FROM THE DEPARTMENT OF SURGERY, UNIVERSITY OF SYDNEY, AND INSTITUTE OF MEDICAL RESEARCH, ROYAL NORTH SHORE HOSPITAL OF SYDNEY, CROWS NEST 2065, AUSTRALIA.

SOME EFFECTS OF DATA ERROR IN THE ANALYSIS OF RADIOTRACER DATA

by

J. MYHILL

Radioactive tracers have been used extensively to study some aspects of biological function in man. The results of a radiotracer study may often be a series of measurements of radioactivity extending over an interval of time after administration of the radioisotope. It is sometimes appropriate to represent the aspects which are under investigation by a simplified compartmental model and to fit the solution function of the model to the experimental data points in order to estimate the model parameters. If the system is in a steady state then it has been previously shown that the behaviour of labelled substances can be represented by linear differential equations (SHEPPARD & HOUSEHOLDER 1951, BERMAN & SCHOENFELD 1956) and that the solution for the activity versus time curve in each compartment consists of a sum of exponentials:

$$f(t) = \sum_{i=1}^{n} N_i e^{-\lambda_i t}$$

Work supported by grants from the New South Wales State Cancer Council, and the Cancer Research Committee of the University of Sydney. Submitted for publication 29 January 1968.

How the errors and omissions in the data affect the estimation of the rate constants, λ_i, and the amplitudes, N_i, and through them the estimation of the transfer rates and compartment sizes, has recently become a subject of study (Myhill, Wadsworth & Brownell 1965, Myhill 1965, 1966, 1967 a-c).

The collection of data is usually terminated for some experimental reason. The radioactivity may become too low to count accurately, the patient may become unavailable for study, or the radioactive label may become detached from its compound. This data truncation, together with the experimental error in each measured data point and the fewness of points, creates uncertainty in the values of the model parameters. This is the case even where a valid, unequivocal model of the system is assured.

Method of error analysis

1. *Computer simulation of data.* Data points were simulated on a high speed digital computer (English Electric KDF 9) using the equation for $f(t)$ and some random number generators. The details of the simulation have been previously described (Myhill 1968).

The situations where λ_1/λ_2 equals 10, 6, 4 and 2 were studied, the cases where the rate constants λ_i are closer to equality being the more difficult. If λ_1/λ_2 is less than about four, some methods of analysis become less effective (Brownell & Callahan 1963) or are subject to large bias (Myhill, Wadsworth & Brownell 1965). The ratio N_1/N_2 was set equal to 10, 6, 4, 2 or 1. The special case of $N_1/N_2 = \lambda_1/\lambda_2$ has been previously reported (Myhill 1967c). The simulated data were terminated when the value of the last data point was 5 % of the value of the initial point. This corresponds to a situation often observed in reports of experiments and is a more difficult case to analyse than if the data were further extended in time (i.e. so that the value of the last data point was less). In an endeavour to approximate to types of data seen in the literature, the cases of 31 and 11 data points equally spaced over the time interval of data collection, or else equally spaced in log-time (more points taken over the early portions of the data), were investigated. Random error with a constant per cent standard deviation was simulated at each point. The statistical distribution of the error was normal since it had been shown previously that the form of the statistical distribution of the error was of minor importance (Myhill 1967, 1968) and the magnitude of the standard deviation of error in a data point in any one set of data was varied from 0.5 % to 10 % between sets. The type of data simulated is shown in Figs 1 and 2 (continuous curves shown there without random errors added).

