

Some Biomedical Applications of a Non-Linear Curve Fit Method¹

James W. Tyson, B.S.,² James H. Meade, Jr., Ph.D.,³
Glenn V. Dalrymple, M.D.,⁴ and Horace N. Marvin, Ph.D.⁵

Little Rock, Arkansas

INTRODUCTION

In modern biology the investigator often wants to know if his experimental data behave according to some theoretical model. Usually a flow diagram of the model is the first step in the analysis. This process focuses attention on the mechanisms underlying the model. Since a flow diagram *per se* cannot be used for the analysis of data, the model must be given in the form of a mathematical equation. Simply writing an equation, however, does not complete the task. At this point the equation is in a general form. To be of any real value, a particular solution must be found that produces a good fit of the theoretical model to the observed data. Also, the values of the parameters must be consistent with the biological principles.

For simple statistical models such as straight lines, the familiar least squares method can be applied directly to give estimates of the parameters of the equation. As more sophisticated models are used to describe actual biological mechanisms, the equations become more involved. Since straight lines cannot usually be fitted to complex non-linear equations, the investigator often tries to find some transformation which will linearize his equations.⁶ After linearization, elementary methods can be used for the estimation of the parameters of the transformed equation. In other instances where a simple transformation cannot be found, graphical solutions give some estimates of the parameters in question.

¹This research was supported by grants FR-00208-02 and NIH-FRS350-05 from the National Institutes of Health and NASA 04-001-014.

²Junior Medical Student, University of Arkansas Medical Center, Little Rock, Arkansas.

³Associate Professor of Biometry, Univ. of Arkansas Medical Center, Little Rock, Arkansas.

⁴Assistant Professor of Radiology and Physiology, Univ. of Arkansas Medical Center, Little Rock, Arkansas.

⁵Professor and Chairman, Department of Anatomy, Univ. of Arkansas Medical Center, Little Rock, Arkansas.

⁶Perhaps the most commonly used transformation is the logarithm. If the data do not give a straight line on linear graph paper, the investigator will often reach for similog paper. Not uncommonly this is done without regard to the meaning of the log transformation on the process under study. Other transformations such as probits, logits, etc. are all too often used because they work, not because they are in any way related to the biological system.

Unfortunately, both transformations and graphical methods have serious limitations. One is never quite sure that assumptions made about the transformed variable can be transferred to the original variable. Also, the same criticism can be leveled against the calculated uncertainties about the parameters of the transformed equations. Purely graphical methods may be confounded by unintentional bias. Although the investigator tries to be as objective as possible, he may unconsciously draw the line to favor the desired result. Graphical methods do not usually allow the estimation of the uncertainties for each of the calculated parameters.

Ideally, an objective method which does not require transformations or graphs, but which can be successfully used for complex non-linear equations, would be of considerable value to the investigator. Toward this end a numerical technique has been applied to the estimation of the parameters of three rather complex equations. The selection of the equations for this paper was completely arbitrary—these equations are being studied in our laboratories.

The method, as described, is sufficiently general to be used for many different types of equations. The wide-spread availability of digital computers for calculations coupled with the simplicity of the method should give investigators a tool which can be readily used for fitting models to data.

SAMPLE EQUATIONS

The equations for the curves have the following general forms:

$$y = (1 - x/\beta) e^{-\alpha x} \quad (1a)$$

$$y = 1 - (1 - e^{-x/\beta})^\alpha \quad (2a)$$

$$y = \alpha/(1 + \beta/x) \quad (3a)$$

In each equation the y's are functions of x, parametric in α and β .

The first equation (1a) is used in hematology. Since the radioisotope Cr^{51} is tightly bound to red blood cells, the disappearance of this label from the blood provides an index of the biological loss of red cells from the circulation. By sampling venous blood at several time intervals, t , and then measuring the percent of the isotope remaining, P , a curve is developed which may be given by:

$$P(t) = (1 - t/L) e^{-Kt} \quad (1b)$$

Where L is the mean life span of the red cell population (days), K is the coefficient of random isotope loss per day (1). For this equation K corresponds to α and L to β in equation 1a.

