Journal of Pharmacokinetics and Biopharmaceutics, Vol. 5, No. 5, 1977

Curve Fitting and Modeling in Pharmacokinetics and Some Practical Experiences with NONLIN and a New Program FUNFIT

Peter Veng Pedersen[1,2]

Received June 15, 1976—Final April 26, 1977

The problems of curve fitting and modeling in pharmacokinetics are discussed. A new nonlinear regression program FUNFIT, written for interactive time sharing, is presented which should be more reliable than programs based on the Gauss–Newton or other related gradient methods. The new program and the well-established program NONLIN were tested on two linear models using human plasma drug level data. FUNFIT found a substantially better solution than NONLIN in the majority of the cases.

KEY WORDS: pharmacokinetics; nonlinear regression; curve fitting; computer program; time sharing; modeling; weighting; least squares; parameter estimation; discrimination between models.

INTRODUCTION

Several programs are available for nonlinear least-squares parameter estimation (1–15). Nearly all are based on the Gauss–Newton or other related gradient methods since these are usually rapidly convergent and provide estimates of the variance–covariance matrix. However, such gradient methods may fail when the residuals are large (16,17), as is often the case in fitting equations to biological data, and they may converge on a nonstationary point (18) if great care is not taken by the user in choosing a suitable value for the step size used in the finite difference approximation of derivatives. The default value specified for this step size in a program may unfortunately apply successfully only in a limited number of cases. Such practical experiences are illustrated in this article in the application of the

[1]University of Sydney, Department of Pharmacy, Sydney, N.S.W. 2006 Australia.
[2]Present address: Department of Pharmacy, School of Pharmacy, University of California, San Francisco, California 94143.

program NONLIN (1), which is based on Hartley's modification of the Gauss–Newton algorithm (19)[3].

A new program, FUNFIT, for general nonlinear regression written for interactive time sharing is presented, which eliminates such problems. The program has implemented the adaptive simplex method of Nelder and Mead (21,22), which is a nongradient method that does not require evaluation of derivatives. This method is less efficient than Gauss–Newton based methods but considerably more robust and reliable. It will essentially never fail even under extreme conditions where the gradient methods may be unstable due to near singularity and ill-conditioning of the matrices used in the iteration procedure.

The possibility of multiple solutions (multiple sums of squares minima) undoubtedly represents the greatest problem in nonlinear estimation. The problem is expected to be particularly pronounced in pharmacokinetic studies, because these often involve the fitting of multiexponential equations to rather variable biological data and because the ratio of number of data points to number of parameters is often quite small. Parameter estimation under such conditions may produce spurious results, and discrimination between pharmacokinetic models may be very difficult and unreliable. The problem can be reduced, but usually not eliminated, if a graphical or numerical method is available which provides good initial parameter estimates for the iteration procedure. This is seldom the case for models describing nonlinear pharmacokinetics. The best approach should therefore be to use an algorithm which is effective in finding the (statistically) best solution in terms of the smallest residual sum-of-squares value.

It is generally accepted that the nongradient search methods perform better than the gradient methods in this respect. In particular, the adaptive simplex method used in FUNFIT appears, because of its unique minimization method, to be very suitable.

Regardless of the choice of algorithm, the question of which starting values the parameters should be given still seems to be the greatest practical problem the user faces in nonlinear estimation. Frequently, when no preliminary estimation technique is available the initial values are simply guesses which all too often produce unacceptable results in the first run. However, by studying these results, corrections can often be made so that acceptable results can be obtained in subsequent runs.

Interactive programs are most convenient in such cases. In particular, FUNFIT has been designed so that it allows a highly interactive and flexible editing of input at any stage.

[3]Numerical techniques in nonlinear parameter estimation have been reviewed by Chambers (20) and discussed excellently by Dennis (17).

THEORY

The parameters in nonlinear regression are most frequently estimated by the method of least squares, i.e., by finding the minimum of the sum of the (weighted) squared residuals, ss(P):

$$\mathrm{ss}(P) = \sum_{j=1}^{\mathrm{NFUNC}} W_j \sum_{i=1}^{\mathrm{NOBS}_j} w_{ij}[Y_{ij}(\mathrm{OBS}) - Y_{ij}(\mathrm{CALC})]^2 \qquad (1a)$$

$$Y_{ij}(\mathrm{CALC}) = f_j[X_{ij}(\mathrm{OBS}), \mathbf{P}_j] \qquad (1b)$$

where NFUNC is the number of functions to be fitted and NOBS_j is the number of observations for the jth of these functions; Y is the dependent variable (response) and X the independent variable(s). The subscript ij denotes the ith element in the jth response system measured; $\mathbf{P}$ is the complete set of parameters and f_j is the jth function describing (modeling) the observation in the jth response system (e.g., blood, urine, bile, or tissue drug levels or a pharmacological response).[4]

Weighting of Observations and Response Systems

The observations should be weighted inversely proportional to their variances. If these variances are assumed constant, the weights would be $w = 1$ ("unweighted data"), but in several cases it is assumed, or verified experimentally, that there is a certain functional relationship between the variances and the dependent or independent variables, and the data are weighted accordingly. However, it would be incorrect in statistical estimation or in mathematical modeling to use weights just to get a better fit. There must be a sound basis for the weighting scheme used.

The problem of weighting is of particular importance in simultaneous fitting (NFUNC > 1) because the error variances may differ considerably between the response systems. The correct approach in such situations would be first to fit each response system *individually* in order to establish a variance estimate for the observations in that system[5] and then fit the systems simultaneously where they are weighted (W_j in equation 1) proportional to their variance estimates.

The function(s), f, to be fitted is usually of explicit form. However, equations of implicit form, which frequently arise in nonlinear pharmacokinetics, and equations where the dependent and the independent variables are given separately in a parametric form can also be fitted.

