

International Journal of Pharmaceutics 184 (1999) 165–172

Incorporating batch effects in the estimation of drug stability parameters using an Arrhenius model

Issa T. Some a.*, Philippe Bogaerts ^b, Raymond Hanus ^b, Michel Hanocq ^a, Jacques Dubois^a

^a *Laboratoire de Chimie Bioanalytique*, *de Toxicologie et de Chimie Physique applique´e*, *Institut de Pharmacie*, *Uni*6*ersite´ Libre de Bruxelles*, *Campus de la Plaine*, *CP* ²⁰⁵/1, *Boule*6*ard du triomphe*, *B*-¹⁰⁵⁰ *Brussels*, *Belgium* ^b *Ser*6*ice d*'*Automatique*, *Faculte´ des Sciences applique´es*, *Uni*6*ersite´ Libre de Bruxelles*, *CP* ¹⁶⁵⁵⁵, *A*6*enue F*.*D*. *Roose*6*elt* ⁵⁰, *B*-1050 *Brussels*, *Belgium*

Received 10 June 1998; received in revised form 7 January 1999; accepted 8 January 1999

Abstract

The nonlinear estimation of drug stability parameters (energy of activation E_a and shelf-life t_y) by conventional approaches employs equations relating drug content determination *C* at time *t* and temperature *T*. The identification procedures lead to the determination of only one initial drug content C_0 for several different experiments. However, it is well known that because of experimental concentration variation or of intentional modification of the experimental schedule, there are as many initial drug contents as experiments. For these reasons, a method which takes into account batch effects is proposed to determine stability parameters and also all initial drug contents C_{0j} where *j* is the index of experiment in one step. This method is more accurate from a statistical viewpoint and is suitable for data treatment in pharmaceutical industries where the initial drug content of each batch entering the stability program can be checked a posteriori. The application of this method is shown on real kinetic data from the hydrolysis of acetylsalicylic acid (ASA). © 1999 Published by Elsevier Science B.V. All rights reserved.

Keywords: Batch effects; Drug stability; Arrhenius model

1. Introduction

Stability studies are designed to give an insight into the drug degradation mechanism and expiry dating (shelf-life) estimation. Shelf-life is defined as the time required for a drug to decompose to

* Corresponding author. Tel.: $+32-2-6505178$; fax: $+32-2-$ 6505187.

E-*mail address*: jdubois@nlb.ac.be (J. Dubois)

90% of its initial concentration at a specific temperature (T) , relative humidity and light energy exposure if required.

The establishment of the prospective expiry date is of prime importance to drug product stability. Accelerated stability testing using the Arrhenius relationship is often employed for stability parameter identification.

The well-known classical approach consists of sequential steps (Garrett and Carper, 1955; Garrett, 1956, 1957). It includes the determination of the kinetic constant k_{pH} *T* for the correct order of the degradation reaction. This is done through the functional relationship between drug content *C* and time *t* at several temperatures *T*.

For a first-order reaction, this relation is

$$
\ln C = \ln C_0 - k_{\text{(pH,T)}} t \tag{1}
$$

Based on the measured times *t* and the corresponding concentrations *C*, it is easy to compute the least-squares estimates of ln *C* and $k_{(pH,T)}$. This identification is repeated for several experiments at different temperatures *T*. Then with the pairs (*T*, $k_{(pH,T)}$), the frequency factor *A* and the energy of activation E_a can also be identified with a leastsquares method using the classical Arrhenius equation

$$
\ln k_{\text{(pH,}T)} = \ln \mathbf{A} - (E_a/RT) \tag{2}
$$

Finally, the rate constant $k_{(pH,T_v)}$ at room temperature T_X (298.16 K) is deduced from Eq. (2) and the predicted shelf-life t_y is determined from the relationship

$$
t_{Y} = \frac{1}{k_{\text{(pH,}T_{X})}} \ln\left(\frac{100}{Y}\right)
$$
 (3)