The second equation (2a) is from radiobiology. A most important consequence of irradiation is that injured cells lose their ability to perform sustained and repeated mitoses. This loss of reproductive integrity can be used to study several of the effects of radiation (2). In the usual experiment (commonly performed with cultured mammalian cells such as HeLa and L cells), known numbers of cells are seeded onto petri dishes. These dishes are given graded single doses of X or γ -radiation and the cells then allowed to grow long enough to produce

macroscopic colonies (10 days to 2 weeks). The results are scored by determining the surviving fraction (the ratio of the number of colonies on the irradiated plates to the number of colonies on the non-irradiated control plates) as a function of radiation dose.

The surviving fraction as a function of dose D rads is given by:

$$S(D) = 1 - (1 - e^{-D/D_0})^n \quad (2b)$$

Where D_0 is the reciprocal of the log-linear portion of the curve, n is the extrapolation number (see Elkind and Sutton (3) for additional details). In equation 2a n corresponds to α and D_0 to β of eq. 2a. This equation is almost always solved by hand graphing. The points are plotted on semilog paper (see Fig. 3) and the points connected. The log linear portion of the curve is then extrapolated to the ordinate to give the extrapolation number n . The reciprocal of this log-linear portion is D_0 .

The third equation (1c) comes from biochemistry. In the more familiar form, it is recognized as the Michaelis-Menton equation (4). The equation gives the velocity of reaction, V , (moles reacted/unit time) as a function of substrate concentration (moles/unit volume).

$$V(S) = V_{\max}/(1 + K_m/S) \quad (3b)$$

Where V_{\max} is the maximal velocity of reaction and K_m is the Michaelis constant (the substrate concentration which gives a velocity equal to $1/2 V_{\max}$). For this equation V_{\max} corresponds to α and K_m to β of eq 3a.

Transformations of one form or another are usually used to solve this equation. One very common technique uses the reciprocals of the variables S and V

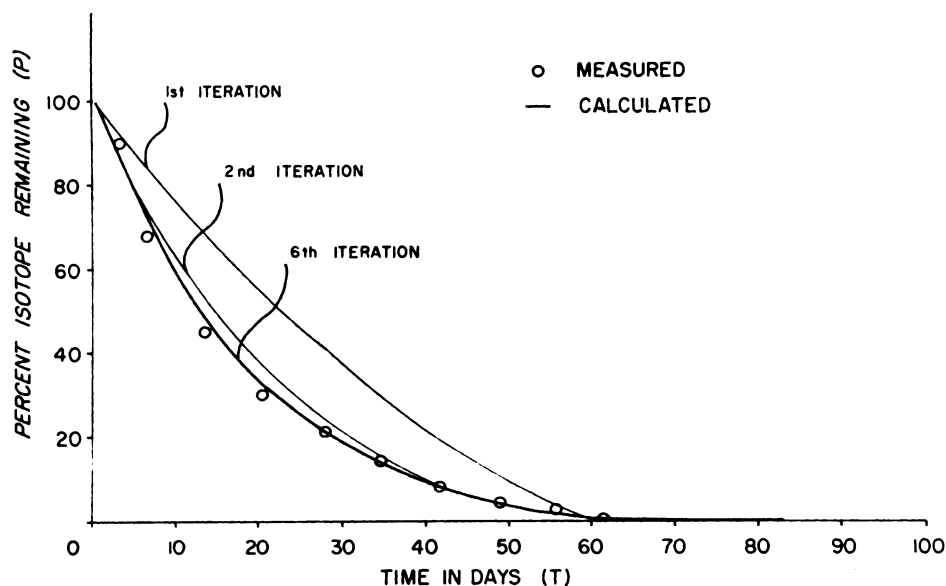


Fig. 1. ^{51}Cr red cell survival data. The fit of equation 1b to the measurements for iterations 1, 2, and 6 is shown

(4); An example is the Lineweaver-Burk in which $\frac{1}{V}$ is plotted against $\frac{1}{S}$ to give a straight line on linear graph paper. The parameter estimates are given by the Y intercept of the curve which is equal to $\frac{1}{V_{\max}}$ and the x intercept which is equal to $\frac{1}{K_m}$.

