Short Communication

Weighting Functions and Data Truncation in the Fitting of Multi-Exponential Functions

By GORDON L. ATKINS

*Department of Biochemistry, University of Edinburgh Medical School,
Teviot Place, Edinburgh EH8 9AG, U.K.*

(Received 17 October 1973)

Artificial data were simulated by using two-exponential functions and normally distributed pseudo-random numbers. The variation corresponded to either constant or relative variance for the data error. These data were used to test (i) three different weighting functions and (ii) the effect of data truncation on the precision of estimating the parameters of two-exponential functions.

The determination of physiological parameters in steady-state systems by using tracer kinetics, or pharmacological parameters by using drug kinetics, often requires the intermediate step of fitting a multi-exponential function to experimental data. The constants of the multi-exponential function are best obtained by a non-linear regression of specific activity or drug concentration on t (time); then the physiological or pharmacological parameters are calculated from these constants by using simple algebraic expressions. The effect of variation in the experimental data on the precision of the constants of the multi-exponential functions or the physiological or pharmacological parameters has been discussed by Myhill (1967, 1969), Glass & de Garreta (1967) and Atkins (1972). During the non-linear regression the data must be weighted by a statistically valid method. The true weighting factor is the reciprocal of the true variance, but since this is not known an estimate is required. Usually replicate analyses are performed so that observed variance can be measured and used as this estimate. However, in some experiments replicate analyses are not possible and it is standard practice to weight the data equally, if previous knowledge suggests that the coefficient of variation is thought to be constant, or to weight them as the reciprocal of the square of the observed value, if the coefficient of variation is thought to be proportional to the magnitude of the observed value. This communication investigates the effect of using the wrong type of weighting function in this situation. Ottaway (1971) suggests a third type of weighting function to be used on these occasions, and this weighting function is also evaluated. The evidence to support his suggestion was published after the present study was completed (Ottaway, 1973).

Llaurado & Smith (1971) have suggested that data truncation, i.e. stopping data collection before the

semi-logarithmic plot becomes linear, does not unduly affect the calculated parameters when a digital computer is used. This conclusion has been re-investigated by using more extensive data.

Methods

Three functions were used: (1) $y = 4 \exp(-0.4t) + \exp(-0.1t)$ ($t = 2, 4, 6, 8, 10, 12, 15, 18, 21, 24, 27, 30$), (2) $y = \exp(-0.1t) + 0.1 \exp(-0.01t)$ ($t = 2, 4, 6, 8, 10, 12, 19, 26, 33, 40, 47, 54$) and (3) $y = \exp(-0.1t) + 0.1 \exp(-0.01t)$ ($t = 2, 4, 6, 9, 12, 18, 26, 34, 42, 57, 72, 87$).

Simulated data sets were calculated by a method which has been well established by me (Atkins, 1971*b*, 1972, 1973).

Sets of normally distributed pseudo-random numbers were calculated (Subroutine RANDOM, Edinburgh Regional Computing Centre Program Library) with either a mean of 1.0 and s.d. ± 0.02 or ± 0.05 , or a mean of 0.0 and s.d. ± 0.005 , ± 0.01 or ± 0.02 . For each function two types of simulated data were generated. Data whose variation was dependent on y were obtained by multiplying values of y by pseudo-random numbers in sequence from the first distribution, and data whose variation was independent of y were obtained by adding pseudo-random numbers from the second distribution to values of y . The simulated data are designated $y_{\text{obs.}}$, where $y_{\text{obs.}} = y + \text{error}$. Forty sets were generated for each function with each type of variation and at each magnitude of variation. Two-exponential functions were then fitted to these data sets by the non-linear regression program of Atkins (1971*a*), by using the IBM 370/155 computer at the Edinburgh Regional Computing Centre. The values $y_{\text{calc.}}$ were calculated from the fitted function at the specified values of t . Each regression was performed three times, first with equal

weighting, secondly with a weighting proportional to $1/(y_{obs.})^2$ and thirdly with the weighting function proposed by Ottaway (1971, 1973). In this communication the correct weighting function means the data were equally weighted if the variation in error was constant or they were weighted by $1/(y_{obs.})^2$ if the variation was proportional to y . The wrong weighting function means that the weighting functions used were $1/(y_{obs.})^2$ and unity respectively. Some regressions were also performed by using $1/(y_{calc.})^2$ as a weighting function. The results were assessed according to three criteria. The first was to note the number of times a two-exponential function was a poor fit to the data, either from diagnostic messages given by the program or by an uneven distribution of residuals about the fitted curve. The second was to calculate, for each set of parameters obtained from acceptable fits of the functions, a mean, an s.d. and thus a coefficient of variation. The third was to estimate bias by calculating a Student's t value.

