MARCEL DEKKER, INC. • 270 MADISON AVENUE • NEW YORK, NY 10016

DRUG DEVELOPMENT AND INDUSTRIAL PHARMACY®
Vol. 29, No. 5, pp. 513–521, 2003

RESEARCH PAPER

Comparison of Shelf-Life Estimates for a Human Insulin Pharmaceutical Preparation Using the Matrix and Full-Testing Approaches

Alexis Oliva,* José Fariña, and Matías Llabrés

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

ABSTRACT

The design of a stability study is intended to establish a shelf-life based on testing a limited number of batches of a drug product, which is applicable to all future batches of the drug product manufactured under similar circumstances. In the FDA guidelines for stability testing of new drug products, the matrix design is suggested as the alternative statistical design to the full testing approach. In the last few years, the matrix design has received increasing attention from pharmaceutical companies in order to reduce the cost of stability studies. For this, the full and matrix stability test plans were used on three batches of human insulin pharmaceutical preparation stored under a kinetic mean temperature of 22.7°C for 3 years. The statistical method was suggested by Ruberg and Stegeman (Biometrics, **1991**, *47*, 1059–1069) and Llabrés et al. (Int. J. Pharm. **2000**, *204*, 61–68), and was used for comparing batches, and the maximum allowable difference in slopes between batches was used as criteria for estimating a common shelf-life. The results obtained show the conservative nature of the matrix approach, although the difference between the matrix and full shelf-life was less than 1 month. The effect of the matrix method on shelf-life was also studied. The results showed that the model for estimating the shelf-life varied as a function of the permutation used, after testing for both slope and intercept poolability at $\alpha = 0.25$. The batches should be pooled for both full testing and matrix design since the critical significance level required to obtain a power of 0.80 was achieved when the maximum allowable difference in slopes between batches was 1%/month, according to Ruberg and Stegeman. In contrast, when the proposed method by Llabrés et al. was used, the results varied in function of the maximum allowable difference in slopes.

Key Words: Stability; Human insulin; Batches; Shelf-life; ANCOVA.

*Correspondence: Alexis Oliva, Departamento Ingeniería Química y Tecnología Farmacéutica, Facultad de Farmacia, Universidad de La Laguna, Avda. Fco. Sanchez, s/n, 38200, La Laguna, Tenerife, Spain; Fax: +34-922-318506; E-mail: amoliva@ull.es.

DOI: 10.1081/DDC-120018640
0363-9045 (Print); 1520-5762 (Online)

INTRODUCTION

One important property of any drug product is its stability over time. Stability testing is performed to ensure that drug products retain their full efficacy up to expiry date under controlled temperature and humidity conditions. Stability information from long-term testing must be provided on at least three batches; it should cover a minimum of 12 months, with samples taken every 3 months over the first year, every 6 months over the second year, and annually thereafter.[1] Matrix stability testing can be applied as an alternative to traditional stability testing. With this design, only one subset of batches is tested at a given time point; at subsequent time points a different subset of batches is tested. As a result, shelf-life is estimated with fewer data points.[1,2]

Typically, multiple batches are manufactured and it is desirable to pool the data from different batches to obtain a single shelf-life for all the batches. There are several statistical methods for establishing a shelf-life for pharmaceutical products, but the decision to pool batches can only be made after their degradation rates (slopes) and initial values (intercepts) are determined to be not statistically significant.[3]

The Tripartite Guideline[1] recommends a significance level of 0.25 in the test for equality of batch degradation slopes and intercepts; this recommendation is based on work described by Bancroft.[4] If the test for equality of slopes and intercepts does not result in rejection for a significance level of 0.25, the data from the batches should be pooled. If it is inappropriate to combine data from several batches, the overall shelf-life depends on the minimum time a batch may be expected to remain within acceptable limits.