This conventional method of data treatment suffers from certain statistical problems as it does not take into account all relevant experimental information in parameter estimation (errors associated with the drug contents and the constants $k_{(pH,T_v)}$ determinations). Therefore other techniques based on the Arrhenius equation have been used. Weighted least-squares analyses were suggested by Bentley (1970), Yang (1981), Yang (1983) and Nash, (1987). Non-linear approaches have been proposed by Davies and Hudson (1982) and King et al. (1984). The method introduced by King et al. takes into account all the drug contents *C* and time *t* at different temperatures *T* in the determination of stability parameters that is the energy of activation E_a , the shelf-life t_y and the initial drug content C_0 . The mathematical development for a first-order reaction is as follows:

A first-order degradation reaction is described by the equation

$$
C = C_0 \exp(-kt) \tag{4}
$$

and the exponential form of the Arrhenius equation is given by

$$
k = A \, \exp\left(\frac{-E_a}{RT}\right) \tag{5}
$$

For a fixed temperature $T = 298.16$ K, Eq. (5) can be written

$$
k_{\text{(pH,298.16)}} = A \exp\left(\frac{-E_a}{R298.16}\right) \tag{6}
$$

Rearrangement of Eq. (6) for *A* gives equation

$$
A = k_{\text{(pH,298.16)}} \exp\left(\frac{E_a}{R298.16}\right) \tag{7}
$$

Substitution of Eq. (7) for *A* into Eq. (5) yields:

$$
k = k_{\text{(pH,298.16)}} \exp\left(\frac{E_{\text{a}}}{R298.16}\right) \exp\left(\frac{-E_{\text{a}}}{RT}\right) \tag{8}
$$

Substituting Eq. (8) for *k* into Eq. (4) and given that

$$
t_{90} = \frac{\ln(100/90)}{k_{\text{(pH,298.16)}}}
$$
(9)

it can be then written that

$$
C = C_0 \exp\{-t \ln(100/90)/t_{90} \exp[(E_a/R)(1/298.16 - 1/T)]\}
$$
\n(10)

Eq. (10) is of prime importance as it allows a direct determination of stability parameters from experimental data. The conventional identification method (King et al., 1984) allows the estimation of only one initial drug content C_0 , but the real statistical significance of this parameter is not often well established.

From a statistical and experimental viewpoint, it is inadequate to estimate one initial concentration in this case. It is well known that in accelerated stability studies, different initial drug contents are used to gather kinetic constants $k_{(pH,T)}$ at different temperatures and therefore there are as many initial drug contents as experiments. For a first-order kinetic, the shelf-life t_y is not dependent on concentrations and as a consequence, the same final values of E_a and t_y could be expected whatever the initial concentrations. This seems not to be the case with the conventional identification procedure.

2. Materials and methods

².1. *Materials*

Analytical grades of acetylsalicylic acid, salicylic acid and 2,4 dihydroxybenzoïc acid (internal standard) were obtained from Sigma-Aldrich (Belgium). HPLC grade of methanol and ethanol were supplied by Merck (Belgium), and distilled. Analytical grade of mono- and disodium phosphate, and potassium chloride were obtained from Merck (Belgium). High purified water from Milli-Q filters system (Millipore, USA) was used for the preparation of all solutions.

².2. *HPLC method*

The HPLC system consisted of a solvent delivery pump (Gilson, France), an injection system with a 20-µl loop (Rheodyne, USA), a 300 mm \times 4.6 mm column (μ Bondapak) packed with 10 μ m irregular particles from Waters (USA) with a temperature control device operating at 25°C, a UV detector (Gilson Holochrome, USA) and a computing system equipped with a data acquisition software program (Borwin Chromatographic software, J.M.B.S., France). The mobile phase consisted of methanol 40% (v/v) and 0.05 M phosphate buffer (pH 2.5) 60% (v/v). The mobile phase was filtered through a 0.45-um pore nylon membrane (Millipore, USA) and deaerated under reduced pressure. The flow rate was maintained at 1.5 ml/min. The detection wavelength was set at 230 nm (the isobestic point of salicylic acid).

