J. Pharm. Pharmacol. 1980, 32: 155–159 Received June 21, 1979

Accelerated storage tests on pharmaceutical products: effect of error structure of assay and errors in recorded temperature

O. L. DAVIES* AND D. A. BUDGETT†

University College of Wales, Aberystwyth, † ICI Ltd., Pharmaceuticals Division, Alderley Park, Macclesfield, Cheshire, U.K.

The likely error structure for accelerated shelf-life studies is discussed. How to fit the Arrhenius relationship in the presence of this structure is shown and it is concluded that test potencies between 20% and 50% are likely to provide most information. The effect of temperature errors on the estimation of shelf-life is also considered.

Accelerated storage tests are used to predict the shelf-life of pharmaceutical products. The product is stored, usually in sealed ampoules, at two or more elevated temperatures. These are sampled at suitable time intervals and their potency determined. Clark & Hudson (1968) and Kirkwood (1977) described the conduct of the test and the statistical analysis of the results. The statistical analysis involves postulating a model to represent the deterioration rates at the various temperatures, fitting this model to the data usually by the method of least squares, verifying the model and using it to extrapolate to ambient temperatures to estimate the shelf-life.

It appears that insufficient attention has been paid to the error structure of the assay methods when fitting the model, and also, there is very little precise information on the possible effects of errors in the measurement of the temperatures of the storage ovens on the estimation of the shelf-life. These two questions are now examined statistically.

Model

For most drugs tested, the deterioration is first order, that is

$$\log y = \log y_0 - K(T)t \quad .. \quad .. \quad .. \quad (1)$$

where y is the potency at time t, y_0 the initial potency and K(T) the rate of deterioration at temperature T. The rate K(T) is related to the temperature T, measured in degrees absolute, by the Arrhenius Law:

$$K(T) = a\, e^{-E/RT} \quad .. \quad .. \quad .. \quad .. \quad (2)$$

where 'a' is the frequency factor, E the energy of activation and R the gas constant.

A more useful form for K(T) to use statistically is

$$K(T) = e^{A-B/T} \quad .. \quad .. \quad .. \quad .. \quad (3)$$

where A and B are parameters.

An ampoule stored at an elevated temperature is usually assayed side-by-side with an ampoule stored at a low temperature where no appreciable deterioration occurs. The latter is regarded as a standard and the potency of the sample is expressed as a per cent or a proportion of the standard. If y is expressed as a proportion the model becomes:

$$Y = \log y = -e^{A-B/T}t \quad .. \quad .. \quad .. \quad (4)$$

Capital Y is used throughout this paper to denote the log of the potency expressed as a proportion of the standard.

Error structure of the assay

Laboratory experience suggests that the standard error of a potency determination when expressed as a percent of the potency, or equivalently the standard error of log y, increases with decreasing potency. There do not appear to be any published data to quantify this, but it is shown in this section that such a result can be expected. It is necessary to take this into account when estimating the parameters of the model. It is reasonable to postulate three sources of error:

(i) those which give rise to a standard error that is proportional to the potency, e.g. dilution errors. Here the standard error is denoted by σ_0.

(ii) those which give rise to a standard error that is proportional to the amount of deterioration, e.g. temperature variations within the storage areas. Here the standard error is denoted σ_1.

* Correspondence.

(iii) those which give rise to a standard error that is independent of the potency, e.g. errors arising in measuring a peak height in a chromatography trace. Here the standard error is denoted by σ_2.

These sources are independent and may be combined giving:

$$\text{Variance of potency} = y^2\sigma_0^2 + (y_0-y)^2\sigma_1^2 + y_0^2\sigma_2^2 \quad (5)$$

$$\text{Variance of log potency} = \sigma_0^2 + (y_0/y-1)^2\sigma_1^2 + (y_0/y)^2\sigma_2^2 \quad (6)$$

This equation is very flexible and covers all situations between a constant variance for y (when $\sigma_0 = \sigma_1 = 0$) to a constant variance for log y (when $\sigma_1 = \sigma_2 = 0$).

The formula can be simplified to:

$$\sigma_0^2 + (y_0/y)^2\sigma_2^2 \quad (7)$$

by assuming that σ_1 is negligible or absorbed in σ_0 and σ_2.