[4]In most pharmacokinetic applications, only one response system, i.e., the plasma level, is considered (NFUNC = 1). However, it is possible using FUNFIT to fit simultaneously up to ten different functions, each containing up to 20 parameters and constants, and 10 variables.

[5]The variance estimate is the residual sum of squares divided by the residual degrees of freedom. The degrees of freedom is the number of observations in the particular response system minus the number of parameters in the function describing that system.

The theory of nonlinear least squares and nonlinear parameter estimation has been extensively discussed in the literature (23,24).

It is most important to emphasize that statistical estimates such as variances, covariances, and related estimations (confidence limits, F and t tests) are biased because the statistical procedure is based on a truncated Taylor series approximation of the nonlinear system. Simulation studies have shown that these estimates can be manyfold different from their true values (25).

Pharmacokinetic Applications of FUNFIT

1. The classical linear, compartmental models are still the most often used models in pharmacokinetic studies. The evaluation of such models is well documented and has become a routine procedure in many investigations.
2. There has, however, been an increasing awareness that linear models cannot adequately describe certain drug disposition phenomena (26,27), and various nonlinear models have been postulated. These mathematical models are often of a form which requires a special technique for least-squares fitting.
3. Often several possible models are investigated to explain a pharmacokinetic phenomenon. There has been increasing interest in discriminating between such models (27,28).

It is appropriate to discuss points 2 and 3 above.

Fitting of Implicit Functions: A Simple Example

Several of the models describing nonlinear pharmacokinetic phenomena can be expressed in an implicit form which can be fitted by defining the functional relationship between the variables explicitly by an iterative procedure.

Consider, for example, a simple one-compartment model with intravenous injection in which the drug is eliminated partly by conversion to a single metabolite according to Michaelis–Menten kinetics and partly by excretion unchanged in the urine (29). The concentration of drug in plasma, c, is given *implicitly* by

$$\ln \frac{c}{c_0} = \left[\frac{V_m}{k_{1u}K_m}\right] \ln \left[\frac{k_{1u}K_m + V_m + k_{1u}c_0}{k_{1u}K_m + V_m + k_{1u}c}\right] - \left[\frac{k_{1u}K_m + V_m}{K_m}\right] t \qquad (2)$$

where V_m and K_m are the Michaelis–Menten parameters, $c_0 = \text{dose}/V_1$, and k_{1u} is the urinary elimination constant. This equation can be written more simply:

$$\ln c + A_1 \ln (A_2 + A_3 c) + A_4 = 0 \qquad (3a)$$

where

$$A_1 = V_m/k_{1u}K_m \tag{3b}$$

$$A_2 = (k_{1u}K_m + V_m)/(k_{1u}K_m + V_m + k_{1u}c_0) \tag{3c}$$

$$A_3 = k_{1u}/(k_{1u}K_m + V_m + k_{1u}c_0) \tag{3d}$$

$$A_4 = \frac{k_{1u}K_m + V_m}{K_m}t - \ln c_0 \tag{3e}$$

The dependent variable, c, cannot be isolated from equation 2 or 3 but must be found by an iterative procedure by solving equation 3 for c.[6] The Newton–Raphson algorithm provides a simple and rapidly convergent method. If the expression in equation 3a is denoted $g(c)$, then c can be determined by the following iteration:

$$c_{i+1} = c_i - g(c_i)/g'(c_i) \tag{4}$$

where

$$g'(c) = \frac{\partial g}{\partial c} = \frac{1}{c} + \frac{A_1A_2}{A_2 + A_3c} \tag{5}$$

It is necessary that c_{i+1} in this iteration does not take a nonpositive value since this will terminate the execution of the program because $\ln c$ is not defined for $c \leq 0$. To prevent this, it is most convenient to define

$$c_{i+1} = c_i/2 \qquad \text{if} \qquad c_{i+1} \leq 0$$

This is an acceptable approach because $g(c)$ is strictly increasing for $c > 0$ since $A_1, A_2, A_3, A_4 > 0$ and therefore $g'(c) > 0$. The term $\ln(A_2 + A_3c)$ in $g(c)$ will also be defined under these conditions.

The above procedure can be used for most implicit mathematical models in nonlinear pharmacokinetics. However, special care must be taken to prevent the parameters' wandering into a parameter space where the function(s) is not defined (e.g., logarithm of a nonpositive number, division by zero).

Discrimination Between Models

The best criterion to use in discriminating between alternate pharmacokinetic models depends on the aim of the investigation and the application of the results (25,30–32). If the main aim is to discriminate

[6]It would be incorrect, as is sometimes done for equations of similar type, to fit directly the equation where t is expressed as a function of the dependent variable. The result would be unreliable because the dependent variable, which accounts for nearly all of the errors in the data, in this way is considered as an independent variable without error.

between models, the experiment should be designed so that the hypothesized models are placed in as much jeopardy as possible.

The problem is nevertheless considerably complicated by the substantial variation and low reproducibility of measurements in a biological system and the limited number of sample points available. Discrimination on a statistical basis requires information about the variability of the observations which can be estimated only by repeated experiments. A likely outcome of such experiments would often be that the system is 'ill-conditioned"; i.e., the variability of the data is too large to allow a discrimination on a significant probability level.

Since the macroparametric representations of linear, compartmental models are often of the multiexponential form

$$c = \sum_{i=1}^{n} A_i e^{-\alpha_i t} \qquad (A_i, \alpha_i > 0) \tag{6}$$

it appears appealing in routine investigations of raw pharmacokinetic data to apply a "multiple regression approach," similar to that used for linear systems, to determine the order, n, of the system.