A detailed study of the Lineweaver-Burk method (as well as other similar methods) showed that using reciprocals can be a source of error *per se*, even if the curves are fitted to the reciprocals by least squares (5).

Therefore, the three equations are examples of equations which must be fitted by some numerical method (eq. 1b), equations which are commonly fitted by graphical means (eq. 2b), and equations which require transformations for linearization (eq. 3b).

This paper will show that these equations can all be fitted by the same basic method. A small digital computer system was used for the calculations.

THE NUMERICAL METHOD

Since in the models just described the parameters enter non-linearly, general linear regression theory cannot be used to estimate them. However, one can "linearize" by expanding a Taylor's series about some initial set of guess estimates¹, truncating the higher order terms, estimating the corrections by least squares, and then adjusting the guess estimates with the corrections. By repeating the process until the corrections become zero (or as close to zero as desired for the particular applications), a least squares fit of the model to the data is achieved.

For ease of presentation the theory below will be for the case in which the response variable (y) is a function of only one independent variable (x) with two parameters (a, b). The general case follows in the same manner (6, 7). The Model:

$$Y_i = f(X_i : a, b) + e_i \quad (4)$$

Where,

Y_i = the response of the i^{th} observation

X_i = the value of the i^{th} measurement

a, b = the parameters to be estimated

e_i = the deviation of the i^{th} response from the model

$f(x : a, b)$ = the particular model describing the biological process

and

$i = 1, 2, \dots, n$; where n is the number of observations on X and Y.

If the e_i are normally and independently distributed, the least squares estimates are the same as Maximum Likelihood Estimates (6). Using a_o and b_o as

¹The initial estimates can be obtained a number of ways. We usually plot the data on graph paper and then make a very rough visual estimate of the parameters. An alternate method is to start with the best available graphical or transformation method and then use the results of these calculations as the initial values. Our experience has been that the rough visual estimates are almost always satisfactory.

guess estimates of a and b , expanding $f(x : a, b)$ about a_o and b_o in a Taylor's series, and truncating after the first order terms we obtain:

$$f(x : a, b) \cong f(x : a_o, b_o) + \gamma \left(\frac{\partial f}{\partial a} \right)_o + \delta \left(\frac{\partial f}{\partial b} \right)_o \quad (5)$$

where,

$$\left(\frac{\partial f}{\partial a} \right)_o \text{ and } \left(\frac{\partial f}{\partial b} \right)_o$$

mean the partial derivative of $f(x : a, b)$ with respect to the parameter evaluated at a_o, b_o and the particular value of X ; $f(x : a_o, b_o)$ means the value of the function evaluated at a_o, b_o and the particular value of X .

Our model now becomes:

$$Y_i = f(X_i : a_o, b_o) + \gamma \left(\frac{\partial f}{\partial a} \right)_o + \delta \left(\frac{\partial f}{\partial b} \right)_o + e_i \quad (6)$$

or, by rearrangement

$$Y_i - f(X_i : a_o, b_o) = \gamma \left(\frac{\partial f}{\partial a} \right)_o + \delta \left(\frac{\partial f}{\partial b} \right)_o + e_i$$

Which is linear in the corrections γ and δ and which can be fitted by least squares using standard linear regression theory.

Letting:

$$Y_i^* = Y_i - f(X_i : a_o, b_o)$$

$$X_{i1}^* = \left(\frac{\partial f}{\partial a} \right)_o$$

$$X_{i2}^* = \left(\frac{\partial f}{\partial b} \right)_o$$

equation 6 then becomes

$$Y_i^* = \gamma X_{i1}^* + \delta X_{i2}^* + e_i$$

The value to be minimized is

$$\sum e_i^2 = \sum (Y_i^* - \gamma X_{i1}^* - \delta X_{i2}^*)^2 \quad (7)$$

The solution which minimizes equation 7 is in the matrix notation

$$B^* = (X'X)^{-1} X'Y$$

Where:

$$Y = \begin{bmatrix} Y_1^* \\ Y_2^* \\ \vdots \\ Y_n^* \end{bmatrix}, B^* = \begin{bmatrix} \gamma \\ \delta \end{bmatrix}, X = \begin{bmatrix} X_{11}^* & X_{21}^* \\ X_{21}^* & X_{22}^* \\ \vdots & \vdots \\ X_{n1}^* & X_{n2}^* \end{bmatrix}$$

X' is the transpose of x and $(X'X)^{-1}$ is the inverse of $(X'X)$.