Table 1

Maximum amount of error in the data that allows convergence[1]

Ratio of rate constants λ_1/λ_2	Ratio of amplitudes N_1/N_2	Number of data points			
		11		31	
		Equal spacing	Log-equal spacing	Equal spacing	Log-equal spacing
10/1	6/1	10 %	10 %	10 %	10 %
10/1	4/1	10 %	10 %	10 %	10 %
10/1	2/1	10 %	10 %	10 %	10 %
10/1	1/1	*	*	10 %	10 %
6/1	10/1	3 %	1 %	6 %	3 %
6/1	4/1	10 %	7 %	10 %	10 %
6/1	2/1	10 %	10 %	10 %	10 %
6/1	1/1	*	*	9 %	10 %
4/1	10/1	1 %	0.5 %	2 %	2 %
4/1	6/1	4 %	1 %	6 %	3 %
4/1	2/1	8 %	4 %	10 %	8 %
4/1	1/1	*	*	10 %	8 %
2/1	10/1	**	**	0.5 %	**
2/1	6/1	**	**	0.5 %	0.5 %
2/1	4/1	**	**	1 %	0.5 %
2/1	1/1	*	*	2 %	1 %

[1]The error entered in the table is the percentage standard deviation at a data point. The error was normally distributed and of constant percent standard deviation throughout a set of data. The data were terminated such that the value of the last data point was 5 % of the value of the first data point. The maximum percent data error employed was 10 %.

* This combination of conditions was not simulated.

** For convergence using these simple methods, require an error less than 0.5 %.

2. *Numerical analysis of simulated data.* The function, $f(t)$, was fitted to each set of data using a valid non-linear least squares method. The estimated values of the λ_i, N_i, and their estimated standard deviations deduced from the variance-covariance matrix of each fit, were obtained. Each set of data was simulated fifty times (with everything the same except the random numbers employed to simulate the error) and analysed fifty times. By this means the variability and bias of both the estimates of the parameters and the estimates of their standard deviations were studied.

The computer programmes were written in Algol and run on an English Electric KDF 9 computer. When this work was commenced the University

Table 2

Average standard deviations of λ_i as per cent of the true λ_i — These standard deviations are the means of those derived from the variance/covariance matrix at each fit

Ratio of rate constants λ_1/λ_2	Ratio of amplitudes N_1/N_2	Error in data	Number of data points: 11, Equal spacing λ_1	11, Equal spacing λ_2	11, Log-equal spacing λ_1	11, Log-equal spacing λ_2	31, Equal spacing λ_1	31, Equal spacing λ_2	31, Log-equal spacing λ_1	31, Log-equal spacing λ_2
10/1	6/1	2 %	2.6	4.8	2.9	6.5	1.8	3.1	1.8	5.1
		5 %	6.5	12.1	7.8	17.3	4.5	7.8	4.5	12.7
		10 %	13.2	23.0	15.2	36.6	9.7	15.9	9.3	26.9
10/1	4/1	2 %	3.1	2.7	3.1	3.6	2.1	1.8	2.0	2.9
		5 %	7.6	6.9	8.2	9.6	5.3	4.4	4.9	7.4
		10 %	15.4	12.9	15.9	18.5	11.6	8.9	10.3	15.4
10/1	2/1	2 %	5.0	1.6	4.8	2.3	3.2	1.0	2.8	1.7
		5 %	12.4	4.1	11.9	5.8	8.3	2.5	6.9	4.2
		10 %	37.8	7.7	23.6	11.1	16.7	5.1	14.5	9.0
10/1	1/1	2 %	*		*		5.5	0.7	4.5	1.2
		5 %	*		*		14.4	1.9	11.1	3.1
		10 %	*		*		34.3	3.9	23.4	6.7
6/1	10/1	1 %	*		5.4	62.7	*		*	
		2 %	4.8	36.5	**		3.2	25.5	3.7	36.6
		5 %	**		**		7.8	67.3	**	
6/1	4/1	1 %	*		*		*		1.3	2.6
		5 %	8.7	13.2	12.7	23.1	6.0	8.5	6.7	13.3
		10 %	18.4	25.3	**		13.5	17.3	13.0	26.7
6/1	2/1	2 %	4.8	2.4	5.2	3.2	3.2	1.6	3.3	3.5
		5 %	11.6	6.2	13.7	8.6	8.1	3.8	8.1	6.3
		10 %	22.5	11.7	28.0	20.1			17.2	13.6
6/1	1/1	2 %	*		*		5.1	10.2	5.0	1.7
		5 %	*		*		13.1	2.6	12.3	4.4
		10 %	*		*		**		26.6	9.5
4/1	10/1	0.5%	1.8	14.5	4.8	54.3	*		1.5	14.1
		1 %	3.5	30.3	**		2.5	20.1	2.9	27.8
		2 %	**		**		4.8	41.7	6.2	58.6
4/1	6/1	1 %	2.9	11.6	6.7	37.0	2.1	8.1	2.4	11.3
		2 %	5.9	24.5	**		4.0	16.6	4.8	22.3
		5 %	**		**		10.6	40.0	**	
4/1	2/1	2 %	5.9	4.3	8.4	7.5	4.0	3.0	4.5	4.5
		5 %	14.3	10.7	**		10.4	7.6	11.6	11.7
		10 %	**		**		22.8	17.1	**	
4/1	1/1	2 %	*		*		5.7	1.8	6.2	2.7
		5 %	*		*		14.4	4.7	15.2	7.0
		10 %	*		*		33.8	12.5	**	
2/1	10/1	0.5%	**		**		4.6	28.8	**	
2/1	6/1	0.5%	**		**		3.5	17.4	4.7	22.2
2/1	4/1	0.5%	**		**		3.4	11.0	4.4	14.1
		1 %	**		**		7.2	22.0	**	
2/1	1/1	0.5%	*		*		4.2	2.7	5.1	3.8
		1 %	*		*		8.3	6.1	9.9	7.8
		2 %	*		*		16.8	12.1	**	

