DRUG DEVELOPMENT AND INDUSTRIAL PHARMACY, 7(1), 63-77 (1981)

STATISTICAL TREATMENT OF STABILITY DATA

Wu-huang Yang
Department of Industrial Pharmacy
Massachusetts College of Pharmacy
Boston, Massachusetts 02115

ABSTRACT

The establishment of the expiration date of pharmaceutical product through the statistical treatment of stability data is discussed in detail for both the stability studies under normal storage conditions and accelerated storage conditions.

EXPIRATION DATING AND STABILITY TESTING

It is stated in the Current Good Manufacturing Practice (CGMP) that the purpose of an expiration date for a pharmaceutical product is to assure that a drug product meets applicable standards of identity, strength, quality, and purity at the time of use. An expiration date is required for all drug products

except some specific exemptions as determined by the Food and Drug Administration (FDA). The expiration date should be established through a suitably designed stability testing program. General guidelines for such a stability testing program are instituted in CGMP. Although there is no formal FDA stability testing guidance established for the drug products for the human uses to date, more specific guideline are available (1-5). These reflect current interpretation and philosophy of FDA in the enforcement of the establishment of an expiration date.

It is emphasized in CGMP that an expiration date should be established through a statistically designed stability testing program. An expiration date projected from a stability testing program conducted under the accelerated conditions (eg. high temperature, high relative humidity etc.) is accepted currently only as a tentative expiration date, which has to be supported by a full-scale stability testing program conducted under normal storage conditions.

Much has been said about the theories involved in a stability testing. General statistical treatment methods for the stability data are also available (6-8). It is the objective of this article to give a detailed illustration of the establishment of an expiration date from the statistical treatment of the stability data. It is believed that using the equations and a calculator, the readers should be able to complete the calculation in a reason-

able period of time and through which have a better understanding of the basic idea involved. It is, of course, understood that various available statistical programs (such as SPSS) and other computer programs can be used to facilitate the calculation.

STABILITY TESTING UNDER NORMAL STORAGE CONDITIONS

Normal storage conditions are defined as the storage conditions specified on the label of the drug product. An expiration date established for a drug prodcut from a stability testing under normal storage conditions is the ultimate goal of all.

Kinetics of Loss of Active Ingredients

As a general rule, the degradation of an active ingredient in a drug product should be monitored to at least 70% completion in order to legitimately determine its kinetic order. For example, it is virtually impossible to differentiate zero, first, and second order if a degredation is followed to only 10% completion. For a possible drug candidate, and thus relatively stable, it would require a relatively long period of time to determine the kinetic order of the degradation reaction. Although it makes little difference, in such cases, to choose any kinetics order to treat the data and establish the expiration date as such, it would be more appropriate to identify the degradation pathway and deter-

mine its kinetic order so that the following statistical treatment is sound both in theory and in practice.

Least Square Linear Regression Method

Assume that the kinetic order of the drug degradation is established and the data can be linearized in an appropriate manner.

A stability testing program will be set up by determining the level of the intact drug in a drug product stored under normal conditions at different time points. This assaying of the intact drug should be conducted using a reliable, meaningful and specific stability-indicating analytical method as required by CGMP. The level of the intact drug is then transformed to appropriate form depending on the kinetic order of the degradation and is then plotted versus time. A best possible straight line is determined for the data points. This is usually done by employing the method of least square linear regression.

Let us assume that in a stability testing program, n chemical assays are carried for k time points ($n \geq k$, therefore there are n/k assays per time point). These n data points are then subjected to linear regression treatment using the following model:

$$Y_i = \beta_0 + \beta_1 t_i + \varepsilon_i \qquad \text{(Eq. 1)}$$

where Y_i is the transformed term for the level of intact drug at time t_i (Table 1), β_0 and β_1 are parameters to be estimated, ε_i is a random error term with mean $E(\varepsilon_i) = 0$ and variance $\sigma^2(\varepsilon_i) = \sigma^2$, and i=1, ...,n.