An alternative basis for the derivation of the error structure of the assay is the following:

Both the sample and the standard ampoules are analysed side-by-side in each assay, and the measurement on the latter, denoted by y, is expressed as a proportion (or percentage) of the measurement on the standard, denoted by y_0. We assume that both the sample being assayed and the standard ampoules are given identical treatment, e.g. there is no differential dilution. If now we denote the standard deviations from the three sources of variation when applied to the measurement on each of standard and sample by $\sigma_A, \sigma_B, \sigma_C$ respectively, it follows that:

$$\text{Variance of } y_0 = (\sigma_A^2 + \sigma_C^2)y_0^2 \quad (8)$$

$$\text{Variance of } y = \sigma_A^2y^2 + \sigma_B^2(y_0-y)^2 + \sigma_C^2y_0^2$$

Therefore: Variance of log potency = Variance of log y_0 + Variance of log y =

$$(2\sigma_A^2 + \sigma_C^2) + \sigma_B^2(y_0/y-1)^2 + \sigma_C^2(y_0/y)^2 \quad (9)$$

This is of the same general form as (6) except that the σ's have a different interpretation and $\sigma_0^2 \geqslant \sigma_2^2$ in (6).

σ_B is likely to be small and can be considered as being absorbed in σ_A and σ_C

Substituting R for $\sigma_C^2/(2\sigma_A^2 + \sigma_C^2)$, (9) then becomes:

$$(2\sigma^2{}_A + \sigma_C^2)(1 + R(y_0/y)^2) \quad (10)$$

R must be less than or equal to 1: the extreme situation R = 0 occurs when $\sigma_C = 0$. For the other extreme when $\sigma_A = 0$, the formulae reduces to $\sigma_C^2(1 + (y_0/y)^2)$. If σ_B^2 is appreciable and considered absorbed in σ_A^2 and σ_C^2, or if the standard is replicated, R could then exceed unity.

Formula (10) is structurally more meaningful than (7), but from the point of view of representing the expected increase in the standard deviation of log potency for decreasing potency the formula are equivalent. Table 1 shows this relationship numerically for formula (7) where R is equal to σ_2^2/σ_0^2. The standard error is expressed as a percent of the potency, and the standard error of the maximum potency is made equal to 1%.

Table 1. Relative standard error as percent of potency.

% Potency	R = 2	1	0·5	0·25	R→∞
100	1·0	1·0	1·0	1·0	1·0
80	1·2	1·1	1·1	1·1	1·3
50	1·7	1·6	1·4	1·3	2·0
30	2·8	2·5	2·1	1·7	3·3
20	4·1	3·6	2·9	2·4	5·0
10	8·2	7·1	6·1	4·6	10·0

Relative standard error = (Relative standard errors at 100%) $\times \sqrt{\left[\frac{1 + R(100/y)^2}{1 + R}\right]}$

The general behaviour of the various situations depicted in Table 1 is similar in that down to 50% potency there is a moderate increase in the standard error of log potency, but for lower potencies the increase is rapid. The rates of increase are larger for larger R. There is an upper limit to this rate of increase when $\sigma_0 = 0$ in equation (7), that is, when $R \to \infty$.

It appears that the increases shown in Table 1 are of the magnitudes which are expected to occur in practice. The value of 1% for 100% potency may be somewhat optimistic and a value of 3% can occur, although 2% may be more likely. The figures of Table 1 should be multiplied by whatever standard error is expected for 100% potency. In the above we assumed no differential dilution which is the usual situation. However, in some cases the standard is diluted to give roughly the same concentration as the sample. If, in addition, both sample and standard are diluted so that roughly the same response is obtained throughout the range of potencies then the standard error will be proportional to the potency. The standard error of log potency will be constant except for any error relating to the amount of deterioration i.e. σ_1. This is because dilution errors will be a consistent proportion of the dilution whilst measurement errors and measurement will both be fairly constant at all potencies. Some test samples may have less potency before dilution than the others

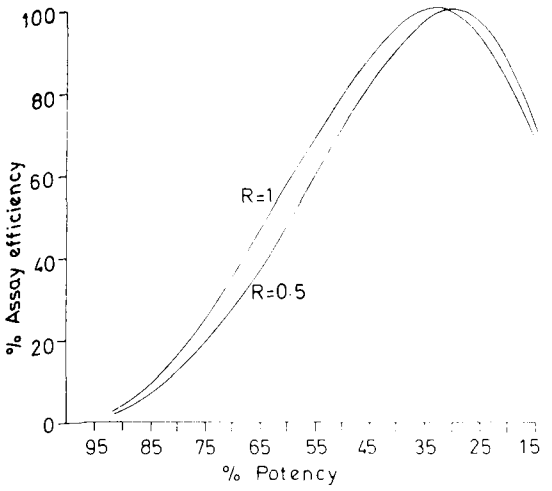

FIG. 1. Relationship between potency and efficiency.

after. In this case the σ_2 error will be constant in the *potency scale* so the standard error of log potency will increase with decreasing potency.

