ABSTRACT


A kinetic model is generally expressed in differential equations. These equations may not be solved analytically. To establish these equations the parameters or coefficients must be estimated directly from experimental data. Innumerable procedures, such as nonlinear least squares, absolute least deviation, quasilinearization, and invariant imbedding have been proposed and used to solve the nonlinear estimation problems.

The purpose of this paper is to demonstrate that SAS - GLM can also be used to provide good parameter estimates through linear least squares, based on the estimation procedure of Hwang and Yeh. Numerical example in a two-compartment open model in clinical pharmacokinetics is given and the results of a comparison with the SAS - NLIN are presented.

INTRODUCTION

Hwang and Yeh (1977) proposed a simple numerical procedure for estimating parameters or coefficients arising in kinetic modeling. The procedure is quite different from the classical ones, such as nonlinear least squares, absolute least deviation, quasilinearization, and invariant imbedding. It involves variable transformation, numerical integration of the differential equations using experimental data directly, and linear least squares. It requires neither initial estimates nor repetitive iterations for convergence. The parameters or coefficients are determined when the least squares estimates are obtained.

Three SAS (Barr, et al., 1976) procedures (i.e., GLM, MATRIX, and NLIN) are used. The GLM procedure uses the principle of linear least squares to fit a general fixed effects linear model to data. When it is used together with the MATRIX procedure, data of almost any type can be handled. MATRIX is a comprehensive matrix handling procedure. It fills the gap between using packaged statistical procedures and writing custom-made programs. Using statements in MATRIX is virtually as efficient and flexible as using APL (Iverson, 1962). The NLIN procedure produces nonlinear least squares estimates of parameters or coefficients of a nonlinear model in algebraic form. It first performs a grid search to determine the starting values for the parameters or coefficients to be estimated. Then, it uses one of the three iterative methods (i.e., a modified Gauss-Newton method, the Marquardt method, or the steepest-descent method) to find the nonlinear least squares estimates of the parameters of coefficients.

Based on the linear least squares approach of Hwang and Yeh, the SAS procedures of GLM and MATRIX are used to estimate parameters in a two-compartment open model in clinical pharmacokinetics. Numerical results compared to those of SAS - NLIN indicate that the linear least squares approach using SAS - GLM is effective.

TWO-COMPARTMENT OPEN MODEL

Consider the following differential equations which represent a two-compartment open model following an IV (intravenous) bolus administration (Greenblatt and Kock-Weser, 1975; Hwang and Yeh, 1977):

\[
\frac{dc_1}{dt} = -(k_{12} + k_e) c_1 + k_{21} c_2 \tag{1}
\]

\[
\frac{dc_2}{dt} = k_{12} c_1 - k_{21} c_2 \tag{2}
\]

where

- \( c_1 \) and \( c_2 \) = drug concentrations in central and peripheral compartments, respectively; \( c_2 \) is not measurable, in general,
- \( k_{12} \) and \( k_{21} \) = first-order rate constants of drug transfer between the two compartments,
- \( k_e \) = first-order rate constant of drug elimination from the central compartment, and
- \( t \) = time elapsed after the drug administration.

The initial conditions when an IV bolus is administered are

\[
c_1(0) = D / V_1 = c_1^0 \tag{3}
\]

\[
c_2(0) = 0 \tag{4}
\]

where

- \( D \) = the single IV dose, and
- \( V_1 \) = the apparent volume of distribution in the central compartment.

The analytical solution can be obtained through Laplace transforms to yield an explicit double-exponential function relating \( c_1 \) and \( c_2 \) to time \( t \) as follows:

\[
c_1 = A e^{-at} + B e^{-bt} \tag{5}
\]
where
\[ k_c = \frac{(A + B)}{(A / a + B / b)} \]
\[ k_{21} = \alpha / k_e \]
\[ k_{12} = (A / B (A + B)) (\beta - a)^2 / k_{21} \]

To estimate the unknown rate constants (i.e., \( k_c, k_{21}, \) and \( k_{12} \)), serum samples are drawn at carefully selected time points after drug administration, assayed for drug concentration, and the parameters are then estimated from the resulting serum concentration-time curves.

