ELSEVIER

International Journal of Pharmaceutics 204 (2000) 61–68

international journal of pharmaceutics

www.elsevier.com/locate/ijpharm

Statistical assessment of between batch stability equivalence

Matías Llabrés*, Alexis Oliva, José Fariña

Departmento de Ingeniería Química y Tecnología Farmacéutica, Facultad de Farmacia, Universidad de La Laguna, 38200 La Laguna, Tenerife, Spain

Received 6 October 1998; received in revised form 17 May 2000; accepted 31 May 2000

Abstract

A statistical method for testing the equivalence between batches regarding their stability is proposed. This method is based on the statistical linear model making use of a set of dummy variables to code the different batches. The method gives us the point estimates of the slope and zero intercept of one batch, and the differences and the corresponding confidence intervals with the remaining batches. In a second step, zero intercepts and slopes are estimated for all the batches. Stability equivalence assessment is based on the comparison of the confidence intervals for the differences between batches with the maximum difference allowable. The main advantages of this method are the possibility to compare several batches, to disclose the equivalence stability criteria from the statistical hypothesis about the equality between slopes, and the joint estimated of the residual variance whatever the decision to pool or not the data from different batches. This method is illustrated with two data set; the first one, previously published by other authors, involved six batches; the second data set include two batches and arose in a stability study of a commercial human insulin conducted in our laboratory.

Keywords: Statistical method; Testing; Batches; Stability

1. Introduction

The Tripartite Guideline on stability testing of new drug substances and products (US Pharmacopoeia, 1995) states that shelf-life must be calculated as the time corresponding to the interception of the lower 95% confidence interval of the regression line with the lowest acceptable limit for drug content, usually 90% of the labeled amount; moreover, stability testing must be conducted using at least three different batches, the shelf-life of the drug product being that of the less stable batch. If no difference in slope or zero intercept is statistically identified, data from the different batches can be pooled; in this way, the degree of freedom of the error variance is increased and the length of the confidence interval will be reduced, leading us to a more precise estimation of the shelf-life.

The method suggested by the Tripartite Guideline is the testing hypothesis of no difference in slope and zero time intercepts using the analysis of covariance (ANCOVA). As long as the main concern is to fail to detect an important difference

* Corresponding author. Tel.: +34-922-318450; fax: +34-922-630095.

E-mail address: mllabres@ull.es (M. Llabrés).

0378-5173/00/$ - see front matter
PII: S0378-5173(00)00469-5

between batches, i.e. type II error, it has been suggested to conduct the hypothesis test with a large significative level, say 0.25 (Bancroft, 1964). However, this procedure has two drawbacks. First, these statistical criteria are independent of what can be considered as an important difference in slope. Second, the chance to pool data from different batches will be larger for poorly designed studies (small number of experimental points, large error variance) than for well designed studies. This approach can be used not only to assess the equality of slopes but also to analyse more complex studies (Fairweather et al., 1995).

Alternative methods for testing the equality of slopes and intercepts have been proposed previously. Norwood (1995) proposed a two-level nested ANCOVA model; in this model, differences between batches are interpreted both through the fixed effect 'initial potency' and by the random interaction term 'batch by time'. Hypothesis test about zero intercept and about the interaction term allows us to test the convenience of pooling the data, but not a specific rule regarding maximum difference allowable in intercepts or slopes. The method proposed by Chow and Shao (1991) also consider the batches as a random effect and, as the same authors stated, their method is only suitable for post-marketing studies in which a large number of batches with relatively few points per batch are compared. Ruberg and Stegeman (1991) specifically addressed the problem of fixing the maximum difference allowable; these authors proposed fixing the power of the test, $1-\beta$, being β the type II error, and to compute the significance level needed to keep the power of the test. To do so, the alternative hypothesis must be established.

In this paper an alternative statistical method is proposed; this method is based on the statistical linear model and it includes a set of dummy variables to code the different batches, leading us to estimate the slope and zero intercept of one batch as well as the differences with the remaining batches and the corresponding confidence intervals. In a second step, intercepts and slopes for each batch can also be computed. This method is illustrated with two data set: the first one corresponding to data previously used by Ruberg and Stegeman (1991); the second data set arises from a room temperature stability test of one commercial preparation of insulin conducted in our laboratory; it is well known that insulin must be stored in the refrigerator, so, the analysis presented here is for illustration purpose.