Fitting the model

The model is

$$\log y = -\exp(A-B/T).t \quad .. \quad .. \quad .. \quad (11)$$

where the potency y is expressed as a proportion. The potency determinations are log normal, and it follows that the appropriate method of fitting is by least squares. The model is non-linear and the procedure for minimizing the sum of squares has to be an iterative one. Any of the recognized minimization procedures may be used, provided:

(i) the model is reparameterized as follows:

$$\log y = -\exp\left(A' - B\left(\frac{1}{T} - \lambda\right)\right)t \quad .. \quad (12)$$

where λ is a weighted mean of 1/T. This is the usual reparameterization used for the Arrhenius model as described previously by Box (1960).

(ii) each observation is weighted by the inverse of the variance of the log potency as given by the error structure formula for the assay.

(iii) good approximate values of the parameters are used as starting values.

The last of these is important and a simple method exists for the purpose. The method is to fit

$$\log K(T) = A' - B\left(\frac{1}{T} - \lambda\right) .. \quad .. \quad .. \quad (13)$$

where K(T) is the deterioration rate for temperature T.

This requires an estimate of K(T) for each temperature. Denote a given temperature by the suffix j, then $K(T_j)$ is obtained by fitting

$$Y = \log y = -K(T_j)\, t. \quad .. \quad .. \quad .. \quad (14)$$

to the observations of the jth temperature by weighted least squares. This gives:

$$\hat{K}(T_j) = \Sigma w_{ji} Y_{ji} t_{ji} / \Sigma w_{ji} t_{ji}^2 \quad .. \quad .. \quad .. \quad (15)$$

where the suffix ji denotes the ith observation for temperature j and w_{ji} its weight.

In order to fit (13) we need the weight to be attached to each log $K(T_j)$. This is derived as follows:

Variance of $\hat{K}(T_j) = \sigma^2/\Sigma w_{ji} t_{ji}^2$

where σ^2 is the combined residual variance over all the temperatures. These are standard statistical procedures. The variance of log $\hat{K}(T_j)$ is obtained by dividing the variance of $\hat{K}(T_j)$ by $(\hat{K}(T_j))^2$. Thus

$$\text{variance of } \log \hat{K}(T_j) = \sigma^2/\Sigma w_{ji} t_{ji}^2 (\hat{K}(T_j))^2 \quad (16)$$

This may be expressed in another form by using the relationship

$$Y_{ji} = -\hat{K}(T_j) t_{ji}$$

Substituting for $K(T_j)$ in (16) gives

$$\text{variance of } \log \hat{K}(T_j) = \sigma^2/\Sigma i w_{ji} \hat{Y}_{ji}^2 \quad .. \quad (17)$$

The weight to be attached to log $K(T_j)$ is the inverse of its variance, that is

$$W_j = \sum_i w_{ji} \hat{Y}_{ji}^2/\sigma^2 \quad .. \quad .. \quad .. \quad .. \quad (18)$$

Since only relative weights are needed, we may drop the divisor σ^2. This formula is of considerable interest because it gives the relative information supplied by each assay. Thus an assay with a potency of y, expressed as a proportion, supplies relative information given by

$W = wY^2$.

where $Y = \log y$.

The relative values of W for the error structures given in Table 1 are tabulated in Table 2 for a wide range of potencies. The maximum value for each structure has been equated to 100: the entries then give the percent efficiency of each potency. The graphs for the various structures are similar: the efficiencies rise from low values for high potencies to a maximum around 30% potency. The maximum is fairly flat giving a wide range from 20% to 50% potency for near optimum values. This has an important bearing on the design of accelerated storage tests. As far as practicable one should aim for potencies within this near optimum range. It is clear that potencies above 60% should be avoided if possible. There may be additional reasons for avoiding low potencies.

Further considerations in fitting the model

If λ is taken to be $\Sigma(W_i/T_i)/\Sigma W_i$ in (13), the equation reduces to

$$\log K(T) = \text{mean log } K(T) - B(1/T - \lambda) \quad .. \quad (19)$$

where mean log $K(T) = \Sigma W_i \log K(T_i)/\Sigma W_i$ and $A' =$ mean log $K(T)$.

The values of A′ and B derived in this way give good starting values for the application of a non-linear estimation procedure. We have found that a simple Gauss procedure, commonly referred to as Gauss-Newton procedure, when applied to the data of many tests, converges in 3 to 5 iterations, even when the model was not a good fit. We have also found that the contours of the likelihood surface over a wide range were near ellipses and the ridge was approximately linear.