The most frequently used statistical method for comparing batches is the analysis of covariance (ANCOVA), proposed by Ruberg and Stegeman,[3] Bancroft,[4] and Norwood.[5] Application of a General Linear Model (GLM, analysis of covariance) to the statistical interpretation of stability data combines the methods of regression and variance analysis in a common model.[6] Alternative methods to test the equality of slopes and intercepts have previously been proposed. Norwood put forward a two-level nested ANCOVA model in which the differences among batches are interpreted both through the fixed effect initial potency and by the random interaction term *batch by time*. The method proposed by Chow and Shao[7] also considers the batch distribution as a random effect and their method is only suitable for postmarketing studies in which a large number of batches with relatively few points each are compared. Ruberg and Stegeman[3] specifically addressed the problem of fixing the maximum allowable difference in slopes. For these authors, the power of the test is fixed and the significance level of the test needed to obtain this power is calculated from the data. To do this, the alternative hypothesis must be established. Llabrés et al.[8] proposed a stability equivalence assessment based on comparison of the confidence intervals for the difference between batches, with the maximum allowable difference. This method allows the slopes and zero time intercept of one batch to be estimated, along with the difference and their corresponding confidence intervals with the remaining batches. In a second step, zero time intercepts and slopes are estimated for all batches.

Therefore, the decision to pool the data from the different batches to obtain a single shelf-life for all batches can be made only after their degradation rates (slopes) and initial values (intercepts) are deemed equal. For this, the objectives of this study were first, to verify if the pooling data for several batches were feasible using the matrix and full-testing approaches to determine a common shelf-life. Second: the statistical models proposed by Ruberg and Stegeman and Llabrés et al. were used for comparing batches, where the maximum allowable difference in slopes between batches is previously fixed. A human insulin pharmaceutical preparation stored at ambient temperature for 3 years was used as drug product; it is well known that insulin must be stored at refrigerated temperatures, but the study presented here is for illustrative purposes investigating both designs in terms of the power of detection of significant difference between slopes and evaluating the precision of estimated drug shelf-life.

EXPERIMENTAL

Analytical Method

A Waters apparatus was used consisting of a pump (Model 600E Multisolvent Delivery System), a UV–VIS detector (Model 490E Programmable Multiwavelength Detector), and the data acquisition software Maxima 820. Reversed-phase separation was carried out using a modified procedure published elsewhere[9] on a C-18 column (Delta Pack, 300 Å, 8×100 mm, Waters) and a mixture (v/v) of 0.2 M sodium sulfate buffer adjusted to pH 2.3 with orthophosphoric acid and acetonitrile (74:26) as mobile phase. The rest of the chromatographic conditions were as follows: flow rate of 1.0 mL/min, detection

at 214 nm, and injection volume 50 μL. Deionized water prepared with a MilliQ apparatus (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). Mobile phase was filtered and degassed. The percentages are given as w/v unless otherwise stated.

To validate the analytical method, seven standard solutions were prepared using pure human insulin (Batch: H01003, Novo Biolabs) at concentrations of 2–8 μg/mL. To quantify human insulin, we used the peak area. Each sample was analyzed four times. The variance analysis (ANOVA) of the linear regression confirmed the linearity of the method through rejection of the null hypothesis of deviation from linearity for a significance level of 0.05 ($\alpha = 0.05$); the coefficient of variation of predicted concentrations was 2.60%.[10] The analytical method used is specific ($r \geq 0.9990$), precise (CV < 1%), reproducible (CV < 1.22%), and robust. The limit of detection was 0.31 μg/mL and limit of quantitation was 0.93 μg/mL.[11]

In the case of the pharmaceutical preparation, 100-μL homogeneous samples were withdrawn from vials and stored at 4°C until analysis. Samples were directly diluted with 0.05 M HCl to obtain concentration values within the calibration range.

Long-Term Stability Study

Three batches of the same formulation and dosage form in the containers and closure intended for marketing were tested, and showed at least 1-year residual shelf-lives.