* This combination of conditions was not simulated. ** Convergence not obtained.

Table 3

Average standard deviations of N_i as per cent of true N_i — These standard deviations are the means of those derived from the variance/covariance matrix at each fit

Ratio of rate constants λ_1/λ_2	Ratio of amplitudes N_1/N_2	Error in data	Number of data points: 11, Equal spacing N_1	N_2	11, Log-equal spacing N_1	N_2	31, Equal spacing N_1	N_2	31, Log-equal spacing N_1	N_2
10/1	6/1	2 %	2.2	3.9	1.1	5.9	1.6	2.5	0.8	4.0
		5 %	5.2	9.9	3.1	15.5	3.8	6.3	1.9	10.1
		10 %	10.4	19.0	6.0	30.6	8.2	13.2	4.0	20.8
10/1	4/1	2 %	2.4	2.7	1.2	4.1	1.9	1.7	0.8	2.8
		5 %	6.0	6.9	3.4	10.4	4.7	4.4	2.1	6.8
		10 %	11.8	13.0	6.5	20.2	10.1	8.9	4.4	14.3
10/1	2/1	2 %	3.1	2.0	1.8	3.1	2.6	1.2	1.1	2.0
		5 %	7.7	5.3	4.8	7.9	6.4	3.2	2.8	4.9
		10 %	14.7	10.0	9.2	15.3	13.0	6.4	6.0	10.3
10/1	1/1	2 %	*		*		3.7	1.1	1.9	1.7
		5 %	*		*		9.3	2.8	4.6	4.1
		10 %	*		*		19.9	5.9	9.6	8.8
6/1	10/1	1 %	*		4.8	49.1	*		*	
		2 %	2.9	33.5	**		1.9	22.6	3.0	31.2
		5 %	**		**		4.5	52.5	**	
6/1	4/1	1 %	*		*		*		0.7	3.0
		5 %	5.7	15.0	6.3	26.4	3.9	9.9	3.7	15.1
		10 %	11.7	28.7	**		8.4	20.3	6.9	28.1
6/1	2/1	2 %	3.1	3.4	2.6	5.1	2.2	2.2	1.3	3.5
		5 %	7.6	8.8	6.8	13.4	5.5	5.6	4.4	8.6
		10 %	15.3	16.7	14.2	28.2			9.0	17.9
6/1	1/1	2 %	*		*		3.3	1.7	2.7	2.7
		5 %	*		*		8.4	4.4	6.6	6.6
		10 %	*		*		**		14.3	14.2
4/1	10/1	0.5%	1.7	19.0	5.9	59.8	*		1.8	17.5
		1 %	3.3	36.3	**		2.3	26.3	3.4	35.1
		2 %	**		**		4.4	49.1	7.2	73.4
4/1	6/1	1 %	2.4	16.2	6.9	42.5	1.6	11.0	2.5	15.1
		2 %	4.5	31.1	**		3.0	21.5	4.8	29.3
		5 %	**		**		7.5	52.6	**	
4/1	2/1	2 %	3.9	7.5	6.5	13.5	2.5	5.1	3.6	7.6
		5 %	9.3	17.6	**		6.3	12.6	9.0	18.6
		10 %	**		**		13.2	25.1	**	
4/1	1/1	2 %	*		*		3.5	3.2	4.8	5.0
		5 %	*		*		9.3	8.8	12.2	12.5
		10 %	*		*		20.3	17.4	**	
2/1	10/1	0.5%	**		**		11.5	11.6	**	
2/1	6/1	0.5%	**		**		7.9	48.5	11.3	67.9
2/1	4/1	0.5%	**		**		7.5	30.9	10.3	41.2
		1 %	**		**		14.6	60.3	**	
2/1	1/1	0.5%	*		*		8.5	8.8	11.8	11.9
		1 %	*		*		17.5	18.2	22.3	22.5
		2 %	*		*		30.7	32.0	**	