Table 2. % Efficiency of a potency determination for various assay error structures derived from formula (7).

% Potency	R = σ_2^2/σ_0^2 2	1	0·5	0·25	$\sigma_0 = 0$
90	4·7	4·1	3·1	2·2	6·7
80	18·8	15·7	12·6	9·3	23·7
70	39·1	34·7	28·4	21·9	45·9
65	50·0	45·5	38·3	30·3	57·7
60	62·5	57·0	49·2	40·0	69·6
55	73·4	68·6	60·7	50·8	79·9
50	82·8	78·3	72·2	62·4	88·8
45	92·2	88·4	82·8	74·1	95·5
40	96·9	95·9	90·4	85·1	99·2
35	100·0	99·2	97·7	94·1	99·9
30	96·9	99·2	99·6	99·7	96·2
25	90·6	93·4	96·2	99·6	88·8
20	79·7	82·6	86·4	92·8	77·0
15	62·5	65·3	69·8	77·2	60·0

The entries for R = 1 and 0·5 for formula (7) are plotted in Fig. 1.

This means that for the cases examined the parameters had only a small degree of non-linearity and their distributions could be assumed approximately normal. Log K(T), being a linear function of the parameters A′ and B, can then also be assumed to be approximately normal. This would follow anyway if the standard error of K(T) is small, say less than 20% of K(T) for all K(T)'s. In practice, this restriction can be relaxed for the smaller K(T). These considerations lead to the conclusion that the starting values derived by the above procedure give satisfactory estimates of A′ and B in most situations without the need to follow with a non-linear minimization procedure. This conclusion is used in the next section.

Effect of errors in the measurement of temperature

The measurement of the temperature of the ovens is subject to error, and it is believed that this could easily amount to a standard deviation of 0·1 °C. It is necessary to find out what effect this has on the estimation of the shelf-life. Defining the shelf-life as the time taken to lose 10% of potency at ambient temperature the log shelf life is given by

$$\log \text{shelf life} = \log(-\log 0{\cdot}9) - A' + B(1/T_0 - \lambda) \quad (20)$$

where T_0 is the ambient temperature in degrees absolute. A′ and B are approximately normal and therefore the estimate of log shelf life is also approximately normal. This was confirmed by Booth (1979) from extensive simulations.

Equation (13) can be fitted to any set of suitable data, and A′ and B can be expressed as joint functions of the temperatures. The standard error of log shelf-life can thus be derived from the differentials of log shelf life with respect to each temperature as follows:

$$DT_i = \frac{\partial(\log \text{shelf life})}{\partial Ti} = -\frac{\partial A'}{\partial T_i} + \frac{\partial B}{\partial T_i}\left(\frac{1}{To} - \lambda\right)$$

Standard error of log shelf life = $\sqrt{\Sigma(DT_i)^2}$ multiplied by the standard error of temperature.

The values of T_i are substituted after differentiation. When the procedure is programmed for a computer, the differentiation is carried out step-wise at intermediate stages.

The above general procedure was used in the following empirical method to find the standard deviations of the log shelf life due to temperature errors in a variety of situations, and compared with the relative standard deviation due to assay errors.

A standard deviation of 0·1 °C was assumed for the temperature. For the assay error, formula (7) was used taking $\sigma_0 = 0{\cdot}0082$ and $\sigma_2 = 0{\cdot}0057$, that is, R = 0·5 and standard error of 100% potency is 1%. Starting with a value for the log frequency factor A and a value for B—the activation energy/gas constant, for a known compound, four consecutive temperatures are chosen in the set 50, 60, 70, 80, 90, 100 °C best suited for the compound. The design is one assay for each temperature taken at the times for which the expected potency is 30%. A time limit of 50 days is imposed which means that for some temperatures the potency of 30% is not reached, in which case the expected potency at 50 days is taken. The above general procedure can then be applied to the expected potencies at the specified sampling times. This was carried out for a number of known drugs and the results are given in Table 3. The ambient temperature for the estimation of shelf life was taken to be 23 °C.

Table 3. Effect of temperature errors and assay errors on the accuracy of the estimate of shelf life at 23 °C.