Results and discussion

Use of weighting factors. The results are summarized in Table 1. They clearly indicate that for both types of variation the use of the correct weighting function compared with the use of the two alternatives gave better results whether assessed by the number of poor fits, the coefficient of variation, or the bias. Also the median (Colquhoun, 1971, p. 26) was generally closer to the mean when the correct weighting function was used, which suggests that the alternatives gave skew distributions for the results. Ottaway's (1971, 1973) function, although not statistically valid

for either type of variation, occasionally gave better results than the correct weighting function if assessed by the first two tests, but the proportion of instances was not high enough and the pattern of distribution was not consistent for there to be any significant difference. When assessed by the third test for bias, then the correct weighting function gave better results. However, if we compare Ottaway's (1971, 1973) function with the wrong weighting function, then the former gave better results when assessed by the first two tests, but there is little difference when assessed by the test for bias.

The conclusion is therefore that if one is fitting exponential functions to experimental data consisting of single determinations at each point then it is preferable to have previous knowledge of the type of variation present so that the correct weighting function can be used. However, if the undesirable situation should occur in which the type of variation is not known then Ottaway's (1971, 1973) function is a good compromise.

All the data for the third exponential function were used for non-linear regressions weighted first by $1/(y_{obs.})^2$ and then $1/(y_{calc.})^2$. A Wilcoxon signed rank test (Colquhoun, 1971, p. 160) applied to the differences between paired parameter estimates (16 sets with 40 values per set) showed that there was no significant difference between the results obtained by the use of these two weighting functions. This observation differs from that of Ottaway (1973) possibly because the variation in the simulated data used in this present study has been kept small. It has previously been shown that a coefficient of variation for the data greater than 5% leads to very imprecise parameter

Table 1. Use of different weighting functions for fitting multi-exponential functions

The results were obtained by using the correct weighting function, the wrong weighting function and Ottaway's (1971, 1973) function when fitting three different two-exponential functions to data which are subject to constant variation or variation related to the magnitude of the observation, and show: the percentage of parameter estimations that gave a poor fit, the percentage of parameter sets that, for a given weighting function, gave the lowest coefficient of variation, and the percentage of parameter sets that gave the least bias.

Function ...	Constant variation			Relative variation		
	1	2	3	1	2	3
% poor fits						
Correct weighting	0	6.3	3.8	0	1.3	0
Wrong weighting	12.5	33.8	5.0	11.3	27.5	5.0
Ottaway's weighting	0	7.5	0	16.3	1.3	1.3
% lowest coefficient of variation						
Correct weighting	31.2	50.0	100	100	25.0	75.0
Wrong weighting	0	37.5	0	0	0	0
Ottaway's weighting	68.8	12.5	0	0	75.0	25.0
% least bias						
Correct weighting	37.5	75.0	50.0	50.0	75.0	58.5
Wrong weighting	37.5	0	25.0	25.0	0	25.0
Ottaway's weighting	25.0	25.0	25.0	25.0	25.0	16.5

estimates (Myhill, 1967, 1969; Glass & de Garreta, 1967; Atkins, 1972).

The three two-exponential functions used were of very different types, so it would be expected that the above conclusions should be applicable to other functions with a similar range of parameter values. The variations in the simulated data used in this study were obtained by using normally distributed numbers, so that the conclusions may only be valid for normally distributed variations.

These results were obtained by using simulated data with a normal distribution for the error, and the two-exponential functions are known to be a good fit to these data. Perhaps for real situations where such functions may only be a good approximation to the data or the error may be systematic the conclusions may be less definite.