* This combination of conditions was not simulated. ** Convergence not obtained.

of Sydney did not possess an IBM 7090, or a Fortran compiler, and thus none of the well known fitting programmes, e.g. BERMAN, SHAHN & WEISS (1962) could be used.

Results

1. *Convergence of the iterations.* The maximum amount of error in the data that permitted convergence of this simple numerical technique was as shown in Table 1. In each case the true parameter values (used in generating the relevant data) were used as starting values for the iteration. From the point of view of starting values, these results thus represent the most favourable situation it is possible to achieve.

No special techniques were used to help the convergence. Each set of data was simulated fifty times, and a failure to converge for at least one set was regarded as meaning that convergence was not obtained for those data. In these regards the results are somewhat less favourable than might be obtained in practice.

Failure of convergence was judged to have occurred if, in the process of calculation, a number exceeded the machine limit ($\sim 10^{38}$) (this was almost always the case) or, rarely, if the change in parameter values was not less than 10^{-4} and the change in variance not less than 10^{-6} after twenty iterations. The parameter values were in the range 0.01 to 1.00.

2. *Error in the λ_i as a function of error in data and number of data points.* Table 2 shows the average error (throughout this paper error means percentage error) in λ_1 and λ_2 as a function of magnitude of error in the data, and number of data points, for normally distributed error. One result is obvious: 31 points give a lower error than 11 points for the same data error.

The lesser rate constant, λ_2, which largely determines the tail of the data, has more error than the greater constant, λ_1, if $N_1/N_2 \geqslant 4$ and $\lambda_1/\lambda_2 \geqslant 4$, and if the data are spaced equally in log-time. The same holds for equally spaced data except that if $\lambda_1/\lambda_2 = 10$ and $N_1/N_2 = 4$, then the error in λ_2 is the least. If $\lambda_1/\lambda_2 \geqslant 4$ and $N_1/N_2 \leqslant 2$, then the error in λ_2 is lower than in λ_1, except that for $\lambda_1/\lambda_2 = 4$ and $N_1/N_2 = 2$, the errors are equal. For $\lambda_1/\lambda_2 = 2$, the error in λ_2 is lower for $N_1/N_2 = 1$ only. These statements apply to the studies with 31 points and the error referred to is the percentage error, as is always the case in this paper.

The same results were observed in studies with eleven points except that for data spaced equally in log-time if $\lambda_1/\lambda_2 = 4$ and $N_1/N_2 = 2$, then the error in λ_2 is the lesser.