².3. *Kinetic method*

A stock solution containing 2 mg per ml of acetylsalicylic acid in pure ethanol was prepared;

5-ml aliquots were transferred into a 50-ml volumetric flask and brought to volume with 0.18 M potassium phosphate buffer (pH 5.98) previously maintained at the desired temperature. The flasks were then shaken and the solutions placed into sealed vials and stored in an oven (Bekso, Belgium). Samples of 1 ml were withdrawn at appropriate time intervals and immediately diluted with 18 ml of 0.18 M phosphate buffer (pH $2.5 \approx$ maximum stability) and 1 ml of internal standard from a 26 μ M stock solution to quench the reaction. The samples were immediately analyzed or frozen and kept at -20 °C until analysis was performed. The time period for each temperature was selected to achieve at least 70% of drug degraded. The initial drug concentrations were 5.55×10^{-5} and 2.775×10^{-5} M. The buffer solution was adjusted to 0.3 M ionic strength with potassium chloride. The among of potassium chloride added was calculated according to van Damme et al. (1979).

².4. *Data analysis*

The reaction was studied at four temperatures (308.16, 318.16, 328.16 and 333.16 K). Two data sets were generated. The first contained kinetic raw data where all the initial drug contents were similar (5.55 \times 10⁻⁵ M) whereas in the second set, some of the initial concentrations were equal to 2.75×10^{-5} M and others to 5.55×10^{-5} M (Tables 4 and 5).

The computer programs used were written in a Matlab environment (version 4.2, MathWork, Natick, MA). The first computer program was based on conventional approach and the second program followed our proposed method. The simplex algorithm (Nelder and Mead, 1965) was used as the nonlinear optimization method for parameter determination.

3. Improved method for stability parameter identification

In their paper, King et al., used a specific extrapolation temperature T_X (298.16 K) and 90% of drug content remaining (Eq. (10)). This equation can be put into general terms for a first-order model

$$
C = C_0 \exp{\{-t \ln(100/Y)/t_Y \exp[(E_a/R)(1/T_X - 1/T)]\}}
$$
\n(11)

where the user can fix the desired extrapolation temperature T_X and the percentage (*Y*) of remaining drug.

For the resolution of this equation, the measurement errors on time *t* values and temperatures *T* are supposed to be negligible. On the other hand, the errors in drug content measurements *C* cannot be considered negligible. We make the a priori assumptions that the measurements have a stationary white noise (i.e. with uncorrelated samples) whose Gaussian distribution has zero mean and unknown variance σ^2 .

These assumptions lead to a least-squares costfunction

$$
J(\theta) = \sum_{j=1}^{M} \sum_{i=1}^{N_i} (C_{ij}(t_{ij}, T_j) - f(C_{ij}, t_{ij}, T_j; E_a, t_Y, C_{0j}))^2
$$
\n(12)

where $\hat{\theta} = [t_{Y}, E_{a}, C_{01}, \dots C_{0M}]$ is the unknown parameter vector, *M* is the number of experiments, N_i is the number of measurement samples in experiment *j*, C_{ii} is the measured concentration and $f(C_{ij}, t_{ij}, T_j; E_a, t_Y, C_{0j})$ is the model given by Eq. (11).

The simplex algorithm allows us to solve the nonlinear optimization problem which consists of minimizing the least-squares cost function (Eq. (12)) with respect to the vector parameter θ .

Based on the least-squares cost function *J* and

the identified parameters
$$
\hat{\theta}
$$
 where $\hat{\theta} = \begin{bmatrix} \hat{t}_Y \\ \hat{E}_a \\ \hat{C}_{01} \\ \hat{C}_{02} \\ \vdots \\ \hat{C}_{0M} \end{bmatrix}$,

the variance–covariance matrix can be approximated by

$$
\hat{E}[\tilde{\theta}\tilde{\theta}^T] \cong \hat{\sigma}^2 P(\hat{\theta})
$$
\n(13)

where $\hat{E}[\tilde{\theta}\tilde{\theta}^T]$ is the estimate of the covariants of

the parametric error $\tilde{\theta} = \theta - \hat{\theta}$,

$$
\hat{\sigma}^2 = \frac{J(\hat{\theta})}{\sum_{j=1}^{M} N_j - \dim(\theta)}\tag{14}
$$

is the estimate of the measurement noise variance which is nothing but the minimized least-squares cost function divided by the difference between the number of measurements and the number of parameters, and

$$
P^{-1}(\hat{\theta}) = \sum_{j=1}^{M} \sum_{i=1}^{N} (J_{ij})^{T} (J_{ij})
$$
\n(15)

is the sensitivity matrix of the model with respect to the parameters where

$$
J_{ij} = \frac{\mathrm{d}f(C_{ij}, t_{ij}, T_j; E_{\rm a}, t_Y, C_{0j})}{\mathrm{d}\theta}\Big|_{\theta = \hat{\theta}}
$$
(16)

These latter relations (Eqs. (13) – (16)) are just given in order to provide details about the whole method. Some commercial software include the computation of this variance–covariance matrix. The reader interested in more theoretical details could refer to Seber and Wild (1989).