**ESTIMATION PROCEDURE - LINEAR LEAST SQUARES**

From Equations (1) - (4), the following can be established:

\[ \frac{dC_1}{dt} = -(k_{12} + k_e)C_1 + k_{21} (C_0 - C_1 - k_e \int_0^t C_1 (u) du) \]
\[ = k_{21} C_0 - (k_{12} + k_e + k_{21}) C_1 \]
\[ - k_{21} k_e \int_0^t C_1 (u) du \]  

(7)

Next, proceed with the numerical integration and place the constant term \( C_0 \) to the left-hand side of the equal sign to obtain the following equation

\[ C_1(t_i) = k_{21} C_0 t_i \]
\[ - (k_{12} + k_e + k_{21}) \int_0^t C_1 (u) du (8) \]
\[ - k_{21} k_e \int_0^{t_i} \int_0^s C_1 (u,s) du ds \]

Then, let
\[ y_1 = C_1(t_i), \]
\[ x_1 = \int_0^{t_i} C_1 (u) du, \]
\[ x_2 = \int_0^{t_i} \int_0^s C_1 (u,s) du ds, \]
\[ x_3 = k_e \int_0^{t_i} \int_0^s C_1 (u,s) du ds \]

(9)

and the following linear regression equation with zero intercept is obtained:

\[ E(y_1) = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 \]

(10)

The estimates for the three rate constants are:
\[ k_{21} = b_1 / C_0 \]
\[ k_e = - b_3 / k_{21} = - C_1 (b_3 / b_1) \]
\[ k_{12} = - (b_2 + k_{21} + k_e) \]

To obtain values for \( x_1, \) and \( x_n, i = 1, 2, ..., n \) in Equation (9), numerical integration scheme is required. For this work the method of cubic spline interpolation (Dunfield and Read, 1972) is used. The cubic spline is an interpolation method giving a smooth curve passing through all the data points. The method consists of the piece-wise fitting of cubics between each pair of data points in such a manner that the two cubics passing through any particular point have the same functional value, first and second derivatives. The third derivative is not required to be continuous. The spline can be considered to follow the data accurately and to give the best approximation to the function represented by the data and hence its derivatives.

**NUMERICAL EXAMPLE**

The following numerical values in (12) and Table 1 are used:

\[ k_{21} = 0.547 hr^{-1}, k_e = 0.25 hr^{-1}, \]
\[ k_{12} = 2.29 hr^{-1}, \]  
\[ C_1(0) = 4.2 \text{ ng/ml}. \]  

(12)

The numerical values for \( C_1(t_j), i = 1, 2, ..., 18 \) were serum concentrations of chlordiazepoxide and its active metabolite, demethylchlordiazepoxide, after administration of a 28 mg IV bolus of chlordiazepoxide HCI, equivalent to 25 mg of chlordiazepoxide base, to a healthy 75 kg male volunteer subject (Greenblatt and Koch-Weser, 1975).

Numerical data with outliers of approximately 15% off the true values (Table 2) are also introduced to test the effectiveness of the linear least squares approach using SAS - GLM versus the nonlinear least squares using SAS - NLIN.

**NUMERICAL RESULTS**

The linear least squares procedure using SAS - GLM and MATLAB (LLS - GLM) is approached directly from the differential equations in Equations (1) and (2), while the nonlinear least squares using SAS - NLIN (NLS - NLIN) is started from the algebraic equation in Equation (5). The SAS - NLIN procedure is limited to solving nonlinear estimation problems with models expressed in a single algebraic equation only.

The results from LLS - GLM using cubic splines are compared with those from NLS - NLIN using the Marquardt Method (Table 3). The LLS - GLM yielded parameter estimates \( k_{21} = 0.8462, k_e = 0.22459, \) and \( k_{12} = 2.29523 \) in 1 iteration in good agreement with the true values \( k_{21} = 0.847, k_e = 0.225, \) and \( k_{12} = 2.29, \) as well as with those from NLS - NLIN \( (E_{21} = 0.84801, E_e = 0.22492, \) and \( k_{12} = 0.2963 \) in 3 to 11 iterations depending on the initial estimates).
When the outlier in Table 2 are introduced into the experimental data, the results for the different patterns obtained for both LLS - GLM and NLS - NLIN are presented in Table 4. It can be seen that the estimated values of \( k_1 \), \( k_2 \), and \( k_3 \) are slightly off the given true values in all outlier patterns whether using LLS - GLM or NLS - NLIN. The underestimation or overestimation is small in magnitude. In any event, in the LLS - GLM case, neither negative estimates nor convergence problems are observed.

Comparable results from both LLS - GLM and NLS - NLIN are also obtained (not shown) when only 9 data points (i.e., \( t_i = 0, 0.5, 1.0, 2.0, 4.0, 8.0, 16.0, 24.0, \) and 48.0 hours postdrug) are used. The magnitude of underestimation or overestimation is slightly greater than that in the case of 18 data points, when the same outlier patterns are imposed upon the data.