2. Methodology

2.1. Statistical model

The proposed statistical model is an extension of the linear model which includes dummy variables to take into account the fixed effect of each tested batch; such models have been described in detail (Draper and Smith, 1981), and will be shortly described in the following paragraphs.

Let's assume that k batches have been included in the trial, and that the number of observations of each bath j is n_j, $j=1\ldots k$; the total number of observations is therefore $N=\Sigma n_j$. The introduction of the dummy variables into the usual linear model lead us to split the model according to the equation

$$\mathbf{y}=\mathbf{X}\,\beta+\sum_{j=1}^{k}\mathbf{z}_j\,\mathbf{X}\,\alpha \qquad (1)$$

In this equation $\mathbf{y}$ is the vector of experimental observations y_l of length N ($l=1\ldots N$), being l equal to $\Sigma n_{j-1}+i_j$ for the ith observation from batch $\#j$. The N by 2 matrix $\mathbf{X}$ contains in its second column the observation times (the first column contains only ones to account for the intercept). The meaning of the vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ will be explained in the following paragraphs. Batch $\#j$ is coded through the vector $\mathbf{z}_j$; these vector are extracted from the matrix of dimensions $k\times(k-1)$

$$\begin{pmatrix} 0 & 0 & .... & 0 \\ 1 & 0 & .... & 0 \\ .. & .. & .... & .. \\ 0 & 0 & .... & 1 \end{pmatrix}$$

Note that this is an identity matrix $\mathbf{I}_{(k-1,\,k-1)}$ expanded with a first row of zeros. Thus, each observation from batch $\#1$ is identified by the

vector $\mathbf{z}_1' = [0, 0, \ldots 0]$, from batch # 2 by $\mathbf{z}_2' = [1, 0, \ldots 0]$, and so on. Therefore, for batch # 1 the model is reduced to

$$\mathbf{y}_1 = \mathbf{X}\,\beta \tag{2}$$

where $\mathbf{y_1}$ is the vector containing the observations from batch # 1, and therefore the vector $\boldsymbol{\beta}' = [\beta_0, \beta_1]$ corresponds to the intercept and slope of this first batch. For the remaining batches, the resulting equation is

$$\mathbf{y}_j = (\beta_0 + \alpha_{j-1}) + (\beta_1 + \alpha_{k+j-1})\mathbf{x}_j, \quad j > 2 \tag{3}$$

where $\mathbf{x}_j$ is the vector corresponding to sampling times for batch # j. Therefore, the first half of the elements of vector $\boldsymbol{\alpha}$ are the differences of intercepts between batch 1 and batches 2, 3,…k, and the second half of elements of vector $\boldsymbol{\alpha}$ the differences in the slopes.

As long as the model in Eq. (1) corresponds to the linear one, and assuming the independence, normality and homogeneity of experimental errors, least squares estimation of the vector $[\boldsymbol{\beta}|\boldsymbol{\alpha}]'$ is done as usually. Let the matrix $\mathbf{C}$ equal to the design matrix (Appendix A). The estimates of the model parameters, $[\mathbf{b}|\mathbf{a}]'$, is computed through the equation,

$$\begin{pmatrix}\mathbf{b}\\ \mathbf{a}\end{pmatrix} = (\mathbf{C}'\,\mathbf{C})^{-1}\,\mathbf{C}'\,\mathbf{y} \tag{4}$$