4. Results and discussion

Table 1

In the framework of this first-order kinetic model validation by the conventional non-linear identification method, the statistical significance of the parameter C_0 appears unclear. Tables 1 and 2 give the results of the fitting of ASA hydrolysis data where all the initial concentrations are identical (first data set) and with different initial contents (second data set) and the two sets lead to

Stability parameters of ASA hydrolysis with King's optimization method (first data set)

Parameter	(θ_0)	Initial estimates Final estimates S.E. $\hat{\theta}$	
t_{90} (h)	10	5.67	0.07
E_a (cal/mol)	16 000	14 629	89
C_0^*	5.55	5.86	0.01
$J(\theta)$	37.41	0.3	

* Same initial concentrations (\approx 5.55 × 10⁻⁵ M).

Table 2

Stability parameters of ASA hydrolysis with the conventional optimization method (second data set)

Parameter	Initial esti- mates (θ_0)	Final estimates S.E. $\hat{\theta}$			
t_{90} (h)	10	6.33	0.08		
	16 000	15 261	88		
$E_{\rm a}$ (cal/mol) C_0^*	5.55	2.932	0.007		
$J(\theta)$	27.85	0.07			

* Different initial concentrations ($\approx 5.55 \times 10^{-5}$ and \approx 2.75×10^{-5} M).

different stability parameters. The energy of activation and the estimated shelf-life are of the samemagnitude but statistically different. The difference between the estimates of initial drug content is more significant. The discrepancy between the identified values for shelf-life and energy of activation is not acceptable.

The analysis of the correlation coefficients from the variance–covariance matrix (Table 3) corresponding to the results of Table 1 indicates a high negative correlation between t_{90} and C_0 ($\rho = -$ 0.67). This high correlation cannot be explained by the Arrhenius relation. The correlation is less significant between E_a and C_0 ($\rho = 0.24$).

These results indicate a fundamental problem in the parameter identification procedure and therefore, a new method is proposed that allows the determination of stability parameters including all initial drug contents. This method is more accurate from statistical and experimental points of view.

Table 4 displays the parameters identified with the first data set and Table 5 the parameters obtained with the second data set by minimizing the cost function (Eq. (12)). The stability parameters are similar whatever data set is used.

Table 3

Correlation coefficient drawn from the variance–covariance matrix of results of the first data set with the conventional identification method

Parameter	l_V	$E_{\rm a}$	\sim 0
l_Y $E_{\rm a}$	0.93	0.93	-0.67 0.24
$\mathsf{c}_\mathfrak{o}$	-0.67	0.24	

Table 4

Stability parameters identified with the proposed method. Similar initial drug contents

Parameter	(θ_0)	Initial estimates Final estimates S.E. $(\hat{\theta})$				
t_{90} (h)	10	5.4	0.3			
E_a (cal/mol)	16 000	14 297	358			
C_{01}^*	5.57	5.8	0.1			
C_{02}	5.57	5.8	0.1			
C_{03}	5.57	5.7	0.1			
C_{04}	5.55	5.9	0.1			
C_{05}	5.55	5.9	0.1			
C_{06}	5.55	5.9	0.1			
C_{07}	5.57	5.8	0.1			
C_{08}	5.58	5.8	0.1			
C_{09}	5.57	5.7	0.1			
C_{010}	5.60	5.6	0.2			
C_{011}	5.60	5.7	0.2			
C_{012}	5.60	5.6	0.2			
$J(\theta) \times 10^{-10}$	37.41	2.474				

* Concentration units (10−⁵ M).

All initial drug contents are determined.