This seems, however, to be an unreliable approach, for several reasons:

1. There exists no computer program which will inevitably find the "best" solution (smallest residual sum of squares) in a nonlinear least-squares estimation which may have several minima.
2. Measurements in biological systems often produce substantial residual values which may give rise to multiple minima.
3. The number of minima will increase very rapidly as the number of exponential terms, n, to be fitted increases.

The problem of multiple minima can be reduced but *not* eliminated by a suitable procedure which gives good initial estimates, or by multiple runs with initial parameter values randomly taken from the parameter space, or by performing a lattice search. The interactive structure of FUNFIT makes it particularly suitable for performing multiple runs and lattice searches.

So-called back-projection or stripping is the technique most frequently used to obtain initial parameter estimates for models of the form described by equation 6. The stripping is done either graphically or automatically by the computer, in some cases employing a spline function representation of the data (27).

However, it is important to realize that this particular technique assumes that one or more exponential terms vanish in certain regions of the total drug concentration–time curve. In other words, the method tends to disregard cases where two or more exponential terms dominate fairly equally throughout the whole time space investigated. Hence the method

may produce biased results. Discrimination between alternative models (equation 6) on this basis must therefore be considered unreliable.

It would be appropriate in this connection to refer to a different method which does not introduce such a bias (33). This method is based on a linear shift operator technique which appears not to have been used previously in pharmacokinetic studies. It seems to be currently the most suitable to use in obtaining initial parameter estimates in linear compartmental models.[7]

In evaluating how well a model describes some data, three points must be considered, which will be discussed: (a) How well do the calculated values agree with those observed, i.e., what is the sum of squared residuals or the correlation coefficient? (b) Does the fit agree with the basic assumptions made about the errors? (c) How predictive is the model?

A comparison of fits entirely in terms of sum of squared residuals (such as an F test) must be considered insufficient.

Analysis of Residuals

The importance of an analysis of residuals (24) seems to have been completely ignored in most computer programs. The basic assumptions in nonlinear least squares are the following: (1) The independent variable(s) is without error. (2) The errors, ε_i, in the dependent variable are independent $[\mathrm{Cov}_{i\neq j}(\varepsilon_i, \varepsilon_j) = 0]$ and normally distributed with zero mean and the same variance $[\varepsilon_i \sim N(0, \sigma^2)]$.

Possibly the best way to examine the residuals is to plot them against the independent and dependent variables (24,25,35). Significant systematic deviations can be visualized in this way. FUNFIT includes such plots and the following statistics which may be helpful in the assessment and comparison of models.

The Kolmogorov–Smirnov statistic (36) is used to test for normality of the residuals. The procedure is simply: Given N residuals, the program calculates

$$D = \max_x |F^*(X) - S_N(X)| \tag{7}$$

where $S_N(X)$ is the cumulative distribution of the residuals and $F^*(X)$ is the cumulative normal distribution function with the same mean and variance as the residual sample. The calculated value of D is compared with the critical value obtained from a Monte Carlo calculation at a given significance level.

The fundamental assumption of random errors is also tested in FUNFIT using two nonparametric tests that will be called the "run test" and the "number test." The run test is based on an analysis of runs of residuals of equal sign (37). For example, the sequence of residuals (+++)(−−)(+)(−−)(+) forms $r = 5$ runs. The least number of residuals with the same sign is

[7]The adaptation of this method to FUNFIT is under investigation.

$L = 4$. If N residuals, with equal probability of being negative and positive, form a sequence with r runs, then the probability of getting $\leq r$ runs is

$$P(\leq r) = \binom{N}{L}^{-1} \sum_{i=2}^{r} f_r \tag{8a}$$

where

$$f_r = 2\binom{L-1}{\frac{i}{2}-1}\binom{N-L-1}{\frac{i}{2}-1} \qquad \text{for } i \text{ even} \tag{8b}$$

and

$$f_r = \binom{L-1}{\frac{i-1}{2}}\binom{N-L-1}{\frac{i-3}{2}} + \binom{L-1}{\frac{i-3}{2}}\binom{N-L-1}{\frac{i-1}{2}} \qquad \text{for } i \text{ odd} \tag{8c}$$

The number statistic is defined as the probability of getting L or fewer residuals of the same sign and is given by

$$P(\leq L) = 2^{-N} \sum_{i=0}^{L+1} \binom{N}{i} \tag{9}$$

If $P(\leq r) < 0.05$ or $P(\leq L) < 0.05$, then the hypothesis that the residuals are random should be rejected (with α error < 0.05).

The analysis of the residuals is somewhat complicated by the fact that there will always be a correlation between the residuals because their number of degrees of freedom is less than their total number. FUNFIT also includes the Durbin–Watson statistic (38) given by

$$d = \sum_{i=2}^{N} (e_i - e_{i-1})^2 \Big/ \sum_{i=1}^{N} e_i^2 \tag{10}$$

to test for excessive serial correlation or systematic deviation among the residuals, e_i.

The normal deviate form of the residuals (24) is used in FUNFIT to detect outliers, that is, data points which, in a statistical sense, are not typical of the rest of the data (39). Outliers should be submitted to particularly careful examination since they may provide information of vital interest. They should be rejected only if they are caused by errors in recording or experimental technique.

If the residual analysis reveals that the residuals do not appear to be significantly random or normally distributed, then *this does not necessarily mean that the model is incorrect.* More exactly, it means one is faced with the problem of either rejecting the hypothesis that the model is "correct,"

rejecting the assumption made about the errors, or rejecting the assumption that the computer program has found the "best" solution in the case of multiple sum of squares minima. In the last two cases, the model cannot be verified. This clearly emphasizes the need for a computer program which is efficient in finding a global minimum, the need for accurate data to reduce or eliminate multiple minima, and the need for carefully designed experiments which do not introduce systematic or cumulative errors.