The variance–covariance matrix of the estimates, $\mathbf{V}$, is done by

$$\mathbf{V} = \mathrm{cov}\begin{pmatrix}\mathbf{a}\\ \mathbf{b}\end{pmatrix} = s^2(\mathbf{C}'\,\mathbf{C})^{-1} \tag{5}$$

being s^2 the mean square of the residuals, the estimate of the variance error, and computed from

$$s^2 = \frac{\sum_{j=1}^{k}\sum_{i=1}^{n_j}(y_{ij} - \hat{y}_{ij})^2}{N - 2k} \tag{6}$$

where $\hat{y}_{ij}$ is the predicted value for the observation # i for batch # j. Intercepts and slopes for each batch can also be computed as can be deduced from Eq. (3), and their variances from the expression $\mathbf{z}_j'\,\mathbf{V}\,\mathbf{z}_j$, being $\mathbf{V}$ the covariance matrix of model estimates (Eq. (5)) and $\mathbf{z}_j$ the vector for coding batch # j defined previously. Confidence intervals from model parameters are therefore derived from Eq. (5); confidence interval for individual batches can also be obtained using the equation

$$\hat{y}_{ij} \pm t_{(1-\alpha,\ N-2k)}\, s[\mathbf{c}_0'(\mathbf{C}'\,\mathbf{C})^{-1}\,\mathbf{c}_0]^{1/2}$$

where $\mathbf{c}_0$ is the vector $[1, x_l, \mathbf{z}_j']'$, x_l the sampling time corresponding to the observation y_{ij}, and $t_{(1-\alpha,\ N-2K)}$ is the Student's statistics with $N - 2k$ degree of freedom and significative level α. Computations were carried out using Mathematica software (Wolfran, 1988).

2.1.1. Pooling criteria

According to Ruberg and Stegeman (1991), maximum allowable difference in slopes between any two batches, $\Delta\beta$, will be equal to the ratio $\Delta C/\theta$, being ΔC the maximum allowable difference in drug potency at the end of the shelf-life, and θ the desired shelf-life. Therefore, data from two different batches will be pooled if two conditions are met: first, the $1-\alpha$ confidence interval ($1-\alpha$ CI) for the difference in slopes is contained in the interval $[-\Delta\beta,\ +\Delta\beta]$, and second, the $1-\alpha$ CI contains the zero value. First condition assures that the power of the test is enough to detect a difference between batches considered relevant while the second one avoids the bias of pooling batches with different slopes.

3. Experimental

3.1. Analytical method

Quantification of insulin was carried out using a slight modification of the HPLC method proposed by Farid et al. (1989) (Water system consisting of a pump, model 600E Multisolvent Delivery System, a UV-VIS detector, a model 490E Programmable Multiwavelenght Detector and data acquisition software, Maxima 820). A reversed-phase C-18 column (Delta Pack, 300 Å, 8·100 mm, Waters) and 74% 0.2 M sodium sulfate buffer (pH 2.3)/26% acetonitrile as the eluent at a flow rate of 1.0 ml/min and UV detection at 214

nm. Deionized water (MilliQ, Millipore Waters) was used throughout; all other chemicals and reagents were HPLC grade. All solvents were filtered with 0.45 μm (pore size) filters (Millipore). The percentages are given as w/v if not otherwise stated.

The validation of the analytical method was done using seven standard solutions of human insulin (Batch: H01003, Novo Biolabs) with concentrations ranging from 2 to 8 μg/ml in 0.05 M HCl. Each concentration was replicated four times. The minimum quantifiable concentration of human insulin was 2.0 μg/ml. The analysis of variance of linear regression (ANOVA) of the calibration line confirms the linearity of the peak area against the concentration; the coefficient of variation the predicted concentration was 2.6% (Hunter and Lamboy, 1981).

Samples from insulin solution preparations (1 ml) were withdrawn from vials and samples were stored at 4°C until analysis. Insulin samples in solution were directly diluted with 0.05 M HCl, as necessary to obtain concentrations within the calibration range. To assess the stability of the analytical methods along the stability study, a standard solution was injected each working day to construct a control chart; no point was outside twice the standard deviation. Analytical quality control chart has been published elsewhere (Oliva et al., 1996).

3.2. Stability study

Two different batches of a commercial human insulin preparation were used. The vials were stored protected from light at room temperature along 2 years. The monthly mean temperature was 20.7, ranging from 20.0 to 24.2°C. Samples were taken at regular time intervals; a total of 50 samples were taken.