The new estimates of a and b then become

$$\begin{aligned} a_1 &= a_o + \gamma \\ b_1 &= b_o + \delta \end{aligned}$$

Using a_1 and b_1 as new estimates the above process is repeated until γ and δ become as small as desired.

The variance of a , $V(a)$, is estimated by $C_{11}\sigma^2$, $V(b)$ by $C_{22}\sigma^2$ where C_{11} , C_{22} are the diagonal elements of the inverse matrix $(X'X)^{-1}$ and σ^2 is the residual variance $\sum e_i^2/(n-2)$ (that is, the variance of the deviations from the model).

NUMERICAL EXAMPLE

Equation 1b will be used for purposes of demonstration.

Recall that this equation has the form:

$$P = (1 - t/L)e^{-kt} \quad (8)$$

The necessary partial derivatives are then:

$$\left(\frac{\partial P}{\partial K}\right) = -te^{-kt} + t^2 e^{-kt}/L \quad (9)$$

and

$$\left(\frac{\partial P}{\partial L}\right) = te^{-kt}/L^2 \quad (10)$$

The deviation of the calculated values from the measured results are given by $(P_i - \hat{P}_i)$. The model in terms of the corrections becomes

$$(P_i - \hat{P}_i) = \gamma \left(\frac{\partial P_i}{\partial K}\right)_o + \delta \left(\frac{\partial P_i}{\partial L}\right)_o + e_i \quad (11)$$

Where P_i and \hat{P}_i mean the observed and the calculated fraction of Cr^{51} remaining at the i^{th} observation; $\left(\frac{\partial P_i}{\partial K}\right)_o$ and $\left(\frac{\partial P_i}{\partial L}\right)_o$ refer to the values of the partial derivatives evaluated at the i^{th} observation with the initial parameter estimates, K_o and L_o .

The observed values [taken from Marvin (9)] are entered in the first two columns of Table 1. The values of $\left(\frac{\partial P}{\partial K}\right)_o$, $\left(\frac{\partial P}{\partial L}\right)_o$, the deviations, and the squares of the deviations are listed in the remaining columns. For the first iteration $L_o = 60$, and $K_o = 0.01$. The necessary sum and the additional cross products are included at the bottom of the table.

From equation 11 normal equations are constructed for the estimation of γ and δ . These are formed by defining

$$G = \sum \left[(P_i - \hat{P}_i) - \gamma \left(\frac{\partial P}{\partial K}\right) - \delta \left(\frac{\partial P}{\partial L}\right) \right]^2$$

after differentiating G with respect to γ and δ and then setting the derivatives equal to zero, the following are developed:

$$\sum \left[\frac{\partial P}{\partial K} (P_i - \hat{P}_i) \right] = \gamma \sum \left(\frac{\partial P}{\partial K} \right)^2 + \delta \sum \left(\frac{\partial P}{\partial K} \frac{\partial P}{\partial L} \right) \quad (12)$$

$$\sum \left[\frac{\partial P}{\partial L} (P_i - \hat{P}_i) \right] = \gamma \sum \left(\frac{\partial P}{\partial K} \frac{\partial P}{\partial L} \right) + \delta \sum \left(\frac{\partial P}{\partial L} \right)^2 \quad (13)$$