Table 1

Transformation of Kinetic Data

Degradation Order	Y	b_0	b_1
0	A*	A_0*	-k*
1	ln A	ln A_0	-k
2	1/A	$1/A_0$	k

*A = the level of the intact drug at time t
A_0 = the initial level of the intact drug
k = the rate constant

For each sample observation (t_i, Y_i), the method of least squares considers the deviation of Y_i from its expected value:

$$Y_i - (\beta_0 + \beta_1 t_i) = \varepsilon_i \qquad \text{(Eq. 2)}$$

In particular, the method of least square requires that we consider the sum of the n squared deviations, denoted by Q:

$$Q = \Sigma(Y_i - \beta_0 - \beta_1 t_i)^2 \qquad \text{(Eq. 3)}$$

where the summation is from i=1 to i=n.

The best estimates of parameter β_0 and β_1 are those values b_0 and b_1 respectively which minimize Q. This is achieved by setting the partial derivative of Q with respect to β_0 and β_1 to zero, (Eq. 4 and 5), and solve the equations simultaneously to obtain the values of b_0 and b_1 (Eq. 6 and 7):

$$\frac{\partial Q}{\partial \beta_0} = -2\Sigma(Y_i - \beta_0 - \beta_1 t_i) = 0 \qquad \text{(Eq. 4)}$$

$$\frac{\partial Q}{\partial \beta_1} = -2\Sigma t_i(Y_i - \beta_0 - \beta_1 t_i) = 0 \qquad \text{(Eq. 5)}$$

$$b_1 = \frac{\Sigma t_i Y_i - (\Sigma t_i)(\Sigma Y_i)/n}{\Sigma t_i^2 - (\Sigma t_i)^2/n} \qquad \text{(Eq. 6)}$$

$$b_0 = \bar{Y} - b_1 \bar{t} \qquad \text{(Eq. 7)}$$

where $\bar{Y} = \Sigma Y_i/n$ and $\bar{t} = \Sigma t_i/n$.

Therefore the regression function is:

$$Y = b_0 + b_1 t \qquad \text{(Eq. 8)}$$

For a specific time point t_i, Y_i is the observed value and $\hat{Y}_i = b_0 + b_1 t_i$ is the fitted value. The difference between Y_i and $\hat{Y}_i$ is called the residual, e_i:

$$e_i = Y_i - \hat{Y}_i \qquad \text{(Eq. 9)}$$

The sum of the squares of residuals, denoted as SSE(error sum of squares), is:

$$SSE = \Sigma e_i^2 = \Sigma(Y_i - \hat{Y}_i)^2 \qquad \text{(Eq. 10)}$$

SSE has n-2 degree of freedom associated with it, since two degrees of freedom are lost when both β_0 and β_1 are estimated in obtaining Y_i. MSE (error mean square) is defined as:

$$MSE = SSE/(n-2) = \Sigma(Y_i - \hat{Y}_i)^2/(n-2) \qquad \text{(Eq. 11)}$$

Calculation of Expiration Date

Expiration date or shelf life is generally defined as the time span required for the level of the active ingredient of a

drug product to reach 90% of the claimed strength. However, one cannot simply insert 90% of the strength into Eq. 8 and solve for t, and use it as the expiration date. The reason is that there are always some uncertainties involved in the estimation during the regression process. Rather, the lower confidence line for new observations at a specified confidence level should be computed and the time for that confidence line to reach 90% of the claimed strength is then considered as the expiration date. This process is illustrated in Figure 1.

Statistically, this is equivalent to the prediction of a new observation Y corresponding to a given level t of the independent variable. Although what we are seeking is exactly the opposite, i.e., to determine the value of t (expiration date) at a known level of Y (90% of the strength), the theories involved are exactly the same.

The new observation on Y is viewed as the result of a new trial, independent of the trials on which the regression analysis is based. This is the same as saying that the expiration date established for a specific sample will be applicable to those drug products manufactured under the same conditions. We shall denote the level of t for the new trial as t_h and the new observation on Y as Y_h. Of course, we assume that the underlying regression model applicable for the basic sample data, as shown in Eq. 8, continues to be appropriate for the new observation.