4. Results and discussion

In order to discuss the principal features of the proposed method, we have re-analyzed the data set 1 published by Ruberg and Stegeman (1991) using the method described above. Estimated parameters (Eq. (4)) along with their univariate 1–0.05 confidence interval (95% CI) are shown in Table 1. The results of the analysis of the variance of regression showed a residual variance s^2 equal to 0.1192 with 25 degrees of freedom (df); the F-test for the origin of variation Regression$|b_1$ was equal to 54.8 with 11 and 25 df, being the null hypothesis rejected for $\alpha < 0.01$. From this test we can conclude that there is at least another model parameter statistically different from zero, excluding b_1. This previous test must be conducted because confidence intervals for individual esti-

Table 1
Model parameter estimates for data published by Ruberg and Stegeman (1991) ($N = 37$; $s^2 = 0.119$ with 25 df)[a]

Model parameter	Estimate	Standard deviation	95% confidence interval	
$b1$	100.5	0.176	100.1	100.8
$b2$	−1.515	0.0903	−1.701	−1.329
$a1$	0.167	0.254	−0.356	0.690
$a2$	−0.357	0.270	−0.914	0.200
$a3$	−0.038	0.309	−0.675	0.599
$a4$	−0.0360	0.332	−0.720	0.648
$a5$	−0.510	0.332	−1.19	0.174
$a6$	0.0663	0.153	−0.249	0.381
$a7$	0.300	0.177	−0.0646	0.665
$a8$	0.123	0.308	−0.511	0.756
$a9$	−0.484	0.452	−1.415	0.448
$a10$	−0.186	0.452	−1.117	0.745

[a] Coefficients $b1$ and $b2$ correspond to the intercept and slope for batch $\#1$. Coefficients $a2$ to $a5$ correspond to the differences between the intercepts of batches $\#2$ to $\#6$ and the intercept for batch $\#1$. Coefficients $a6$ to $a10$ correspond the slope differences.

Table 2
Intercepts and slopes estimated from data depicted in Table 1 for each of the six tested batches[a]

Batch	Intercept	Slope
#1	100.5 (0.176)	−1.515 (0.093)
#2	100.6 (0.182)	−1.449 (0.123)
#3	100.1 (0.205)	−1.215 (0.152)
#4	100.4 (0.254)	−1.392 (0.294)
#5	100.4 (0.281)	−1.999 (0.443)
#6	99.98 (0.281)	−1.701 (0.443)

[a] Standard deviations (intro bracket) have been computed from the variance–covariance matrix.

mates are univariate and they do not account for the correlation among estimates. As can be seen in Table 1, 95% CI excluded the zero value only for b_1 and b_2; this means that the drug is effectively degrading in batch #1 and that neither the differences in the intercepts nor in the slopes between batch #1 and the remainder ones are statistically significative.

Table 2 shows the estimates of intercepts and slopes as well as their standard deviations computed from coefficients in Table 1. These estimates agree with those published by Ruberg and Stegeman (1991), but differ in the standard deviations where we have assumed that the residual variance and the degrees of freedom are the same for every batch. Observed and predicted drug concentration as well as the lower 95% CI for each of the six batches are depicted in Fig. 1. It is important to realize that as for the computation of the matrix of variance–covariance of estimates (Eq. (5)), confidence intervals are constructed using a common variance for the residuals (Eq. (6)). This has two advantages: first, the confidence intervals for the predicted drug potencies are shorter; and second, the influence of outlyers in the prediction for a particular batch is decreased, which is particularly important when the number of points by batch is not large enough to avoid the bias introduced by the outlyers.

Following the example discussed by Ruberg and Stegeman (1991) we will set the proposed shelf-life equal to 3 years and the maximum allowable difference in the predicted concentration at the end of the shelf-life equal to 3%; these values give a maximum difference in slopes $\Delta\beta = \pm 1\%$/year. Data depicted in Table 1 show that 95% CI for the differences in slopes (a_6 to a_{10} coefficients) are inside the interval $[-1, +1]$ for batches #2, #3 and #4, but not for batches #5 (a_9) and #6 (a_{10}). In other words, there is a lack of power in data from these two batches to show what we have defined as an important difference in slopes. From Fig. 1 it is readily apparent that these results arise because the sampling interval (0–1.063 years) and the number of points (four in both cases) were too short. Concluding, only data from the first four batches can be combined to get a joint estimate of the slope.