DISCUSSIONS

The linear least squares approach of Hwang and Yeh using SAS - GLM appears to be as effective as the nonlinear least squares using SAS - NLIN, in the modeling of a two-compartment open model in clinical pharmacokinetics. The two-compartment open model is chosen for this illustration, simply because the model can be worked in both ways (i.e., using differential equations for LLS - GLM and using an analytical solution in an algebraic equation for NLS - NLIN). This allows a direct comparison in the effectiveness of these two approaches. However, in model situations where the differential equations cannot be solved analytically, the change of concentrations may not be completely measurable, or the equations may not be directly differentiable, the LLS approach using SAS - GLM with MATRIX is still effective.

Variance estimates of parameters are not attempted for the two-compartment open model. Because the observed \( y_i \)s in (9) are taken over a successive sequence of times, they are not independent as required for classical regression. Nor are the \( x_j \)s truly independent in a strictly statistical sense. Thus, though the linear least estimator is unbiased, it may not have a minimum variance. In addition, the relationship between the linear least squares estimator and the kinetic rate constant is model-dependent. Discussions about the paradox of variance estimates of parameters in nonlinear models can be found in Hwang and Yeh (1977).

The estimated computer time required was investigated for neither LLS - GLM nor NLS - NLIN. In a multiprogramming/time-sharing environment, the actual run time in CPU seconds for a particular job is executed, and output via TSO (Time Sharing Option). In any event, the actual CPU seconds required for the SAS - GLM should be much less than those needed for the SAS - NLIN, since the convergence is always obtained in one iteration.

REFERENCES


Table 1. Numerical Values Used as Experimental Data for $C_i(l_1)$, $i = 1, 2, ..., 18$

<table>
<thead>
<tr>
<th>$i$</th>
<th>$L_i$ (cm)</th>
<th>$C_i$</th>
<th>$y_i$</th>
<th>$x_{2i}$</th>
<th>$x_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0.25</td>
<td>1.3956</td>
<td>-1.8054</td>
<td>0.7665</td>
<td>0.1682</td>
</tr>
<tr>
<td>3</td>
<td>0.50</td>
<td>1.9598</td>
<td>-2.6012</td>
<td>2.6868</td>
<td>0.3723</td>
</tr>
<tr>
<td>4</td>
<td>1.00</td>
<td>1.0791</td>
<td>-3.1209</td>
<td>1.9172</td>
<td>1.1829</td>
</tr>
<tr>
<td>5</td>
<td>1.50</td>
<td>0.9579</td>
<td>-3.2421</td>
<td>2.4025</td>
<td>2.2696</td>
</tr>
<tr>
<td>6</td>
<td>2.00</td>
<td>0.9134</td>
<td>-2.3566</td>
<td>2.5870</td>
<td>1.5275</td>
</tr>
<tr>
<td>7</td>
<td>2.50</td>
<td>0.8585</td>
<td>-3.3415</td>
<td>3.7116</td>
<td>6.9133</td>
</tr>
<tr>
<td>8</td>
<td>3.00</td>
<td>0.8104</td>
<td>-3.3969</td>
<td>4.6058</td>
<td>11.1739</td>
</tr>
<tr>
<td>9</td>
<td>3.50</td>
<td>0.7724</td>
<td>-3.4776</td>
<td>6.1308</td>
<td>21.0960</td>
</tr>
<tr>
<td>10</td>
<td>4.00</td>
<td>0.7379</td>
<td>-3.5561</td>
<td>7.5017</td>
<td>35.5060</td>
</tr>
<tr>
<td>11</td>
<td>4.50</td>
<td>0.6940</td>
<td>-3.6280</td>
<td>8.9182</td>
<td>79.3458</td>
</tr>
<tr>
<td>12</td>
<td>5.00</td>
<td>0.6530</td>
<td>-3.6884</td>
<td>9.8026</td>
<td>79.3458</td>
</tr>
<tr>
<td>13</td>
<td>5.50</td>
<td>0.6068</td>
<td>-3.7533</td>
<td>11.6507</td>
<td>113.3250</td>
</tr>
<tr>
<td>14</td>
<td>6.00</td>
<td>0.5592</td>
<td>-3.8770</td>
<td>13.0823</td>
<td>160.8523</td>
</tr>
<tr>
<td>15</td>
<td>6.50</td>
<td>0.5154</td>
<td>-3.9635</td>
<td>14.2375</td>
<td>217.6220</td>
</tr>
<tr>
<td>16</td>
<td>7.00</td>
<td>0.4748</td>
<td>-4.0183</td>
<td>15.3394</td>
<td>337.1700</td>
</tr>
<tr>
<td>17</td>
<td>7.50</td>
<td>0.4380</td>
<td>-4.0713</td>
<td>16.4514</td>
<td>403.2400</td>
</tr>
<tr>
<td>18</td>
<td>8.00</td>
<td>0.3965</td>
<td>-4.1315</td>
<td>17.5799</td>
<td>608.1990</td>
</tr>
</tbody>
</table>