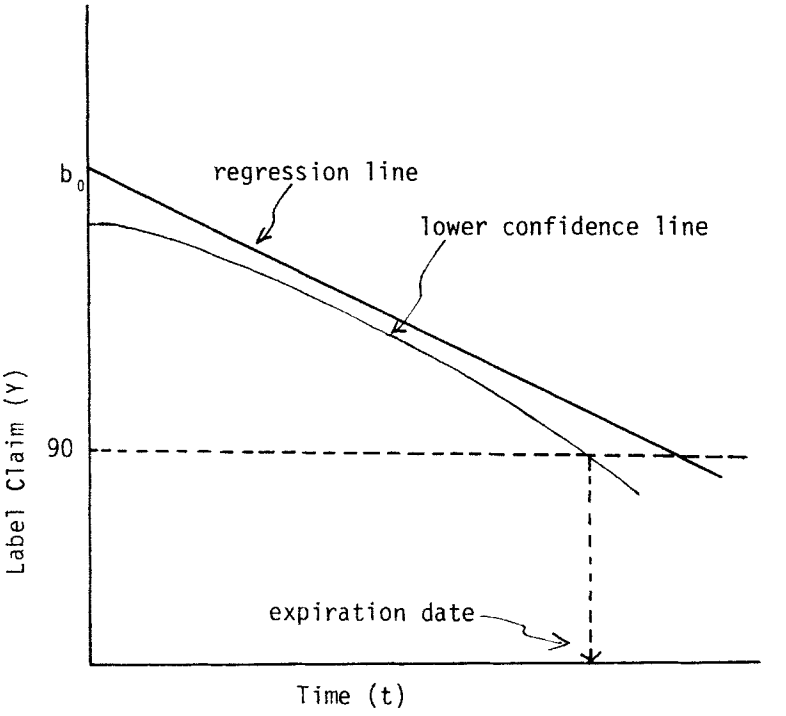

Figure 1

The establishment of expiration date. A drug undergoing zero order degradation is illustrated.

The point estimator for Y_h is:

$$\hat{Y}_h = b_0 + b_1 t_h \qquad \text{(Eq. 12)}$$

The $(1-\alpha)$ confidence interval of $\hat{Y}_h$ is:

$$\hat{Y}_h - t(1-\alpha/2;n-2)s(Y_h) \leq Y_h \leq \hat{Y}_h + t(1-\alpha/2;n-2)s(Y_h) \qquad \text{(Eq. 13)}$$

where $t(1-\alpha/2;n-2)$ is the t critical value for $1-\alpha$ confidence level with $n-2$ degree of freedom, and

$$s^2(Y_h) = (MSE)\left[\, 1 + \frac{1}{n} + \frac{(t_h - \bar{t})^2}{\Sigma(t_i - \bar{t})^2} \,\right] \qquad \text{(Eq. 14)}$$

The lower confidence level of Y_h in Eq. 13 is the only concern to us. If t_h is the expiration date, this lower confidence level is 90% of the strength. Therefore,

$$90 = \hat{Y}_h - t(1-\alpha/2;n-2)s(Y_h) \qquad \text{(Eq. 15)}$$

Substitution of Eq. 12 and 14 into Eq. 15 gives,

$$90 = b_0 + b_1 t_h - t(1-\alpha/2;n-2)\{(MSE)[1 + \frac{1}{n} + \frac{(t_h-\bar{t})^2}{\Sigma(t_i-\bar{t})^2}]\}^{1/2} \qquad \text{(Eq. 16)}$$

Eq. 16 is a quadratic equation for t_h and subsequently there are two possible solutions. One should always choose the smaller t_h value to have the smallest risk in setting up the expiration date.

Calculation of Level of Active Ingredient for a Projected Expiration Date

Suppose the expiration date established from the samples containing the active ingredient at the claimed level is not acceptable to a manufacturer, it is possible to use the available data to determine a new active ingredient level in a drug product for a new projected expiration date. Since most drug products listed in USP have 10% tolerance in the active ingredient level, this is indeed a feasible way.

