharm.* **17**:527 (1989).
8. S. Riegelman and P. Collier. The application of statistical moment theory to the evaluation of in vivo dissolution time and absorption time. *J. Pharmacokin. Biopharm.* **8**:509-534 (1980).
9. M. K. Charter. The estimation of moments: a technical note. *J. Pharmacokin. Biopharm.* **17**:203-208 (1989).
10. A. Dunne and P. King. Estimation of noncompartmental parameters: a technical note. *J. Pharmacokin. Biopharm.* **17**:131-137 (1989).
11. P. Veng-Pederson and W. Gillespie. Mean residence time in peripheral tissue: a linear disposition parameter for evaluating a drug's tissue disposition. *J. Pharmacokin. Biopharm.* **12**:535-543 (1984).
12. P. Veng-Pederson. Mean time parameters in pharmacokinetics. Definition, computation and clinical implication. (Part II). *Clin. Pharmacokin.* **17**:424-440 (1989).
13. W. L. Chiou. Critical evaluation of the potential error in pharmacokinetic studies of using the linear trapezoidal rule method for the calculation of the area under the plasma level-time curve. *J. Pharmacokin. Biopharm.* **6**:539-546 (1978).
14. D. Z. D'Argenio and D. Katz. Sampling strategies for noncompartmental estimation of mean residence time. *J. Pharmacokin. Biopharm.* **11**:435-446 (1983).