The Predictive Power of the Model

The ultimate goal in mathematical modeling in pharmacokinetics seems to be to establish models with significant predictive power. A similar goal exists in modeling of economic systems. The very voluminous literature in this area can undoubtedly give inspiration to future approaches in pharmacokinetics.

One of the best ways to test the predictive property of a model is to test the hypothesis underlying a model that the parameters do not depend on the model variables. This can readily be done if sufficient data are available. The total set of data is first partitioned into two or more subsets. The parameters are then estimated separately for each subset and the parameter subsets are tested for any trend or for a functional relationship with the independent variable(s) by a suitable correlation analysis. A test to establish whether the parameter subsets are significantly different from the parameter set obtained for the whole sample can also be employed (40,41). The highly interactive structure of FUNFIT readily facilitates such a partitioning of the data so that the above tests can be made.

Standard Deviation of a Quantity Which Is Expressed in Terms of the Parameters

A pharmacokinetic model which is fitted by least squares is usually expressed in terms of the macroparameters (e.g., equation 6) or the microparameters (rate constants, etc.) for which the regression program calculates an estimate of the variance–covariance matrix. However, the investigator may have a special interest in other pharmacokinetic quantities, which can be expressed in terms of the calculated parameters, and may want an estimate of their variability.

To give a simple example, it may be of interest to get an estimate of the variability of the total area under the curve which, assuming a two-compartment model [$c = p_1 \text{ EXP } (-p_2 t) + p_3 \text{ EXP } (-p_4 t)$], is

$$A = p_1/p_2 + p_3/p_4 \tag{11}$$

This can be done in two ways:

Use of Variance–Covariance Matrix

The variability can be estimated from elements of the variance–covariance matrix using the following general formula, which is based on a Taylor series expansion (34):

$$V(g) = \sum_{i,j=1}^{n} \left(\frac{\partial g}{\partial p_i}\right)\left(\frac{\partial g}{\partial p_j}\right) \mathrm{Cov}\,(p_i, p_j) \tag{12}$$

where $V(g)$ is the variance of a quantity, $g = g(p_1, p_2, \ldots, p_n)$, which is expressed in terms of n parameters and where the summation extends over all n^2 choices of the two indices i and j.

The formula is exact when g is linear in the parameters but is otherwise an asymptotic approximation which should be sufficiently accurate for most purposes, provided that the coefficient of variation of the parameters does not exceed about 20%. Equation 12 applied to equations 11 gives

$$\begin{aligned} V(A) = {} & \frac{V(p_1)}{p_2^2} + \frac{p_1^2 V(p_2)}{p_2^4} + \frac{V(p_3)}{p_4^2} + \frac{p_3^2 V(p_4)}{p_4^4} \\ & - \frac{2p_1 \mathrm{Cov}\,(p_1, p_2)}{p_2^3} + \frac{2\,\mathrm{Cov}\,(p_1, p_3)}{p_2 p_4} - \frac{2p_3 \mathrm{Cov}\,(p_1, p_4)}{p_2 p_4^2} \\ & - \frac{2p_1 \mathrm{Cov}\,(p_2, p_3)}{p_2^2 p_4} + \frac{2p_1 p_3 \mathrm{Cov}\,(p_2, p_4)}{p_2^2 p_4^2} - \frac{2p_3}{p_4^3} \mathrm{Cov}\,(p_3, p_4) \end{aligned} \tag{13}$$

However, the calculation of $V(A)$ using this expression will be subject to truncation errors because A is nonlinear in the parameters p_2 and p_4. Furthermore, the application of equation 12 to more complex expressions than equation 11 may easily produce very large expressions.

Transformation Technique

The transformation technique, which seems not to have been used in pharmacokinetic analysis, produces results which are without the truncation errors inherent in equation 12 and considerably more reliable.

Any of the four parameters in equation 11 can be expressed in terms of A and the remaining three parameters; as, for example,

$$p_1 = p_2(A - p_3/p_4) \tag{14}$$

This expression can then be substituted into equation 11, yielding

$$c = p_2(A - p_3/p_4)\, e^{-p_2 t} + p_3\, e^{-p_4 t} \tag{15}$$

The purpose of this transformation is to introduce the quantity of interest

(A) as a *formal* parameter, at the expense of another parameter (p_1), so estimates of its variability can be obtained directly from the computer.

This transformation technique can be used in all cases (either directly or by use of an iteration procedure) and is highly recommended. If it is difficult to get a good initial estimate of the quantity of interest, this can be done simply by first fitting the untransformed equation.

RESULTS AND DISCUSSION

FUNFIT was applied to obtain parameter estimates of the following simplified two- and three-compartment models:

$$c = p_1 e^{-p_2 t} + p_3 e^{-p_4 t} \qquad p_i > 0 \tag{16}$$

$$c = p_1 e^{-p_2 t} + p_3 e^{-p_4 t} + p_5 e^{-p_6 t} \qquad p_i > 0 \tag{17}$$

which were used to describe the plasma profile of pancuronium after intravenous bolus injection in four human subjects.

Parameter limits and initial estimates were chosen identical to those used in applying the 1969 version of NONLIN (1), which appears to be the most commonly used nonlinear regression program in pharmacokinetic investigations. The stopping criterion for FUNFIT was 0.1 (percent),[8] which gives approximately the same relative change in the SS value at convergence as NONLIN with its differently defined stopping criterion set at TEST = 0.0001. The step size used in NONLIN to approximate derivatives was DEL = 0.001, which is the same value as that used in the test problems given in the NONLIN user's manual (1). The experimental plasma levels were recorded to three significant digits after the decimal point so the precision factor in NONLIN was chosen as IDIG = −9 to avoid significant truncation of calculated values. In this way, NONLIN and FUNFIT should give identical SS values for identical parameter values and this was verified in all the runs.