The samples were stored protected from light at thermostatically controlled room temperature for 3 years. The monthly mean temperature was 20.7 ± 1.6°C, while the mean kinetic temperature (MKT) was 22.7°C.[12] An enthalpy of activation of 67.0 KJ/mol was used, calculated according to the Arrhenius relationship using accelerated testing data.[13]

A full stability plan proposed by the Tripartite Guideline[1] was used, and the matrix stability plan proposed by Golden et al.[2] was used as an alternative to a full stability test plan. Both designs are shown in Table 1.

Statistical Model

The statistical model used for comparing k batches was analysis of covariance, following the model proposed by Ruberg and Stegeman[3]

$$Y_{ij} = \tau_i + \beta_i X_{ij} + \varepsilon_{ij}$$

where

Y_{ij} = jth response for the ith batch
τ_i = batch effects (intercept)
β_i = degradation rate of the ith batch
X_{ij} = time of the stability sample corresponding to Y_{ij}
ε_{ij} = random error corresponding to Y_{ij}.

The hypothesis to be tested is

$$H_o : \beta_i = \beta_j, \quad \text{for all } i, j$$

versus

$$H_a : \beta_i \neq \beta_j, \quad \text{for some } i \neq j.$$

If H_o is not rejected, a shelf-life for each batch is calculated assuming a common slope for all the batches. If H_o is rejected, shelf-life is determined by taking the shortest shelf-life calculated from the individual batches studied.

The Tripartite Guidelin[1] recommends using a significance level of 0.25 to determine whether the batches have a common slope and intercepts.[4] This test is conducted using a high level of significance in order to compensate for the low statistical power of the test. However, Ruberg and Stegeman[3] suggest that the power of the test must be fixed and the critical significance level calculated to reject the alternative hypothesis (H_a). To do so, the alternative hypothesis and the power to detect this alternative must be established.

The maximum allowable difference in slopes between batches, Δ, will be equal to the ratio $\Delta C/\theta$,

Table 1. Full and matrix stability test design used in this study proposed in the Tripartite Guideline[1] and by Golden et al.,[2] respectively.

	Full plan			Matrix plan		
Time (months)	Batch 1	Batch 2	Batch 3	Batch 1	Batch 2	Batch 3
0	X	X	X	X	X	X
3	X	X	X	X		
6	X	X	X		X	X
9	X	X	X		X	
12	X	X	X	X		X
18	X	X	X			X
24	X	X	X	X	X	
36	X	X	X	X	X	X

ΔC being the maximum allowable difference in drug potency at the end of the shelf-life and θ the desired shelf-life, assuming equivalent intercepts. In our case, three batches are involved in the analysis; all their slopes must be defined. The first alternative hypothesis assumes that one batch degrades more rapidly than the rest. We thus have:

$$H_{a1} : \beta_1 = \beta_2 = 0, \quad \beta_3 = \Delta.$$

The noncentrality parameter for H_a is computed with $\beta_k = \Delta$ corresponding to the batch that deviates from the rest, having sufficient power to detect the least favorable alternative.[3]

The noncentrality parameter, λ, corresponding to H_a is calculated as:

$$\lambda = \mathrm{SSH}/\sigma^2$$

$$\mathrm{SSH} = \sum_i w_i(\beta_i - \overline{\beta})^2$$

$$\overline{\beta} = \sum_i w_i\beta_i \Big/ \sum w_i$$

$$w_i = \sum_j (X_{ij} - \overline{X})^2$$

$$\overline{X} = \sum_j X_{ij}/n_i$$

where σ^2 is taken from the ANCOVA model allowing different slopes for each batch and the summations are for $i = 1,2,\ldots,k$ and $j = 1,2,\ldots,n_i$, the number of data points for the ith batch.[3] The significance level needed to achieve the specified power is given by

$$\alpha_c = 1 - F(f_c, k-1, \nu_\varepsilon, 0)$$

where

$$f_c = F^{-1}(\gamma, k-1, \nu_\varepsilon, \lambda)$$

$k-1$ being the degrees of freedom for H_o, ν_ε are the degrees of freedom for error, γ is the Type II error probability, and F is the noncentral F distribution function with noncentral λ.