after substituting the values obtained from Table I,

TABLE I
NUMERICAL EXAMPLE

$$L_o = 60.0$$

$$K_o = 0.01$$

First Iteration

Days after Inject.	Observed Fct. Cr ⁵¹ Remaining	Calculated Fct. Cr ⁵¹ Remaining				
(t)	(P _i)	(\hat{P}_i)	(P _i - \hat{P}_i)	(P _i - \hat{P}_i) ²	($\partial P / \partial K$) _o	($\partial P / \partial L$) _o
3.0	.895	.922	-0.02692	.00072	-2.76577	0.00081
7.0	.678	.824	-0.14561	.02120	-5.76530	0.00181
14.0	.450	.667	-0.21651	.04688	-9.33111	0.00338
21.0	.300	.527	-0.22688	.05147	-11.06448	0.00473
28.0	.206	.397	-0.19708	.03884	-11.28637	0.00588
35.0	.138	.294	-0.15562	.02421	-10.27670	0.00685
42.0	.084	.194	-0.11311	.01279	-8.27879	0.00767
49.0	.048	.122	-0.06431	.00414	-5.50343	0.00834
56.0	.034	.038	-0.00408	.00002	-2.13251	0.00889
63.0	.016	.012	0.04263	.00182	1.67767	0.00932

$$\begin{aligned} \sum \left[\frac{\partial P}{\partial K} (P_i - \hat{P}_i) \right] &= 10.638791 & \sum \left[\frac{\partial P}{\partial L} (P_i - \hat{P}_i) \right] &= -0.005358 \\ \sum \left(\frac{\partial P}{\partial K} \right)^2 &= 589.561440 & \sum \left(\frac{\partial P}{\partial L} \right)^2 &= 0.000413 \\ \sum \left(\frac{\partial P}{\partial K} \frac{\partial P}{\partial L} \right) &= -0.345969 & \sum (P_i - \hat{P}_i)^2 &= 0.20209 \end{aligned}$$

TABLE II
NUMERICAL EXAMPLE—SUMMARY

Iteration No.	K	L	Std error K	Std error L	Residual Variance
1	.010	60.000	.00918	10.96003	.02526
2	.031	64.211	.00484	9.85763	.00136
3	.041	79.096	.00360	11.23441	.00031
4	.042	80.513	.00410	16.28042	.00029
5	.042	80.533	.00418	17.36419	.00029
6	.042	80.534	.00418	17.37728	.00029