FUNFIT found a different solution than NONLIN in all the cases where a three-compartment model (equation 17) was fitted, and in half of the cases where a two-compartment model (equation 16) was fitted (Table I). The residual sum of squares values obtained using FUNFIT were substantially lower than those obtained using NONLIN in all the cases where there was a difference. The average percentage difference was −55% and −29% when fitting equations 17 and 16, respectively. The differences were also reflected in the parameter values. Furthermore, the run test indicates that the residuals are overall more randomly distributed in the FUNFIT results.

[8]For details about convergence and expansion criteria, see Nelder and Mead (21).

Table I. Least-Squares Fitting of Equations 16 and 17 Using NONLIN (N) and FUNFIT (F)

Patient/ points		SS	ΔSS%[a]	Runs[b]	p_1	p_2	p_3	p_4	p_5	p_6
IA/12	N	$0.7131E-1$		6	$0.1452E+1$	0.2871	0.4553	$0.3022E-1$	0.3986	$0.4165E-2$
	F	$0.2410E-1$	−66.2	9	$0.4924E+1$	0.3026	$0.4081E-1$	0.1686	0.6013	$0.6632E-2$
	F[c]	$0.2869E-1$	−59.8	9	$0.2940E+1$	0.2370	$0.4855E-3$	$0.3568E-1$	0.6025	$0.6712E-2$
JC/11	N	$0.1825E-1$		8	$0.1038E+2$	0.9543	$0.1139E+1$	$0.7307E-1$	0.6039	$0.5048E-2$
	F	$0.6631E-2$	−63.7	9	$0.4649E+1$	0.5826	0.8223	$0.5242E-1$	0.5597	$0.4721E-2$
	F[c]	$0.6074E-2$	−66.7	9	$0.1303E+2$	0.7920	0.8369	$0.5411E-1$	0.5647	$0.4744E-2$
BA/10	N	$0.8973E-2$		6	0.2614	$0.4341E-1$	0.3327	$0.3473E-1$	0.1697	$0.1873E-2$
	F	$0.4653E-2$	−48.1	6	$0.5603E-1$	0.5668	0.5911	$0.2526E-1$	0.1041	$0.2758E-3$
	F[c]	$0.4447E-2$	−50.4	6	$0.7253E-1$	$0.2467E-1$	0.5253	$0.2478E-1$	$0.9702E-1$	$0.3492E-5$
MC/10	N	$0.1382E-1$		4	0.8224	$0.4488E-1$	$0.7388E-1$	$0.6326E-1$	0.3148	$0.2521E-2$
	F	$0.7844E-2$	−43.2	6	$0.1658E+1$	0.9200	0.9374	$0.3623E-1$	0.2283	$0.1936E-2$
	F[c]	$0.8346E-2$	−39.6	6	0.8740	$0.3600E-1$	$0.818E-1$	$0.6357E-1$	0.2336	$0.2035E-2$
IA/12	N	$0.5102E-1$		5	$0.1983E+1$	0.2367	0.6586	$0.7449E-2$		
	F	$0.2199E-1$	−55.9	8	$0.1066E+2$	0.3999	0.6202	$0.6858E-2$		
	F[c]	$0.2199E-1$	−55.9	9	$0.1049E+2$	0.3978	0.6198	$0.6853E-2$		
JC/11	N	$0.1621E-1$		7	$0.1240E+1$	$0.9424E-1$	0.6496	$0.5616E-2$		
	F	$0.1621E-1$	0	7	$0.1240E+1$	$0.9425E-1$	0.6497	$0.5616E-2$		
BA/10	N	$0.4528E-2$		6	0.5998	$0.2569E-1$	0.1012	$0.2100E-3$		
	F	$0.4446E-2$	−1.8	6	0.5975	$0.2470E-1$	$0.9677E-1$	$0.5866E-5$		
	F[c]	$0.4449E-2$	−1.8	6	0.5962	0.2484	$0.9835E-1$	$0.5402E-4$		
MC/10	N	$0.8194E-2$		6	0.9464	$0.3760E-1$	0.2373	$0.2093E-2$		
	F	$0.8194E-2$	0	6	0.9464	$0.3761E-1$	0.2374	$0.2095E-2$		

[a] $\Delta SS\% = 100(SS_{FUNFIT} - SS_{NONLIN})/SS_{NONLIN}$.
[b] See text for definition of runs.
[c] FUNFIT was in these cases started using NONLIN's final parameter estimate as initial estimates and with initial parameter step sizes equal to 0.1% of these parameter values.

To test NONLIN's results, FUNFIT was, in the cases where NONLIN gave different results, also started with initial parameter values identical to NONLIN's *final* estimates and with initial parameter step sizes equal to 0.1% of these parameter values. At the first iteration, FUNFIT gave exactly the same SS value as NONLIN's value at convergence, *but it did not accept this solution as a stationary point* and converged at a significantly different solution (Table I).

The detailed minimization report chosen in the investigation of this phenomenon, in fact, showed that the SS function in NONLIN's convergence region had a significant gradient value, indicating that NONLIN's solutions in these cases could not be considered to be sufficiently close to the true sum of squares minimum.

The most likely reason that NONLIN failed in these cases to find a satisfactory solution appears to be substantial errors in the approximation of derivatives.