The comparison of these results with those of Table 1 shows any significant differences. This clearly indicates that, in the conventional identification method, the estimated initial drug content is meaningful only if all the initial drug concentra-

Table 5

Stability parameters obtained by the proposed method. Different initial drug contents

Parameter	Initial esti- mates (θ_0)	Final estimates S.E. $\widehat{(\theta)}$			
t_{90} (h)	10	5.4	0.2		
E_{\circ} (cal/mol)	16 000	14 25 1	366		
C_{01}^*	2.79	2.91	0.09		
C_{02}	2.79	2.90	0.09		
C_{03}	2.79	2.86	0.09		
C_{04}	2.78	2.93	0.09		
C_{05}	5.55	5.9	0.1		
C_{06}	5.55	5.9	0.1		
C_{07}	2.79	2.88	0.09		
C_{08}	5.58	5.81	0.09		
C_{09}	5.57	5.66	0.09		
C_{010}	2.80	2.8	0.1		
C_{011}	5.60	5.7	0.1		
C_{012}	5.60	5.5	0.1		
$J(\theta) \times 10^{-10}$	27.85	1.402			

* Concentration units (10−⁵ M).

Table 6														
	Correlation coefficients between identified parameters drawn from the variance-covariance matrix with the proposed method (results from Table 4)													
Parameters	t_{90}	$E_{\rm a}$	C_{01}	C_{02}	C_{03}	C_{04}	C_{05}	C_{06}	C_{07}	C_{08}	C_{09}	C_{010}	C_{011}	C_{012}
	1.000	0.929	0.175	0.174	0.172	0.049	0.049	0.049	-0.147	-0.148	-0.144	-0.315	-0.323	-0.313
t_{90} (h) E_a (cal/mol)	0.929	1.000	0.291	0.291	0.287	0.132	0.133	0.133	-0.072	-0.072	-0.070	-0.263	-0.270	-0.261
	0.175	0.291	1.000	0.150	0.148	0.089	0.089	0.092	0.036	0.036	0.034	-0.028	-0.028	-0.028
	0.174	0.291	0.150	1.000	0.148	0.090	0.089	0.089	0.036	0.036	0.034	-0.028	-0.028	-0.028
C_{02}	0.172	0.287	0.148	0.148	1.000	0.087	0.090	0.090	0.034	0.034	0.034	-0.028	-0.028	-0.026
C_{03} C_{04}	0.049	0.132	0.089	0.090	0.087	1.000	0.057	0.059	0.035	0.035	0.032	0.004	0.004	0.004
C_{05}	0.049	0.133	0.089	0.089	0.090	0.057	1.000	0.059	0.035	0.035	0.035	0.004	0.004	0.004
	0.049	0.133	0.092	0.089			0.059	1.000			0.035	0.004	0.004	0.004
C_{06}		-0.072	0.036	0.036	0.090 0.034	0.059 0.035	0.035	0.035	0.035 1.000	0.035 0.052	0.052	0.061	0.063	0.061
	-0.147 -0.148	-0.072	0.036	0.036	0.034	0.035	0.035	0.035	0.052	1.000	0.052	0.061	0.063	0.061
		-0.070	0.034	0.034	0.034	0.032	0.035	0.035	0.052	0.052	1.000	0.059	0.061	0.059
			-0.028	-0.028	-0.028	0.004	0.004	0.004	0.061	0.061	0.059	1.000	0.107	0.105
	-0.144					0.004	0.004			0.063				0.107
C_{07} C_{08} C_{09} C_{010}	-0.315	-0.263						0.004	0.063		0.061	0.107	1.000	
C_{011} C_{012}	-0.323 -0.313	-0.270 -0.261	-0.028 -0.028	-0.028 -0.028	-0.028 -0.026	0.004	0.004	0.004	0.061	0.061	0.059	0.105	0.107	1.000

^a Errors are the differences between the experimental and calculated values of drug contents. The values given are the experimental ones.

tions used in the stability study are strictly equal. This is impossible to achieve even for a very skillful analyst.

The conventional identification method leads to biases in the \hat{t}_Y and \hat{E}_a values due to the experimental variations C_{0j} which are all described with one degree of freedom \hat{C}_0 .

