LETTER TO THE EDITOR

In "A Least Squares Test of General Use in the Radioisotope Laboratory," *J.N.M.*, p. 711, 1966, Dr. Donald W. Brown has presented some interesting and useful comments on data analysis. However, there are qualifications in the application of the least squares technique which he has not pointed out and which could be traps for the unwary. The statement that "Regression analysis provides a direct and completely objective method," is valid only if particular assumptions apply.

The danger in the computerization and automation of research laboratory procedures is that they come to be considered an end in themselves. These very useful and powerful tools tend to lull the user into thinking that they can correct for theoretical deviations or inaccurate data. Statistical manipulation of the data only serves to estimate the error or to obtain most probable values of the parameters.

The use of least squares, or any other statistical method of analysis for that matter, is justified only if it serves some purpose. For example:

A. The data is very poor and graphical correlations are uncertain. The numerical approach may then extract some slightly usable numbers.

B. It is known that the data contains large errors, but it is essential for the purposes of the experiment to estimate the most probable values and the statistical reliability of certain parameters. This is the most usual biological application. In Dr. Brown's letter, for example, it would be the variance of λ (rather than of log y, as indicated) that is of interest.

C. The most useful application of the technique is the case where both the equation of the curve and the statistics of the error are well known and there is a great deal of data available. It is then possible to determine the coefficients and their probable errors with high precision. This is rarely the case with biological data.

Dr. Brown has indicated that the numerical method of analysis of the data from plasma iron clearance and chromium-labelled red blood cell survival tests leads to a more correct answer than that obtained by graphically fitting a line to the data. This assumption is not necessarily valid. To apply the least squares method, there must be *a priori* knowledge of the true curve (3), in this case a straight line on semilogarithmic paper. The observed data must also be subject only to errors whose statistical properties are known. It is, indeed, essential that a graph be plotted first to indicate that these assumptions hold. If the data does not appear to lie on a straight line by graphical analysis, no measure of statistical manipulation, such as a least squares fitting technique, will make it fit a straight line. In many cases the data from the above mentioned tests is clearly a sum of two exponentials rather than a single exponential (1). This is precisely one of the pieces of information to be learned from these tests. One would suspect that the large errors shown in Fig. 1 of the letter (two cases with over 30% errors) are due to the mechanical attempt by calculation to force two exponentials into a single straight-jacket.

It may be well to look a little more generally into the equations cited by Dr. Brown. His equations are those for least squares fitting to a straight line, merely substituting the log of the counts for the counts. This is valid if certain assumptions are made: (1) There is only a single exponential function. (2) The errors in time measurement are negligible. (This is often not as true as we like to think.) (3) The total count and count-rate above background of each measurements are made: (1) There is only a single exponential function. (2) The errors involved are not systematic, i.e., they are not due to physiological variation of the patient, but to such things as sample volume measurement, where the error in each count is proportional to the count.

If these assumptions hold, the transformed equations cited by Dr. Brown are valid because the weighting factors are equal. In the more general case, a weighting factor W_i must be calculated for each observation. It is defined as (2, 3):

$$W_i = \frac{1}{\sigma_i^2}$$

where σ_i^2 is the variance of the transformed *ith* ordinate.

If conditions 1 through 4 hold, the weighting factor is the same for each observation and factors out. If condition 3 does not hold, i.e., the Poisson counting fluctuation is appreciable, then the weighting factor would be:

$$W_i = \frac{D_i}{\log e}$$

With other statistics, other weighting factors would have to be determined.

One final comment on measure of error. The standard log error cited by Dr. Brown is the standard deviation of log y about the line determined by the averaged λ and y_0. What is really desired is the variance of the slope and intercept. These can be calculated from the observed data and the appropriate weighting factors.

SAUL ARONOW, PH.D.
ALAN B. ASHARE, M.D.
PHYSICS RESEARCH LABORATORY
MASSACHUSETTS GENERAL HOSPITAL
BOSTON, MASSACHUSETTS 02114

REFERENCES

1. BECK, W. S., Hematology Research Laboratory, Massachusetts General Hospital. Personal communication.

2. DEMING, W. E., "Statistical Adjustment of Data," Wiley & Sons, New York, 1943.

3. WENTWORTH, W. E., "Rigorous Least Squares Adjustment," *J. Chem. Educ.*, **43**:96, 162, 1965.