$$(10.638791) = \gamma (589.561440) + \delta (-0.345969) \quad (14)$$

$$(-0.005358) = \gamma (-0.345969) + \delta (0.000413) \quad (15)$$

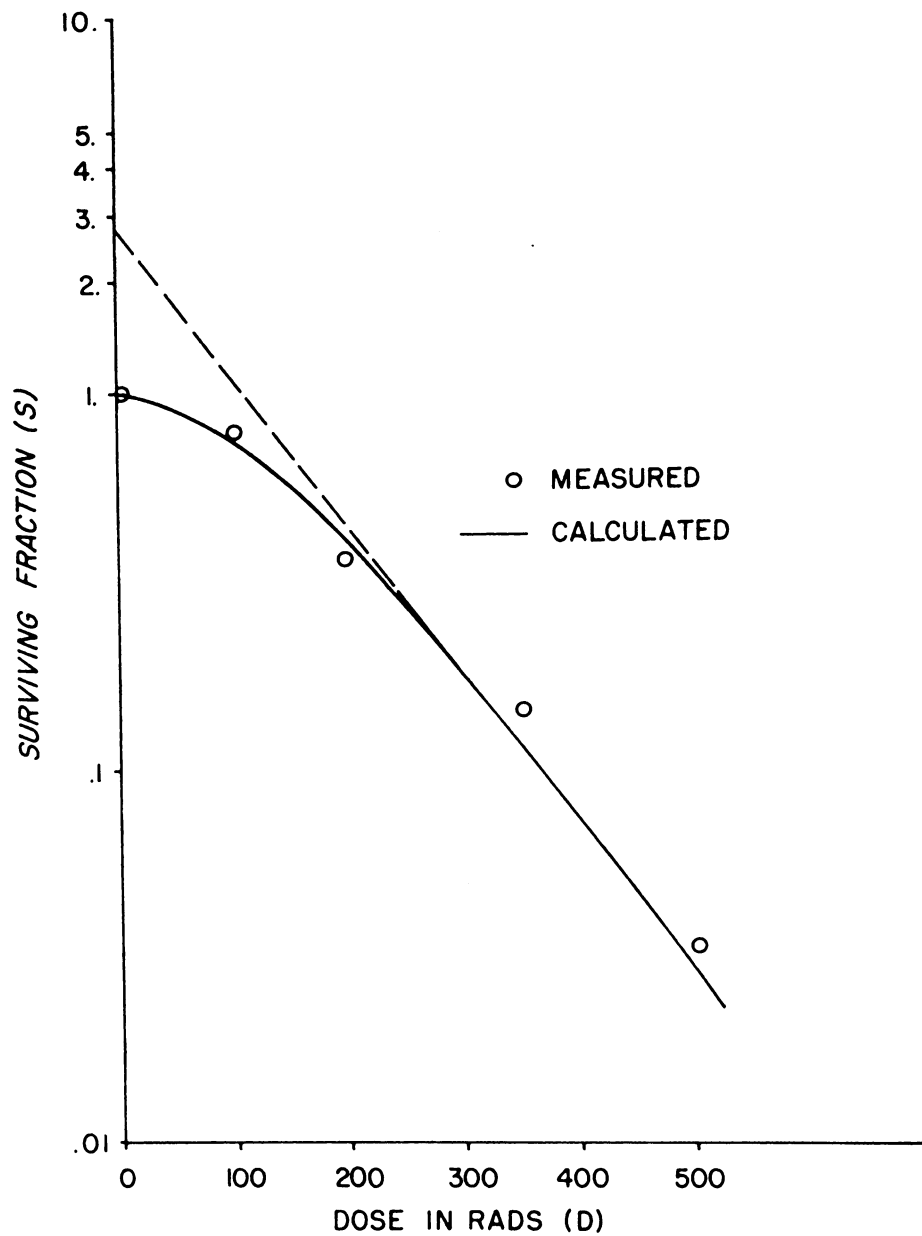


Fig. 2. Irradiated L cell survival data. The fit of equation 2b is shown.

Solution of equations 14 and 15 gives

$$\gamma = 0.02052 \quad \delta = 4.21078$$

Therefore the new estimates of K and L are

$$K_1 = K_o + \gamma = 0.01 + 0.02052 = 0.03052$$

$$L_1 = L_o + \delta = 60. + 4.21078 = 64.21078$$

The residual sum of squares is calculated by summing the fifth column in Table I.

The residual variance, σ^2 , is given by:

$$\sigma^2 = \sum (P_i - \hat{P}_i)^2 / (N - 2) = \frac{0.20209}{8} = 0.02526$$

Where n is the number of observations.

The variance of K, is calculated by:

$$\begin{aligned} V(K) &= \left\{ \sum \left(\frac{\partial P}{\partial L} \right)^2 / \left[\sum \left(\frac{\partial P}{\partial K} \right)^2 \sum \left(\frac{\partial P}{\partial L} \right)^2 - \left[\sum \left(\frac{\partial P}{\partial K} \frac{\partial P}{\partial L} \right) \right]^2 \right] \right\} (\sigma^2) \\ &= (0.00333) (0.02526) \\ &= 8.41158 \times 10^{-5} \end{aligned}$$