The derivative with respect to the ith parameter, p_i, of the function $f(X, \mathbf{P})$ to be fitted in NONLIN is approximated by a one-sided difference formula:

$$\frac{\partial f}{\partial p_i} \simeq \frac{f(X, p_1, p_2, \ldots, p_i + \delta p_i, \ldots, p_n) - f(X, \mathbf{P})}{\delta p_i} \tag{18}$$

where the step size δ (DEL) is chosen by the user. The value of this quantity is critical for the accuracy of the derivative. In choosing a proper value for δ, one has to steer between two hazards: (a) If the value chosen for δ is too *small*, the derivative will be substantially inaccurate because of the *rounding error* which arises when the two f values in equation 18 are too close. (b) If δ is set too *large*, the derivative approximation will be too inaccurate because of the *truncation error* (equation 18 is accurate only in the limit as $\delta \to 0$).

This indicates that there must be an optimal value for δ. It can be shown that this value, for the ith parameter, is approximately given by

$$|\delta_i| \simeq \sqrt{\left|\frac{4\varepsilon f}{p_i H_{ii}}\right|} \tag{19}$$

where ε is the relative error in the computed f value and H_{ii} is the ith diagonal element of the Hessian matrix.[9] This formula shows that the optimal δ value differs from parameter to parameter. It is not uncommon to find a very large value for the ratio $\max(p_i H_{ii})/\min(p_i H_{ii})$, indicating that the choice of a *single* common δ-value as is done in NONLIN may not be good enough for all derivative evaluations and may cause convergence to a nonstationary point (18).

[9]The Hessian matrix is the symmetrical matrix of the second-order derivatives and is proportional to the inverse of the variance–covariance matrix.

The above problem can be overcome in several ways: (a) By abolishing difference approximations and using exact analytical derivatives. This, however, may be of considerable inconvenience for the user who must define the analytical derivatives. It also limits the use of the program to equations for which analytical derivatives can be obtained. (b) By modifying the initial choice of δ according to equation 19 or by other means (42). However, even if alterations are made according to (a) or (b) the Gauss–Newton methods may still converge in some cases to a point at which the gradient does not vanish (16). (c) By abolishing the linearization approach in the Gauss–Newton methods and using a general function minimization approach (43–46). (d) By the use of an algorithm which is not based on derivatives or derivative approximations as is done in FUNFIT. The disadvantage of the last approach is that more function evaluations are required to reach convergence. For most pharmacokinetic applications, this disadvantage is not significant. However, in cases where many parameters (more than about 12) are to be estimated or where the equation(s) to be fitted are very time consuming to evaluate (e.g., in the fitting of a functional relationship described by a system of differential equations) the disadvantage may become significant.

The presence of multiple SS minima in fitting the three-compartment model is evident from the fact that in the cases where FUNFIT was started with NONLIN's final estimate it converged on a different solution (Table I). The difference between SS values found by FUNFIT in consecutive runs was much smaller than the difference between NONLIN and FUNFIT's values.

To test for multiple minima in fitting the two-compartment model to IA's data, FUNFIT was started randomly ten times in the chosen parameter space. In nine of these cases it found the same solution (Table I), but in one case it converged to

$$\mathrm{SS} = 0.2209E-1 \qquad \text{and}$$

$$p_{1\text{–}4} = 0.8780E+1,\ 0.3744,\ 0.6162,\ 0.6797E-2$$

It is encouraging that this minimum is larger than that found in the $9+2$ other cases. The more frequent occurrence of different solutions in the three-compartment fitting confirms that the problem of multiple minima increases with an increasing number of parameters.

In only about half the cases investigated did FUNFIT and NONLIN find a smaller SS value for the three-compartment model than for the two-compartment model. This clearly emphasizes the problems in discriminating between nonlinear mathematical models as discussed.

If, in fitting linear compartmental models (equation 6), the lower limits for the coefficient parameters A_i are set to zero, then, in theory, the fit in

Table II. Least-Squares Fitting of Equations 16 and 17 Using NONLIN with Precision Factors −3 and −9

Patient/ points	Precision factor	SS	$\lvert\Delta SS\%\rvert$[a]	Runs[b]	p_1	p_2	p_3	p_4	p_5	p_6
IA/12	−9	$0.7131E-1$	48.7	6	$0.1452E+1$	0.2871	0.4553	$0.3022E-1$	0.3986	$0.4165E-2$
	−3	$0.3658E-1$		8	$0.1793E+1$	0.1807	0.3155	$0.1097E-1$	0.2949	$0.3697E-2$
JC/11	−9	$0.1825E-1$	35.2	8	$0.1028E+2$	0.9543	$0.1139E+1$	$0.7307E-1$	0.6039	$0.5048E-2$
	−3	$0.1183E-1$		6	$0.112E+1$	0.1813	0.5571	$0.3310E-1$	0.5060	$0.4237E-2$
BA/10	−9	$0.8973E-2$	68.8	6	0.2614	$0.4341E-1$	0.3327	$0.3473E-1$	0.1697	$0.1873E-2$
	−3	$0.1515E-1$		4	0.1396	$0.5172E-1$	0.4649	$0.4265E-1$	0.1734	$0.1953E-2$
MC/10	−9	$0.1382E-1$	582.0	4	0.8224	$0.4488E-1$	$0.7388E-1$	$0.6326E-1$	0.3148	$0.2521E-2$
	−3	$0.9427E-1$	—[c]	3	0.8000	0.6000	0.3000	0.1000	0.4000E−1	0.2000E−2
IA/12	−9	$0.5102E-1$	74.9	5	$0.1983E+1$	0.2367	0.6586	$0.7449E-2$		
	−3	$0.8924E-1$		4	0.8795	$0.4713E-1$	0.3312	$0.2706E-2$		
JC/11	−9	$0.1621E-1$	165.0	7	$0.1240E+1$	$0.9424E-1$	0.6496	$0.5616E-2$		
	−3	$0.4306E-1$		4	$0.1114E+1$	$0.4835E-1$	0.4251	$0.2649E-2$		
BH/10	−9	$0.4528E-2$	17.1	6	0.5998	$0.2569E-1$	0.1012	$0.2100E-3$		
	−3	$0.5303E-2$		6	0.5547	$0.3019E-1$	0.1610	$0.2121E-2$		
MC/10	−9	$0.8194E-2$	2.78	6	0.9464	$0.3760E-1$	0.2373	$0.2093E-2$		
	−3	0.8422		5	0.9226	$0.4107E-1$	0.2760	$0.2779E-2$		