Statistically, this is equivalent to obtaining a new b_0 value, b_0', for the new projected t_h value, t_h'. The slope of the

new regression line should remain as b_1 of the original regression line if the new samples are tested in the same manner. The value of b_0' is obtained by solving Eq. 16 after the known values of b_1 and t_h' are inserted. This process is illustrated in Figure 2.

Whether or not this new level for the active ingredient (i.e. b' value in Figure 2) is acceptable depends on how close this level is to 110 if a 10% tolerance is allowed. If this new level is not feasible to manufacture and the new projected expiration date is desired, then a reformulation becomes necessary. This practice should prove useful in the preformulation period.

STABILITY TESTING UNDER ACCELERATED STORAGE CONDITIONS

The projection of an expiration date of a pharmaceutical product from a stability testing under accelerated storage conditions is based on the applicablity of the Arrhenius theory.

The Arrhenius relation, which is basically a relationship between the rate constant and the temperature, is expressed as following,

$$k = A\exp(-E_a/RT) \qquad \text{(Eq. 17)}$$

where k is the rate constant of the degradative pathway, E_a the activation energy of the degradation, T the absolute temperature, R the gas constant, and A the frequency factor. Eq. 17 can be con-

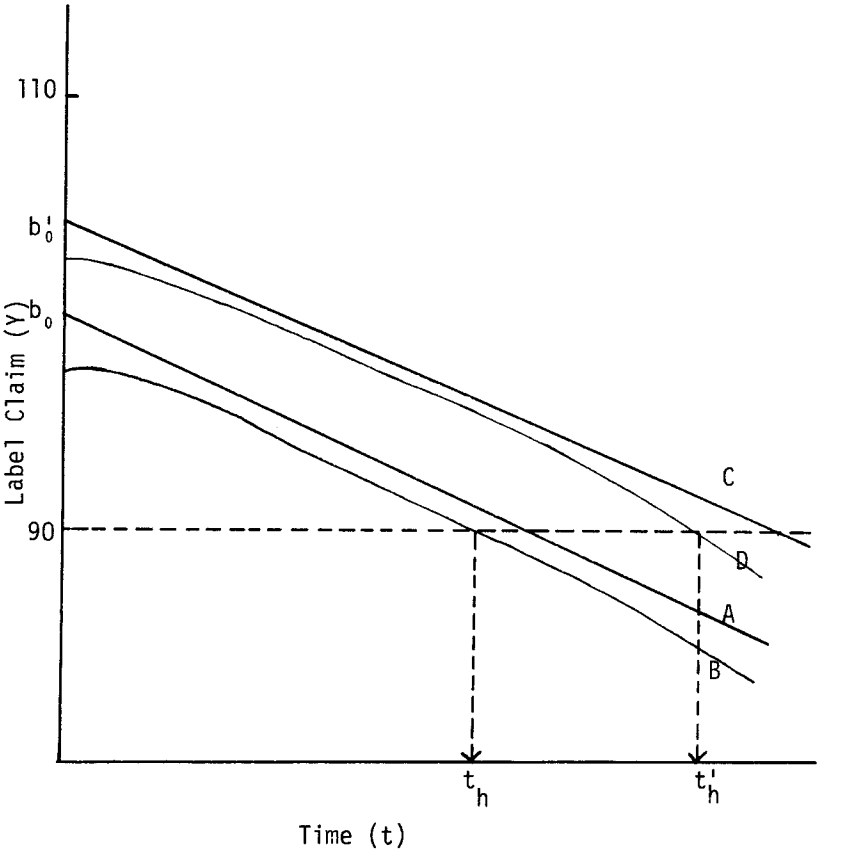

Figure 2

The calculation of the level of the active ingredient for a new projected expiration date. Line A is the regression line for stability data; line B is the lower confidence line for line A; line C is the projected regression line for the new product with the active ingredient level of b_0' with new projected expiration date t_h'; line D is the lower confidence line for line C.

verted to its linearized form by taking the natural logarithm of both sides,

$$\ln k = \ln A - \frac{E_a}{R}\frac{1}{T} \qquad \text{(Eq. 18)}$$

Briefly, in an accelerated stability testing the drug products are stored under several higher temperatures with all other conditions fixed and the level of the active ingredient is assayed at different time points. The data for each temperature are then subjected to the statistical treatment as elaborated previously. It is the rate constant, obtained from the slope of the regression line, which is our concern now.