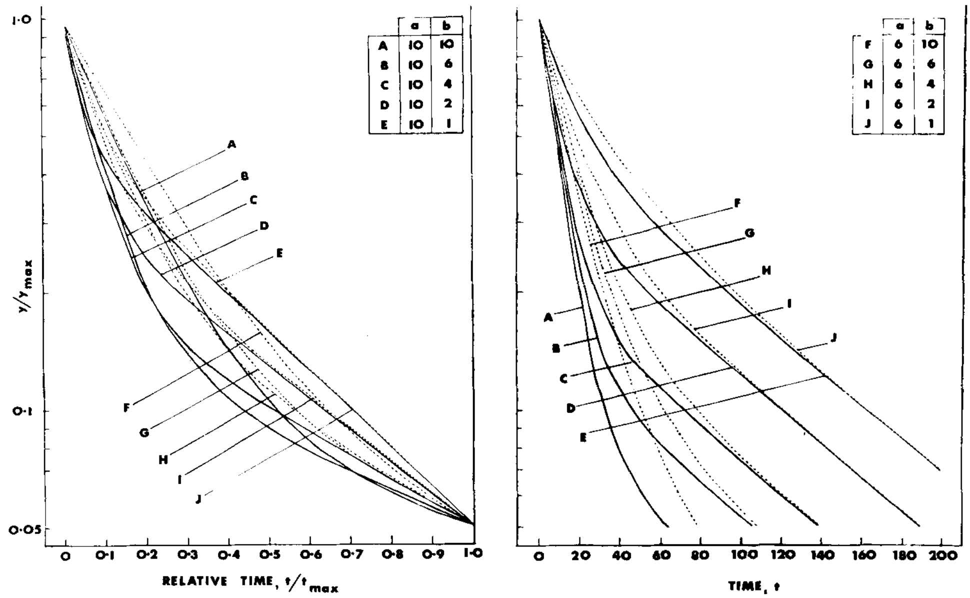

Fig. 1. Sums of two exponentials having rate constants of ratio 10 and 6. The amplitude ratios are 10, 6, 4, 2 and 1. A sum is terminated when it has declined to 5 % of its initial value. Each curve is plotted twice: on the left both magnitude and time are normalised; on the right the magnitude alone is normalised.

The percent error in λ_1 was only sometimes lessened by employing log-equal spacing of the data, while the percent error in λ_2 was always increased.

3. *Error in N_i as a function of error in data and number of data points.* The average errors in N_1 and N_2 as a function of magnitude of error in the data, and number of data points for normally distributed error, are shown in Table 3.

With 31 points and log-equally spaced data the percent error in N_2 is equal to that in N_1 when $N_1/N_2 = 1$, irrespective of λ_1/λ_2, and is greater than that in N_1 for all other values of N_1/N_2. The same holds for 11 points.

For equally spaced data, if $N_1/N_2 \geqslant \lambda_1/\lambda_2$, then the percent error in N_2 is the greater. The errors are about equal for $\lambda_1/\lambda_2 = 10$ and $N_1/N_2 = 4$; $\lambda_1/\lambda_2 = 6$, and $N_1/N_2 = 2$; $\lambda_1/\lambda_2 = 4$ or 2, and $N_1/N_2 = 1$. For each of these values of λ_1/λ_2, if N_1/N_2 exceeds the value quoted, then the percent error in N_2 is the greater, otherwise it is the lesser.

For log-equal spacing the absolute errors in the N_i are approximately equal.