[a] $\lvert\Delta SS\%\rvert = 100 \times \lvert SS_{(-9)} - SS_{(-3)}\rvert / SS_{(-9)}$.
[b] See text for definition of "runs."
[c] NONLIN failed to find a proper solution. At convergence after five iterations, the final parameter estimates were identical to the initial estimates.

terms of SS of a higher-order model (e.g., $n = 3$ vs. $n = 2$) should always be better or at least as good as the fit of a lower-order model.[10] Therefore, if under such conditions it is found that the higher-order model does not improve the fit (SS), then there are reasons to believe that a better minimum exists. In such cases, the higher-order model should be refitted with initial estimates of the common parameters equal to the final parameter estimates of the lower-order models. If this procedure also fails to give a lower SS value, then a third run should be made where the common parameters are restricted within narrow limits around the optima values for the lower-order model, and where the parameters in the higher-order term (e.g., p_5 and p_6 above) are much less restricted. The above approach will be successful in most cases and its use is recommended not only for fitting to sums of exponentials but also for any other "order system" where the terms are allowed to vanish. It should reduce significantly the problems associated with multiple minima and the problems of finding suitable initial estimates.

Truncation

In the NONLIN program, it is possible by using the IDIG parameter to specify various degrees of truncation of the calculated values for the dependent variable. For example, if IDIG is set at -3, then all calculated values of the dependent variable will be truncated to three significant digits after the decimal point. The philosophy behind the use of this parameter is that there is no reason to calculate the predicted values to any higher precision than the observed values. In adopting such a philosophy it must be realized that the results so obtained will be specific to the NONLIN program and in general cannot be compared with results obtained using other nonlinear regression programs. The difference between results obtained specifying virtually no truncation (IDIG $= -9$) and specifying truncation to the precision of the observations (IDIG $= -3$) was found to be very pronounced (Table II). The substantial difference was reflected not only in the SS values and parameter values but also in the randomness of the residuals.

Truncation (IDIG $= -3$) furthermore strongly affects the errors in the derivative approximations since the values of ε and hence δ in equation 19 will be affected. This may explain why NONLIN in one case (Table II) failed to converge properly.

The relative precision with which the parameter can be calculated will also be affected by truncation since it will not be possible to improve their estimates further when the computer has reduced the residual sum of

[10]Provided, of course, that the parameter space for the lower-order model is a subset of that of the higher-order model.

squares to a value, SS, for which

$$SS/SS(true) < 1 + \varepsilon \tag{20}$$

where the error, ε, depends on the degree of truncation chosen by IDIG (20). Furthermore, the truncation procedure cannot be justified on a statistical basis since the theory of least squares assumes no errors in the calculated values. In fact, great effort is often made to program the function to be fitted so that it can be evaluated with minimum errors to avoid biased results.

The user of the program NONLIN is therefore advised to use a value for the precision factor IDIG such that minimum truncation takes place.

The many nonlinear regression programs available have provided the scientist with a powerful tool useful for a great variety of problems. However, the results obtained have too often been accepted and used without an awareness of the limitations and possible unreliability of the program used, ignoring the numerical problems involved. The complex structure of the program used has often resulted in an authoritativeness which may go so far as using the program as a substitute for rational thought. It is to be hoped that these investigations will mobilize a higher degree of skepticism and a more modest interpretation of the results obtained. There is a definite need for programs which are more reliable and which allow greater interaction between the user and the program, with the user in a more dominant role.

ACKNOWLEDGMENT

I wish to thank Dr. Andrew A. Somogyi for the use of his pancuronium data.

REFERENCES

1. C. M. Metzler. *A User's Manual for NONLIN*, The Upjohn Co., Technical Report 7292/69/7292/005, Kalamazoo, Mich., 1969.
2. G. W. Booth and T. L. Peterson. *Nonlinear Estimation*, IBM Share Program, No. 687 WLNL7, 1958.
3. D. W. Marquardt. *Least Squares Estimation of Nonlinear Parameters*, IBM Share Library Program, Distribution No. 309401, 1966.
4. D. A. Meeter. *Program GAUSHAUS*, Numerical Analysis Laboratory, University of Wisconsin, Madison, 1966.
5. M. Pfeffer. COMPT: A time-sharing program for nonlinear regression analysis of compartmental models of drug distribution. *J. Pharmacokin. Biopharm.* **2**:137–163 (1973).
6. M. Berman and M. F. Weiss, Users Manual for SAAM, National Institutes of Health, Bethesda, Md., 1968.
7. J. M. Chambers. Computer system for fitting models to data. *Appl. Stat.* **18**:249–263 (1969).
8. R. Fletcher. *Report R6799*, Atomic Energy Research Establishment, Harwell, U.K., 1971.
9. G. L. Atkins. A versatile digital computer program for nonlinear regression analysis. *Biochim. Biophys. Acta* **252**:405–420 (1971).