Table 2. Outlier Patterns of $C_i(l_1)$ $i = 1, 2, ..., 18$

<table>
<thead>
<tr>
<th>$i$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>4.2</td>
<td>6.2</td>
<td>4.2</td>
</tr>
<tr>
<td>2</td>
<td>0.25</td>
<td>1.3956</td>
<td>1.2956</td>
<td>1.3956</td>
</tr>
<tr>
<td>3</td>
<td>0.50</td>
<td>1.9598</td>
<td>1.9598</td>
<td>1.9598</td>
</tr>
<tr>
<td>4</td>
<td>1.00</td>
<td>1.0791</td>
<td>1.0791</td>
<td>1.0791</td>
</tr>
<tr>
<td>5</td>
<td>1.50</td>
<td>0.9579</td>
<td>0.9579</td>
<td>0.9579</td>
</tr>
<tr>
<td>6</td>
<td>2.00</td>
<td>0.9134</td>
<td>0.9134</td>
<td>0.9134</td>
</tr>
<tr>
<td>7</td>
<td>2.50</td>
<td>0.8585</td>
<td>0.8585</td>
<td>0.8585</td>
</tr>
<tr>
<td>8</td>
<td>3.00</td>
<td>0.8104</td>
<td>0.8104</td>
<td>0.8104</td>
</tr>
<tr>
<td>9</td>
<td>3.50</td>
<td>0.7724</td>
<td>0.7724</td>
<td>0.7724</td>
</tr>
<tr>
<td>10</td>
<td>4.00</td>
<td>0.7379</td>
<td>0.7379</td>
<td>0.7379</td>
</tr>
<tr>
<td>11</td>
<td>4.50</td>
<td>0.6940</td>
<td>0.6940</td>
<td>0.6940</td>
</tr>
<tr>
<td>12</td>
<td>5.00</td>
<td>0.6530</td>
<td>0.6530</td>
<td>0.6530</td>
</tr>
<tr>
<td>13</td>
<td>5.50</td>
<td>0.6068</td>
<td>0.6068</td>
<td>0.6068</td>
</tr>
<tr>
<td>14</td>
<td>6.00</td>
<td>0.5592</td>
<td>0.5592</td>
<td>0.5592</td>
</tr>
<tr>
<td>15</td>
<td>6.50</td>
<td>0.5154</td>
<td>0.5154</td>
<td>0.5154</td>
</tr>
<tr>
<td>16</td>
<td>7.00</td>
<td>0.4748</td>
<td>0.4748</td>
<td>0.4748</td>
</tr>
<tr>
<td>17</td>
<td>7.50</td>
<td>0.4380</td>
<td>0.4380</td>
<td>0.4380</td>
</tr>
<tr>
<td>18</td>
<td>8.00</td>
<td>0.3965</td>
<td>0.3965</td>
<td>0.3965</td>
</tr>
</tbody>
</table>

Table 3. Numerical Results: Estimation of $b_1$, $b_2$, and $b_3$ - No Outlier

<table>
<thead>
<tr>
<th>ESTIMATION PROCEDURE</th>
<th>Initial</th>
<th>Final</th>
<th>NUMBER OF ITERATIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td>True Value</td>
<td>0.867</td>
<td>0.225</td>
<td>1.19</td>
</tr>
<tr>
<td>LSE - GLM</td>
<td>0.867</td>
<td>0.225</td>
<td>1.19</td>
</tr>
<tr>
<td>LSE - NLIM</td>
<td>0.867</td>
<td>0.225</td>
<td>1.19</td>
</tr>
<tr>
<td>LLS - GLM</td>
<td>0.35</td>
<td>0.18</td>
<td>0.368</td>
</tr>
<tr>
<td>LLS - NLIM</td>
<td>0.35</td>
<td>0.18</td>
<td>0.368</td>
</tr>
</tbody>
</table>

Table 4. Numerical Results: Estimation of $b_1$, $b_2$, and $b_3$ - With Outlier

<table>
<thead>
<tr>
<th>OUTLIER PATTERN</th>
<th>$b_1$</th>
<th>$b_2$</th>
<th>$b_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>LSE - GLM</td>
<td>0.9051</td>
<td>0.8172</td>
<td>0.225</td>
</tr>
<tr>
<td>LSE - NLIM</td>
<td>0.9051</td>
<td>0.8172</td>
<td>0.225</td>
</tr>
</tbody>
</table>

The initial estimates for $b_1$, $b_2$, and $b_3$ are 1.0, 1.0, and 0.0, respectively, for LLS - NLIM. The convergence takes about 12-15 iterations.