The standard error of K = $\sqrt{V(K)} = 0.0092$

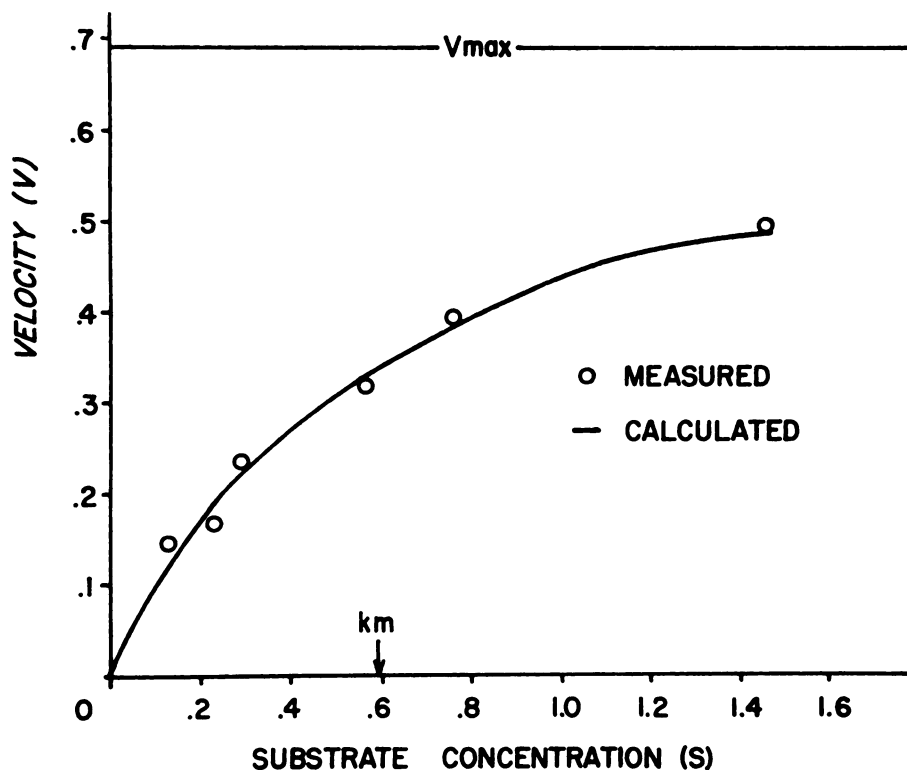


Fig. 3. Michaelis-Menton kinetics data. The fit of equation 3b is shown.

The variance of L is similarly calculated by:

$$\begin{aligned} V(L) &= \left\{ \sum \left(\frac{\partial P}{\partial K} \right)^2 / \left[\sum \left(\frac{\partial P}{\partial K} \right)^2 \sum \left(\frac{\partial P}{\partial L} \right)^2 - \left[\sum \left(\frac{\partial P}{\partial K} \frac{\partial P}{\partial L} \right) \right]^2 \right] \right\} (\sigma^2) \\ &= (4754.8749) (0.02526) \\ &= 120.1221 \end{aligned}$$

The standard error of $L = \sqrt{V(L)} = 10.96003$

The calculations are then repeated using K_1 and L_1 as the estimates of K and L .

The results of the successive iterations are given in Table II. After the sixth iteration, the correction factors γ and δ become as small as desired (less than 1×10^{-5}) and iteration using the computer is abandoned.

RESULTS AND DISCUSSION

Figure 1 shows the fit of the model to the data given by equation 1b on iterations 1, 2 and 6. The actual parameter values are listed in Table II. While the initial estimates cause the curve to miss the points rather badly, subsequent iterations rapidly pulled the curve into place. Although six iterations were required to reduce the correction below 1×10^{-5} , almost no difference can be seen between iterations 3 and 6. Figure 2 (the computer output) lists the final fit of the model to the data.

TABLE III

COMPUTER OUTPUT FOR ^{51}Cr RED CELL SURVIVAL.
THE SIXTH (FINAL) ITERATION IS SHOWN.

RED BLOOD CELL CURVE

ITERATION NUMBER 6

<i>Time (Days)</i>	<i>Fct. Isotope Remaining</i>	<i>Calculated Fct. Isotope Remaining</i>	<i>Deviation</i>
3.0	.8950	.8500	.0450
7.0	.6780	.6826	— .0046
14.0	.4500	.4618	— .0118
21.0	.3000	.3089	— .0089
28.0	.2060	.2038	.0022
35.0	.1380	.1321	.0059
42.0	.0840	.0836	.0004
49.0	.0480	.0511	— .0031
56.0	.0340	.0297	.0043
63.0	.0160	.0159	.0001

Residual variance 0.0002916

K 0.042 L 80.533

Std error K 0.0042 Std error L 17.378

Figure 3 shows the curve of equation 2b as fitted to some cultured cell data from one of our laboratories. Figure 4 lists the calculated and observed post-irradiation surviving fractions for this experiment along with the residual variance and the parameter standard errors. Figure 5 gives the final fit of equation 3b to some data published by Wilkinson (8). Figure 6 contains the observed and the calculated velocities of reaction together with the residual variance and the standard errors.

These three somewhat diverse examples indicate that the same numerical process can be effectively used for several different models.