For $\lambda_1/\lambda_2 \geqslant 6$, the log-equal spacing (i.e. crowding the points near the

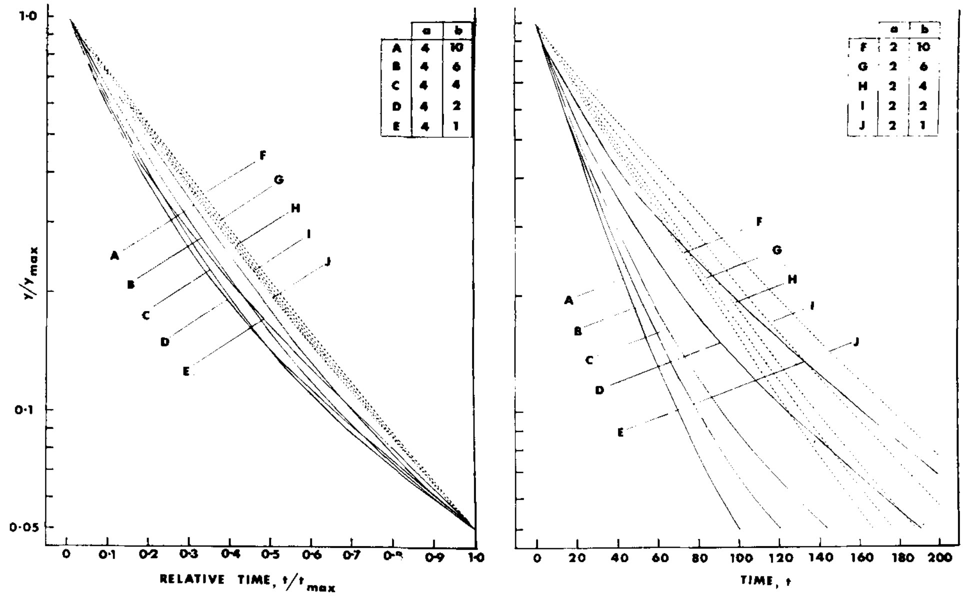

Fig. 2. Sums of two exponentials having rate constants of ratio 4 and 2. The amplitude ratios are 10, 6, 4, 2 and 1. A sum is terminated when it has declined to 5 % of its initial value. Each curve is plotted twice: on the left both magnitude and time are normalised; on the right the magnitude alone is normalised.

beginning of the data) always improved the per cent error in N_1, over that obtained in equally spaced data. For $\lambda_1/\lambda_2 \leqslant 4$ the opposite was true.

Discussion

The tracer activity curve in any steady-state two-compartment open system can be described by a sum of two exponential terms. The relationships between the transfer rates and compartment sizes, on the one hand, and the rate constants (λ_i) and amplitudes (N_i), on the other hand, are expressible by simple algebraic equations. The form of the equations depends on the configuration of the system, and these equations can be used to transform information on the errors obtained in λ_i and N_i to information about the associated errors in the transfer rates and compartment sizes. By this means this error information can be applied to a range of biological problems, mainly tracer studies in biochemistry and physiology.

Within the conditions here investigated convergence is readily obtained by this simple numerical method if the ratio of λ_1 to λ_2 is not less than about four. For lower ratios (i.e. rate constants more nearly equal) a very good experiment is necessary, that is, the standard deviation in each data point should be about 1 %. In practice, since the best starting values for the iteration are of course unknown, convergence may be difficult to obtain even when these requirements are satisfied. However, the employment of various techniques to help convergence, and trial of various starting values, would improve the situation over that shown in Table 1.

For any desired error in λ_i or N_i, the allowable error in the data may be found from Tables 2 and 3. As the data error increases the parameter error increases proportionately. For the same error in each point, taking more points results in a smaller parameter error. The variation of error in the λ_i and N_i depends on the spacing of the data points and the ratios λ_1/λ_2 and N_1/N_2. The parameter errors could be reduced by collecting data extending further in time, as well as by lowering the error in each point or taking more points over the same time interval. The present study, however, was to examine the situation where, for experimental reasons, the data could not be extended further in time.

The conclusions of this paper apply to any two-compartment system from which data has been collected until $f(t) = qf(0)$, $q = 0.05$, in one of the compartments, and for which $\lambda_1 > \lambda_2$ implies $N_1 \geq N_2$. The results do not depend on the absolute magnitudes of the N_i or λ_i. In practice, if q is larger, then the analysis is more difficult, the errors in N_i and λ_i will be greater, and the situation is worse in other respects. If q is smaller, i.e. the data are sampled for a longer time, the results will be better. If more data points are taken the results will also be better. The results quoted herein can thus afford a guide in evaluating other situations. If three compartments are being studied then the results would be the best that one could possibly achieve for the errors in four of the six parameters.