The results in Tables 2 and 5 are obtained with the same set of data. The significant differences in the results exhibit in a clear way the necessity for the more rigorous identification approach, which is proposed here.

The variance–covariance matrix corresponding to the results in Table 4 is displayed in Table 6. The analysis shows a strong correlation between E_a and t_y ($\rho = 0.93$). This strong correlation is the consequence of the Arrhenius relation. On the other hand, there is no significant correlation between the shelf-life and the initial drug content $(\rho=0.05)$ or between the energy of activation and initial drug content ($\rho=0.2$). The initial drug contents are also independent ($\rho=0.01$). This is more in accordance with the chemical meanings of these parameters.

Because of the non-linearity of the model, different final estimates can be obtained depending on initial estimates. Therefore tests of the influence of initial values on the final estimates are the simplest way to confirm the validity of new models or the reliability of algorithms. In stability studies, good initial values can be obtained from the two-stage Garrett approach, which offers the simplest way to obtain them. The robustness of the proposed identification method with respect to the initial estimates of E_a and t_y has been tested and shown that the same values were identified with a wide range of initial estimates (± 100 times final values) and the simplex method seems very robust for the fitting process.

The correctness of the estimates is checked by cross-validation. The cross-validation in a triplicate experiment consists of determining the model parameters with the two first experiments and using these results as the initial values of the third experiment. The final results must not be significantly different from the first one. The cross-validation test confirmed the reliability of the estimated stability parameters. The results were not significantly different from those presented in Table 4.

Table 7 shows the first data set and the errors associated with the determination of concentrations as a function of time and temperature. The a posteriori error distribution shows a Gaussian error distribution and the calculated values of drug content are in good agreement with the experimental ones.

5. Conclusion

The non-linear equation proposed by King et al. is adequate and usually used for the direct estimation of drug stability parameters. The identification procedure used is not in accordance with the experimental approach, as it does not take into account the experimental variations in initial drug content

The method proposed in this paper avoids this problem and allows the determination of stability parameters as well as all initial drug contents in one step whatever initial concentrations are used during the experiments. This method is in good accordance with the experimental program and is more accurate from a statistical viewpoint. It is then more suitable for data treatment in pharmaceutical industries where each batch entering the stability program can be monitored for its initial drug content.

Acknowledgements

We acknowledge the 'Service de la Coopération de l'Universite´ Libre de Bruxelles' for grant support and Didier Bien for his skillful computer program management.

References

- Bentley, D.L., 1970. Statistical techniques in predicting thermal stability. J. Pharm. Sci. 59, 464–468.
- Davies, O.L., Hudson, H.E., 1982. Stability of drugs: accelerated storage tests. In: Statistics in Pharmaceutical Industry, 2nd edn. Marcel Dekker, New York, pp. 445– 480.
- Garrett, E.R., 1956. Prediction of stability in pharmaceutical preparation. II. J. Am. Pharm. Sci. 45, 171.
- Garrett, E.R., 1957. Prediction of stability in pharmaceutical preparation. III. J. Am. Pharm. Sci. 46, 584.
- Garrett, E.R., Carper, R.F., 1955. Prediction of stability in pharmaceutical preparation. I. J. Am. Pharm. Sci. 44, 515.
- King, S-Y.P., Kung, M.S., Fung, H.L., 1984. Statistical prediction of drug stability based on non-linear parameter estimation. J. Pharm. Sci. 73, 657–662.
- Nash, R.A., 1987. A new linear model for stability prediction. Drug Dev. Ind. Pharm. 13, 487–499.
- Nelder, J.A., Mead, R., 1965. A simplex method for function optimization. Comput. J. 7, 308–313.
- Seber, G.A.F., Wild, C.J., 1989. Nonlinear Regression. Wiley, New York.
- van Damme, M., Hanocq, M., Molle, L., 1979. Méthodes de calcul de la force ionique de solutions tampons. J. Pharm. Belge. 34, 5–12.
- Yang, W.H., 1981. Statistical treatment of stability data. Drug Dev. Ind. Pharm. 7, 63–67.
- Yang, W.H., 1983. A discusssion on room temperature for stability testing. Drug Dev. Ind. Pharm. 9, 849–859.