Curtailment of data collection. Semi-logarithmic plots of the data sets of the first and third two-exponential functions have linear 'tails'. However, plots for the second function do not, because there are insufficient data at higher values of t . Llaurodo & Smith (1971), however, suggested that the use of a digital computer program should still allow one to calculate all four parameters of the functions. Table 2 compares the results obtained by using the second and third functions, the former being a truncated version of the latter, yet both having the same number of data points. It clearly shows that the mean values obtained with the extended time-scale were in all instances closer to the true value, although a t test failed to show any significant biases. It also shows that the precision of the mean values, as assessed by their coefficients of variation, was better with the extended data. In all instances the bias was always lower for the extended data. These conclusions differ from those of Llaurodo & Smith (1971). This is because,

first, these authors used experimental data and therefore did not know the characteristics of the theoretically correct function, and so could not estimate or test bias, and, secondly, the amount of data they analysed was insufficient, being derived from only eight experiments.

Conclusions

This communication shows that (i) for fitting multi-exponential functions to experimental data a statistically valid weighting function must be used, (ii) the function proposed by Ottaway (1971, 1973) can be of value in restricted circumstances, and (iii) the truncation of data collection produces less precise and possibly biased parameter estimates. It also illustrates that for theoretical investigations into parameter estimation large amounts of artificial data of known characteristics must be used rather than small amounts of data obtained experimentally.

I thank Miss Caroline Thompson for her valuable technical assistance.

Atkins, G. L. (1971a) *Biochim. Biophys. Acta* **252**, 405-420
Atkins, G. L. (1971b) *Biochim. Biophys. Acta* **252**, 421-426
Atkins, G. L. (1972) *Biochem. J.* **127**, 437-438
Atkins, G. L. (1973) *Eur. J. Biochem.* **33**, 175-180
Colquhoun, D. (1971) *Lectures on Biostatistics*, Clarendon Press, Oxford
Glass, H. I. & de Garreta, A. C. (1967) *Phys. Med. Biol.* **12**, 379-388
Llaurodo, J. G. & Smith, G. A. (1971) *Int. J. Biomed. Comput.* **2**, 265-275
Myhill, J. (1967) *Biophys. J.* **7**, 903-911
Myhill, J. (1969) *J. Theor. Biol.* **23**, 218-231
Ottaway, J. H. (1971) *Biochem. J.* **125**, 44p-45p
Ottaway, J. H. (1973) *Biochem. J.* **134**, 729-736

Table 2. Effect of data truncation on the accuracy and precision of parameter estimation

The results were obtained by fitting the function $y = X_1 \exp(-\lambda_1 t) + X_2 \exp(-\lambda_2 t)$ to data over a good time-range (NT) and a truncated time-range (T). The types of variation were: relative, ± 0.02 (A); relative, ± 0.05 (B); constant, ± 0.005 (C); and constant, ± 0.01 (D). The data were weighted by $1/(y/y_{obs})^2$ (A and B) or equally weighted (C and D). The true values of the parameters were $X_1 = 1$, $\lambda_1 = 0.1$, $X_2 = 0.1$ and $\lambda_2 = 0.01$. Each pair of values is the mean value and its coefficient of variation.

Data	Type of variation	No. of successful fits	X_1		λ_1		X_2		λ_2	
T	A	40	0.9605	1.66	0.0931	3.78	0.0779	21.08	0.00556	64.27
NT	A	40	0.9970	1.35	0.1001	2.30	0.1001	7.07	0.00999	9.58
T	B	39	0.9548	4.22	0.0965	9.65	0.0913	42.79	0.00738	90.65
NT	B	40	0.9921	3.31	0.1006	5.61	0.1025	15.75	0.01034	19.76
T	C	38	0.9923	2.83	0.1010	3.81	0.1090	27.52	0.01090	54.51
NT	C	37	0.9963	1.57	0.1002	2.60	0.1033	16.30	0.01026	28.28
T	D	38	0.9070	11.31	0.1043	19.22	0.1548	77.80	0.01462	94.33
NT	D	40	0.9941	3.10	0.1013	4.60	0.1082	28.13	0.01033	47.67