Least Square Linear Regression Method

It is recognized that Eq. 18 is the equation of a straight line. The best line when ln k is plotted versus 1/T can be estimated again using the method of least square regression. All the equations (Eq. 1 to 16) are applicable to this treatment if it is recognized that ln k and 1/T of Eq. 18 are equivalent to Y and t of Eq. 1 respectively. After the regression line is obtained, it is then possible to extrapolate the data at higher temperatures to find the rate constant at a lower temperature or the rate constant at the normal storage conditions. The rate constant thus obtained can be substituted into the appropriate rate equation, as

listed in Table 1, to project a tentative expiration date for a given initial level of the active ingredient or vice versa.

Weighted Least Square Linear Regression Method

It has been pointed out (6,7) that a simple least square linear regression treatment violates one basic assumption inherent in the process, namely the constancy of the variance for the error term in Eq. 1. This violation can be explained by the following argument.

Eq. 19 gives the variance of the estimate of β_1 (from which the rate constant is obtained) in Eq. 1:

$$s^2_{b_1} = (MSE)/\Sigma(t_i - \bar{t})^2 \qquad \text{(Eq. 19)}$$

Assuming the Arrhenius relationship is applicable, a higher rate of degradation is expected to occur at higher temperature; hence assays at a higher temperature might be made at shorter time intervals for a shorter period of time. This has an effect on the value of $s^2_{b_1}$ since the denomenator in Eq. 19 is smaller at a higher temperature. Thus non-constant variances for the slopes of the regression lines are built in the derivation of the accelerated stability data.

One method to overcome this violation is to use weighted least square regression method. It was shown (7) that a better statistical model for such treatment using Arrhenius relationship is:

$$\ln k_i = \beta_0 + \beta_1(1/T) + \varepsilon_i/k_i \qquad \text{(Eq. 20)}$$

where the new error term is ε_i/k_i and the variance ε_i^2/k_i^2. In weighted least square regression method, the estimators of β_0 and β_1 are obtained by minimizing the quantity (9):

$$Q = \Sigma w_i[\ln k_i - \beta_0 - \beta_1(1/T)]^2 \qquad \text{(Eq. 21)}$$

where the w_i's are weights, which should be the inverse of the variance. Thus the appropriate weights in this case are:

$$w_i = k_i^2 \qquad \text{(Eq. 22)}$$

The general normal equations which arise from minimizing Eq. 21 with respect to β_0 and β_1 seperately are:

$$\Sigma w_i(\ln k_i) = b_0\Sigma w_i + b_1\Sigma w_i(1/T_i) \qquad \text{(Eq. 23)}$$

$$\Sigma w_i(\ln k_i)(1/T_i) = b_0\Sigma w_i(1/T_i) + b_1\Sigma w_i(1/T_i)^2 \qquad \text{(Eq. 24)}$$

The estimators b_0 and b_1 can then be obtained by solving Eq. 23 and 24 simultaneously. The rest of the treatment in obtaining a tentative expiration date is the same as described in previous section.

REFERENCES

1. Joel S. Davis, Pharmaceutical Technology, 65, Oct 1979
2. Quality Control Reports, The Gold Sheets, 12(11)(1978)
3. ibid., 13(5)(1979)
4. ibid., 13(11)(1979)
5. FDA Bureau of Veterinary Medicine, "Drug Stability Guidelines", 2nd revision(1979)

6. Edward R. Garrett, J. Amer. Pharm. Asso., Sci, Ed., 45, 171 (1956)

7. Donald L. Bentley, J. Pharm. Sci., 59,464(1970)

8. J.T. Carstensen and E. Nelson, J. Pharm. Sci.,65, 311(1976)

9. John Neter and William Wasserman, "Applied Linear Statistical Models," Irwin, Illinois, 1974