10. W. Murray *et al.* Division of Numerical Analysis, National Physical Laboratory, Teddington, Middlesex, U.K.
11. J. L. Kuester and J. H. Mize. *Optimization Technique*, McGraw-Hill, New York, 1973, Chap. 6.
12. H. Eisenpress, A. Bomberault, and J. Greenstadt. *Nonlinear Regression Equations and Systems, Estimation and Prediction*, 7090 G2IBM0035 IBM, Hawthorne, N.Y.
13. F. S. Wood. *Nonlinear Least Squares Curve-Fitting Program*, 360D 13.6.007 IBM/360, 1971.
14. SHARE Program Library Agency. Triangle Universities Computation Center, P.O. Box 12076, Research Triangle Park, North Carolina 27709.
15. *Nonlinear Parameter Estimation and Programming.* Catalog of Programs for IBM System 360 Models 25 and Above, IBM, White Plains, N.Y.
16. M. J. D. Powell. A hybrid method for non-linear equations. In P. Rabinowitz (ed.), *Numerical Methods for Non-linear Algebraic Equations*, Gordon and Breach, London, 1970, pp. 87–114.
17. J. E. Dennis. Some computational techniques for the nonlinear least squares problem. In G. D. Byrne and C. H. Hall (eds.), *Numerical Solution of System of Nonlinear Algebraic Equations*, Academic Press, New York, 1973, pp. 157–183.
18. J. E. Dennis. Algorithms for nonlinear problems which use discrete approximations to derivatives. *Proc. ACM Natl. Conf. 1970*, ACM, New York. Also Tech. Rep. 71–98, Department of Computer Science, Cornell University, Ithaca, N.Y., 1971.
19. H. O. Hartley. The modified Gauss–Newton method for the fitting of nonlinear regression functions by least squares. *Technometrics* **3:**269–280 (1961).
20. J. M. Chambers. Fitting nonlinear models: Numerical techniques. *Biometrica* **60:**1–13 (1973).
21. J. H. Nelder and R. Mead. A simplex method for function minimization. *Comput. J.* **7:**308–313 (1965).
22. W. Spendley. Nonlinear least squares fitting using a modified simplex minimization method. In R. Fletcher (ed.), *Optimization*, Academic Press, New York, 1969, pp. 259–270.
23. Y. Bard. *Nonlinear Parameter Estimation*, Academic Press, New York, 1974.
24. N. R. Draper and A. Smith. *Applied Regression Analysis*, Wiley, New York, 1966.
25. H. G. Boxenhaum, S. Riegelman, and R. M. Elashoff. Statistical estimation in pharmacokinetics. *J. Pharmacokin. Biopharm.* **2:**123–147 (1974).
26. E. Kruger-Thiemer and R. R. Levine. The solution of pharmacological problems with computers part 8: Non first-order models of drug metabolism. *Arzneimittel-Forsch.* **18:**1575–1579 (1968).
27. J. Wagner. *Fundamentals of Clinical Pharmacokinetics*, Drug Intelligence Publications, Hamilton, Ill., 1975.
28. J. Wagner. Do I need a pharmacokinetic model, and, if so, which one? *J. Pharmacokin. Biopharm.* **3:**457–478 (1975).
29. J. Wagner. *Biopharmaceutics and Relevant Pharmacokinetics*, Hamilton Press, Hamilton, Ill., 1971.
30. M. Berman. The formulation and testing of models. *Ann. N.Y. Acad. Sci.* **108:**182–194 (1963).
31. G. C. Nooney. Mathematical models in medicine: A diagnosis. *J. Chron. Dis.* **19:**325–332 (1966).
32. M. Berman, E. Shahn, and M. F. Weiss. The routine fitting of kinetic data to models. *Biophys. J.* **2:**275–287 (1962).
33. J. Hiller. Exponential fitting and general technique of the Prony form. In R. S. Andersen and M. R. Osborne (eds.), *Data Representation*, University of Queensland Press, Queensland, Australia, 1970, pp. 62–76.
34. M. G. Kendall. *The Advanced Theory of Statistics*, Charles Griffin & Co., London, 1963.
35. C. Daniel and F. S. Wood. *Fitting Equations to Data*, Wiley, New York, 1971.
36. H. W. Killifors. On the Kolmogorov–Smirnov test for normality with mean and variance unknown. *J. Am. Stat. Assoc.* **62:**399–402 (1967).

37. F. S. Sweed and C. Eisenhart. Tables for testing randomness of grouping in a sequence of alternatives. *Ann. Math. Stat.* **14**:66–87 (1943).
38. J. Durbin and G. S. Watson. Testing for serial correlation in least squares regression. *Biometrica* **37**:409–428 (1950).
39. F. J. Anscombe. Rejection of outliers. *Technometrics* **2**:123–147 (1960).
40. W. G. Hunter and R. Mezaki. A model building technique for chemical engineering kinetics. *AIChE J.* **10**:315–322 (1964).
41. G. E. P. Box and W. G. Hunter. The experimental study of physical mechanisms. *Technometrics* d7:23–42 (1965).
42. G. W. Stewart. A modification of Davidon's minimization method to accept difference approximations of derivatives. *J. Assoc. Comput. Mach.* **14**:72–83 (1967).
43. R. Fletcher. A survey of algorithms for unconstrained optimization. In W. Murray (ed.), *Numerical Methods of Unconstrained Optimization*, Academic Press, New York, 1972, pp. 123–129.
44. R. Fletcher. A review of methods for unconstrained optimization. In R. Fletcher (ed.), *Optimization*, Academic Press, New York, 1969, pp. 1–12.
45. K. M. Brown and J. E. Dennis. *New Computational Algorithms for Minimizing a Sum of Squares of Non-linear Functions*, Yale University Report 71–6, 1971.
46. M. J. D. Powell. Problems related to unconstrained optimization. In W. Murray, *Numerical Methods for unconstrained Optimization*, Academic Press, New York, 1972, pp. 29–42.