SUMMARY

The effect of error in radiotracer data on the accuracy of derived parameters in two-compartment systems was studied. Computer simulation was employed to produce data with error of different magnitude for each of several values of rate constant and amplitude. The data were arranged to simulate experimental situations and analysed by non-linear least squares. For a given error in the data the calculated errors in the rate constants and amplitudes were defined in terms of their ratios and the number and spacing of the data points.

ZUSAMMENFASSUNG

Der Einfluss der bei der Analyse mit radioaktiven Trägern entstandenen Datafehler auf die Genauigkeit der derivierten Parameter wurde studiert. Bei Simulation mit einer Datenverarbeitungsmaschine wurden Datafehler verschiedener Grösse für jeden von mehreren

Zahlwerten der Geschwindigkeitskonstanten und Amplituden produziert. Die Daten sollten experimentelle Bedingungen simulieren und wurden mit der Methode nicht-linearer Mindest-Quadraten analysiert. Für einen bestimmten Datafehler wurden die berechneten Fehler in Geschwindigkeitskonstanten und Amplituden als Funktionen der Verhältniszahlen der Konstanten und Amplituden sowie der Anzahl und Abstände der Datapunkte ausgedrückt.

RÉSUMÉ

L'auteur a étudié l'effet d'erreurs de mesures de radioactivité de traceurs sur l'exactitude des paramètres calculés à partir de ces mesures. Il a utilisé la simulation sur ordinateur pour obtenir des mesures entachées d'erreurs de différentes grandeurs pour chaqune des différentes valeurs des constantes de vitesse et des amplitudes. Il a combiné ces mesures de façon à simuler des situations expérimentales et les a analysées par la méthode non-linéaire des moindres carrés. Pour une erreur donnée de la mesure, il a calculé les erreurs sur les constantes de vitesse et sur les amplitudes pour différentes valeurs de leurs rapports et pour différents nombres et différents espacements des points de mesures.

REFERENCES

BERMAN M. and SCHOENFELD R.: Invariants in experimental data on linear kinetics and the formulation of models. J. appl. Phys. 27 (1956), 1361.

— SHAHN E. and WEISS M. F.: The routine fitting of kinetic data to models. Biophys. J. 2 (1962), 275.

BROWNELL G. L. and CALLAHAN A. B.: Transform methods for tracer data analysis. Ann. New York Acad. Sci. 108 (1963), 172.

KENDALL M. G. and STUART A.: The advanced theory of statistics. 2 (1963) p. 83.

MYHILL J.: Analysis of radioisotope tracer data. Proc. Symposium on Use of Computers in Medicine and Biology, Melbourne 1965.

— Computer analysis of radioisotope tracer kinetic studies. Proc. 6th Ann. Meeting on Physics in Medicine & Biology, Melbourne 1966.

— (a) Effects of data error in two-compartment analysis. Proc. 7th Ann. Meeting on Physics in Medicine and Biology, Adelaide 1967.

— (b) Effects of error on compartmental analysis. Abstracts 7th Intern. Conf. Med. Biol. Eng., Stockholm 1967.

— (c) Investigation of the effect of data error in the analysis of biological tracer data. Biophys. J. 7 (1967), 903.

— Investigation of the effect of data error in the analysis of biological tracer data from three-compartment systems. J. theoret. Biol. (in press 1968).

— WADSWORTH G. P. and BROWNELL G. L.: Investigation of an operator method in the analysis of biological tracer data. Biophys. J. 5 (1965), 89.

PIKE M. C.: Random normal deviate. Comm. A. C. M. 8 (1965), 606.

SHEPPARD C. W. and HOUSEHOLDER A. S.: The mathematical basis of the interpretation of tracer experiments in closed steady-state systems. J. appl. Phys. 22 (1951), 510.