Fig. 2 and Table 3 show the results for the insulin solution stability test. Results from the analysis of variance of regression showed that the F-test for the source of variation Regression$|b_1$ was equal to 358.9 with 3 and 46 df, large enough to reject the null hypothesis for $\alpha < 0.01$. As before, null hypothesis of no difference between slopes can be accepted as long as the 95% CI includes the zero value. Maximum difference between slopes was set equal to 1.311×10^{-4}% days^{-1}; this values arise assuming the shelf-life equal to 350 days and maximum difference in concentration at time equal to shelf-life in 3%. As can be seen in Table 3, the confidence interval for the difference between slopes is $(-5.83 \times 10^{-5}, 3.38 \times 10^{-5})$; therefore, we can conclude that this stability test has power enough to detect the maximum difference allowable and at the same time we conclude the equivalence of both batches.

In summary, statistical analysis of stability tests must address two problems: first, to test if the differences between the intercepts and between the slopes are equal to zero, and therefore data from different batches can be pooled; second, to test if the batches can be considered equivalents regarding their degradation rate. This second point means that the differences between degradation rates lack practical consequences, and as in bioequivalence trials; this does not mean that the differences must be equal to zero. It is important to realize that the requisite of differences between intercepts and between slopes equal to zero arise only if we pool the data in order to avoid the bias introduced by pooling data from batches with

different degradation rates, regardless of the relevance of such differences. This is why it is necessary to have a powerful test to detect the alternative hypothesis, and the reason to set α equal to 0.25 when the homogeneity of slopes is tested through the ANCOVA[1]; however, this strategy penalizes the studies involving a large number of data because the probability to reject the null hypothesis increases. The method outlined in this paper overcomes this problem because stability equivalence criteria are disclosed from the null hypothesis testing about slopes and intercepts. Moreover, this method always makes use of all the data analyzed to compute a global residual variance, and therefore, there is not need to pool the data to increase the number of degrees of freedom of the residual variance. Last but not least, we can analysis the data taking into consideration a stability equivalence criteria, for example, those defined by Ruberg and Stegeman (1991).