No.	A log frequency factor	B = E/R	Shelf life (days)	Standard error of log shelf life due to: temp. errors σ = 0·1 °C		assay errors* σ_0 = 0·008165, σ_2 = 0·005774	
1	33·95	13297	6142	0·0386	(3·9)	0·0760	(7·9)
2	21·33	8455	146	0·0229	(2·3)	0·0585	(6·0)
3	24·09	9059	71	0·0180	(1·8)	0·0386	(3·9)
4	21·57	9703	7788	0·0437	(4·5)	0·2545	(29·0)
5	33·67	13286	7823	0·0413	(4·2)	0·0827	(8·6)
6	19·18	8475	1342	0·0283	(2·9)	0·1126	(11·9)
7	15·31	6771	203	0·0184	(1·9)	0·0670	(6·9)
8	27·40	9864	39	0·0196	(2·0)	0·0379	(3·9)
9	34·16	11978	58	0·0238	(2·4)	0·0379	(3·9)
10	33·87	13286	6413	0·0390	(4·0)	0·0772	(8·0)
11	19·11	8455	1345	0·0283	(2·9)	0·1134	(12·0)
12	25·93	9764	122	0·0194	(2·0)	0·0419	(4·3)
13	35·89	13035	364	0·0261	(2·6)	0·0443	(4·5)
14	33·80	13186	4895	0·0368	(3·7)	0·0717	(7·4)
15	14·94	6694	227	0·0184	(1·8)	0·0708	(7·3)
16	20·01	7851	71	0·0185	(1·9)	0·0479	(4·9)

* R = 0·5, s.d. of Y_0 0·01 which is equivalent to a standard error of 1% for y_0.
(The figures in brackets are the equivalent standard deviations expressed as per cent of shelf-life.)

The contribution of temperature errors to the shelf life is around 2–4%. These cannot be considered serious particularly when the standard errors of 3–4% apply to very long shelf lives for which the deterioration rates at the elevated temperatures are not sufficient to give expected potencies of 30% within the time limit of the test.

Booth (1979) applying somewhat different methods on a different but related model, reached similar conclusions on the effect of temperature errors.

Replication will reduce the standard error due to sampling and assay errors but will not affect the standard error due to errors in temperature. There is therefore an economic limit to the replication. For example, four replications would halve the effect of assay error. In most cases the temperature error would then predominate and little would be gained by further replication.

It is evident from the above data that assay and temperature errors can make an important contribution to the error associated with a shelf life prediction from an accelerated stability study. After consideration of the Arrhenius equation, this is not unexpected but we have endeavoured to show that good experimental design can significantly minimize the potential errors involved and also possibly economize in time. For example, once the order of reaction has been established, preliminary experimentation would determine the time for almost 55% decomposition to occur at the temperatures chosen for the study. If samples were monitored between 55 and 75% decomposition instead of between 10 and 75% decomposition, which is usual, the number of samples necessary to assay may be significantly reduced due to the increase in efficiency of the assay at this level of potency. It is appreciated that all drugs would not fit this design since it is not always practical to obtain 40% decomposition over a reasonable time period particularly at one or more of the lower temperatures employed. However, its value at the higher temperatures usually employed cannot be ignored.

It has been shown that with low potencies (see Table 1) the assay error increases and it may be advisable not to exceed 75% decomposition. It is unlikely that errors other than those discussed would interfere at this level of potency. At still lower potencies it is feasible that other chemical reactions may become more significant and produce products which the assay method may not be capable of differentiating from the parent drug.

Table 3 shows that temperature errors of the order ±0·1 °C do not seriously affect shelf life predictions. Larger errors e.g. ±4 °C usually result in non-linearity of the Arrhenius regression. If this type of error is suspected it becomes increasingly difficult to decide if the system under study obeys the Arrhenius law. Inaccuracy in monitoring temperature is only going to be a small contribution to the overall temperature error. The major contribution will come from the sensitivity of the oven thermostat, the fall in temperature experienced when the oven door is opened to remove samples and the time taken to return to the temperature employed, and non-uniform temperature distribution through samples. The use of a water bath to store the samples may eliminate some of the problems but may also introduce others. It may be concluded however that the design of accelerated storage tests can be improved by a statistical consideration of some of the variables involved.

Acknowledgements

Acknowledge is due to Dr H. E. Hudson for extensive discussion of the pharmaceutical aspects of this problem, and to Mr S. H. Ellis for suggestions in relation to the error structure of the assay.

REFERENCES

Booth, A. (1979) Unpublished Thesis of the University of Wales.

Box, G. E. P. (1960) Fitting Experimental Data, M.R.C. Technical Summary Report No 151

Clark, C. J., Hudson, H. E. (1968) Manuf. Chem. Aerosol News 39: 1: 25–27

Kirkwood, T. B. L. (1977) Biometrics 33: 736–742