Convergence of solution occurred for each of the examples just given. By this we mean that with each succeeding iteration, the parameter estimates approached some stable value and the residual variance approached a minimum. The final iteration gives the best estimates of the parameters of the model (and their associated standard errors) as fitted to the set of data being studied. Assuming that one has an adequate and proper model to describe the biological system, a set of well-behaved data, convergence of the solution, and a sufficiently small residual variance¹, the investigator should have grounds to state that his data are fitted by the model.

On occasions non-convergence may occur. There are several possible causes for non-convergence. Perhaps the most important reason is the wrong model. Although an improper or incorrect model can cause non-convergence, initial non-convergence is not *a priori* evidence that the model is wrong. Such factors as

¹Exactly how small the residual variance needs to be is a question that must be answered by the individual investigator for his particular problem. Obviously, what is a good fit for one investigator's problem may not be good enough for another. That is to say, goodness of fit can only be satisfactorily evaluated in light of the biological process itself and the conditions under which the experiment is run.

TABLE IV

COMPUTER OUTPUT FOR L CELL DATA.

CELL SURVIVAL CURVE ITERATIVE METHOD

EXPERIMENT III L CELLS PE = 0.50 2/18/66

ITERATION NUMBER 14

<i>Dose (Rads)</i>	<i>Surviving Fraction</i>	<i>Calculated Surv. Fct.</i>	<i>Deviation (Data-Calc)</i>
0.00000	0.99990	1.00000	-0.00010
100.00000	0.79000	0.77025	0.01975
200.00000	0.36000	0.39537	-0.03537
350.00000	0.15000	0.11278	0.03722
500.00000	0.03400	0.02948	0.00452

Residual variance 0.001013

N 2.8694 DO 109.4451

Std error of N 0.1955 Std error of DO 4.2614

heterogeneous data, insufficient computer accuracy, or vastly incorrect initial estimates may cause non-convergence.

The lack of a sufficient number of experimental points to define the entire structure of the curve is a frequent cause of non-convergence. Not uncommonly we have found that while a given set of data produced non-convergence, a repeat of the experiment with additional points gave convergence.

If repeated trials fail to result in convergence, the method described in this paper cannot be used for estimating the model parameters and the investigator must seek other avenues of analysis.

SUMMARY

A numerical method, applicable to some of the problem of non-linear curve fitting as it occurs in biology and medicine is described. Three different equations, theoretical considerations and a sample calculation are described. Given a proper model to describe the biological system, well-behaved data, convergence of solution and a sufficiently small residual variance, the investigator has objective justification that his data have been fitted by the model.

TABLE V

COMPUTER OUTPUT FOR THE DATA OF FIGURE 3.

MICHAELIS-MENTON KINETICS

ITERATION NUMBER 3

<i>Substrate Concentration</i>	<i>Velocity</i>	<i>Calculated Velocity</i>	<i>Deviation</i>
0.13800	0.14800	0.12971	0.01829
0.22000	0.17100	0.18601	-0.01501
0.29100	0.23400	0.22636	0.00764
0.56000	0.32400	0.33429	-0.01029
0.76600	0.39000	0.38813	0.00187
1.46000	0.49300	0.49013	0.00287

Residual variance 0.000184

Km 0.596535 Vmax 0.690393

Std error Km 0.068256 Std error Vmax 0.036824

REFERENCES

1. BERLIN, N. I., WALDMANN, T. A., AND WEISSMAN, S. M.: *Physiological Review* 39:577-609, 1959.
2. PUCK, T. T., AND MARCUS, P. J.: *J. Exptl. Med.* 103:653-666, 1956.
3. ELKIND, M. M., AND SUTTON, H.: *Radiat. Res.* 13:556-593, 1960.
4. DIXON, M., AND WEBB, E. C.: *Enzymes*, Academic Press, New York, 1958.
5. DOWD, J. E., AND RIGGS, D. S.: *J. Biol. Chem.* 240:863-869, 1965.
6. LUCAS, H. L., JR.: Personal Correspondence. North Carolina State University, 1962.
7. SCARBOROUGH, J. B.: *Numerical Mathematical Analysis*, Johns Hopkins Press, 1962.
8. WILKINSON, G. N.: *Biochem. J.* 80:324-332, 1961.
9. MARVIN, H. N.: *J. Nutrition* 80:185-190, 1963.