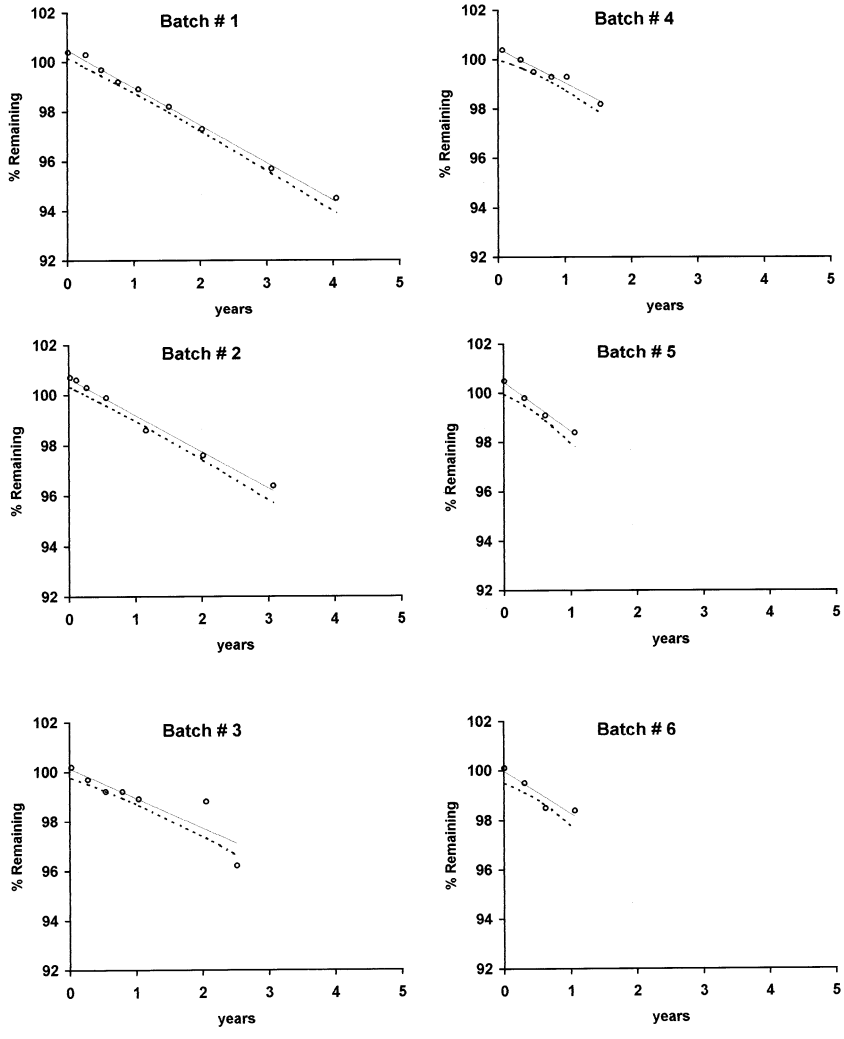

Fig. 1. Ruberg and Stegeman data. Experimental values, fitted linear model and lower bound one-side 95% confidence interval computed from joint estimate of residual variance. Ordinate, percent remaining; abscise, time in years.

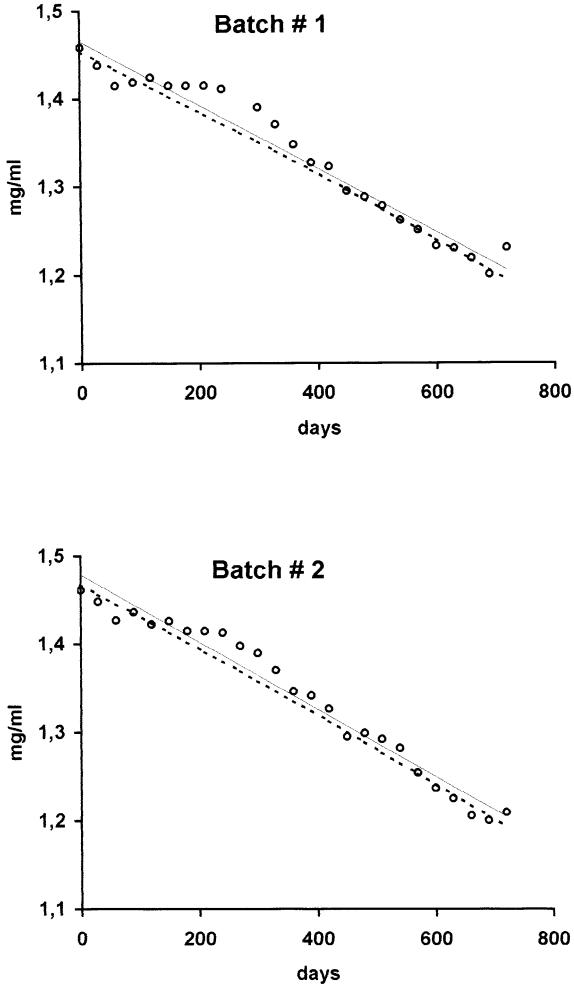

Fig. 2. Insulin data. Experimental values, fitted linear model and lower bound one-side 95% confidence interval computed from joint estimate of residual variance. Ordinate, concentration remaining in mg/ml; abscise, time in days.

Acknowledgements

This work have been financed by CICYT, project no. SAF 94–1379.

Appendix A

The following matrix corresponds to the general structure of the matrix **C** in Eq. (4) (except for the first row that has been included to facilitate this explanation).

$$\begin{pmatrix} \beta_1 & \beta_2 & \alpha_1 & \alpha_2 & \cdots & \alpha_{k-1} & \alpha_k & \alpha_{k+1} & \cdots & \alpha_{2k-2} \\ 1 & t_{1,1} & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & t_{n_1,1} & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 1 & t_{1,2} & 1 & 0 & \cdots & 0 & t_{1,2} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & t_{n_2,2} & 1 & 0 & \cdots & 0 & t_{n_2,2} & 0 & \cdots & 0 \\ 1 & t_{1,3} & 0 & 1 & \cdots & 0 & 0 & t_{1,3} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & t_{n_3,3} & 0 & 1 & \cdots & 0 & 0 & t_{n_3,3} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & t_{1,k} & 0 & 0 & \cdots & 1 & 0 & 0 & \cdots & t_{1,k} \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & t_{n_k,k} & 0 & 0 & \cdots & 1 & 0 & 0 & \cdots & t_{n_k,k} \end{pmatrix}$$

First column contains only ones, and the second one the sampling times for the k batches, from the first sampling time for the batch 1, $t_{1,1}$, to the last one for batch k, $t_{nk,k}$. These two columns account for the intercept, β_1, and the slope, β_2, of the first batch. The columns headed with α_1, α_2, α_{k-1}, account for the differences between the intercepts of the batches 2, 3…k and the intercept of the batch #1. Column headed with α_1 contains ones in those position corresponding to sampling times for batch #2 and zero elsewhere; column headed with α_2 contains

Table 3
Model parameter estimates for case II, insulin solution ($N = 50$; $s^2 = 3.065 \times 10^{-4}$ with 46 df)[a]

Model parameter	Estimate	Standard deviation	95% confidence interval	
b_1	1.47	6.79×10^{-3}	1.46	1.48
b_2	-3.69×10^{-4}	1.62×10^{-5}	-4.02×10^{-4}	-3.37×10^{-4}
a_1	9.21×10^{-3}	9.61×10^{-3}	-1.01×10^{-2}	2.86×10^{-2}
a_2	1.22×10^{-5}	2.29×10^{-5}	-5.83×10^{-5}	3.38×10^{-5}

[a] Coefficients b_1 and b_2 correspond to the intercept and slope for batch #1. Coefficients a_1 and a_2 correspond to the differences in intercept and slope, respectively.

ones in those position corresponding to sampling times for batch # 3 and zero elsewhere, and so on until column headed with α_{k-1}.

Columns headed with α_k to $\alpha_{2\cdot k-2}$ accounts for the differences between the slopes of batch 2, 3…k and the slope of the first batch. Column headed with α_k contains the sampling times of the batch # 2 in those positions corresponding to the sampling times of this batch. The same pattern is followed for the remaining columns.

If we are using standard statistical software or a multiple regression tool in a spreadsheet to do the calculations, we can omit the first column because the default models include the calculation of the intercept.

References

Bancroft, T.A., 1964. Analysis and inference for incompletely specified models involving the use of preliminary test of significance. Biometrics 20, 427–442.

Chow, S.C., Shao, J., 1991. Estimating drug shelf-life with random batches. Biometrics 47, 1071–1079.

Draper, N.R., Smith, H., 1981. In Applied Regression Analysis, 2nd ed. Wiley, New York.

Fairweather, W.R., Lyn, T.Y.D., Kelly, R., 1995. Regulatory, design and analysis aspects of complex stability studies. J. Pharm. Sci. 84, 1322–1326.

Farid, N.A., Atkins, L.M., Becker, G.W., Dinner, A., Heiney, R.E., Miner, D.J., Riggin, R.M., 1989. Liquid chromatographic control of the identity, purity and potency of biomolecules used as drug. J. Pharm. Biomed. Anal. 7, 185–188.

Hunter, W.G., Lamboy, W.F., 1981. A bayesian analysis of the linear calibration problem. Technometrics 23, 323–328.

Norwood, T.E., 1995. Statistical analysis of pharmaceutical stability data. Drug Dev. Ind. Pharm. 12, 553–560.

Oliva, A., Fariña, J.B., Llabrés, M., 1996. Influence of temperature and shaking in stability of insulin pharmaceutical preparations: degradation kinetics. Int. J. Pharm. 143, 163–170.

Ruberg, S.J., Stegeman, J.W., 1991. Pooling data for stability studies: testing the equality of batch degradation slopes. Biometrics 47, 1059–1069.

US Pharmacopoeia XXIII, 1995. US Pharmacopoeia Convention, Inc. Rockville, MD, 1959–1963.

Wolfran, S., 1988. Mathematica: A System for Doing Mathematics by Computer. Wolfram Research Inc. & Addison-Wesley Publishing Company, Inc.