The second alternative hypothesis is defined such that it minimizes SSH, and thereby λ, over all possible alternatives for a fixed Δ value.

$$H_{a2} : \beta_1 = 0; \quad \beta_{k-1} = \overline{\beta}; \quad \beta_k = \Delta$$

where batch 1 and k have the smallest weights (w_i).

Data Analysis

For statistical data analysis, the program SPSS version 8.0 for Windows by SPSS Inc. (Chicago, Illinois, USA) was used. Stability data processing was performed with Mathematica® Software for Windows.[14]

RESULTS AND DISCUSSION

The stability data set for human insulin pharmaceutical preparation is listed in Table 2 and fitted regression lines for each batch are plotted in Fig. 1 according to a full stability design. Before pooling

Table 2. Long-term stability testing data for human insulin pharmaceutical preparation.

Time (months)	% Remaining		
	Batch 1	Batch 2	Batch 3
0	100.1	100.2	100.8
3	97.2	98.6	99.8
6	96.7	96.2	98.1
9	95.7	96.0	96.4
12	92.6	91.8	92.7
18	86.0	87.0	88.7
24	83.1	83.1	83.4
36	74.0	75.0	76.0

Note: Three samples for each batch at each time point were taken, and the mean value was calculated. A total of 9 vials for each batch were used in this study.

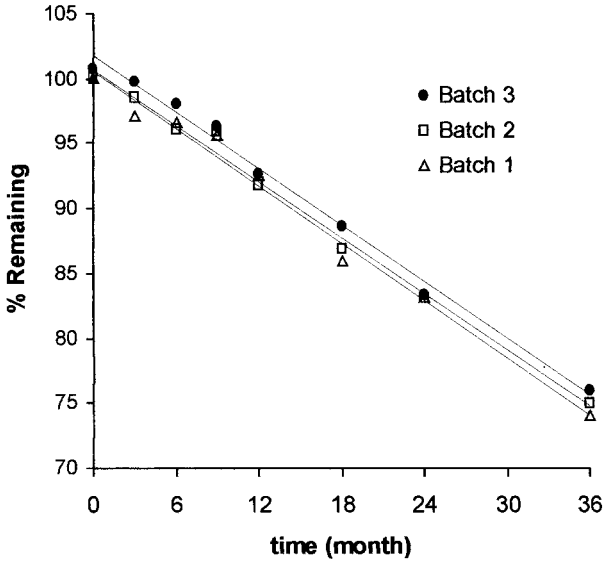

Figure 1. Stability data and individual regression lines corresponding to three batches of human insulin pharmaceutical preparation stored under a mean kinetic temperature of 22.7°C for 3 years according to the full stability test plan.

the data from several batches to determine a common shelf-life, a preliminary test is carried out to verify if the batches have a common slope and intercept. If batch-to-batch variability is small, i.e., the relationship between the parameter of interest and time is essentially the same from batch to batch, stability data should be combined into an overall estimate. The decision to combine the data for several batches is made if there are strong reasons in favor.

Table 3 lists the estimated slopes, intercepts, and shelf-life calculated separately for each batch, and Table 4 shows the ANCOVA results of the equality test for batch degradation slopes and intercepts for the full and matrix testing approaches, respectively.

The shelf-life estimated for both the full and matrix data sets are evidence of the stability of this human insulin pharmaceutical preparation at room temperature. The matrix stability plan applied offers a reduction in analytical testing without a significant loss in the ability to detect stability trends over time. This fact is reflected in the length of the estimated shelf-life. The results obtained indicate the conservative nature of the matrix approach, although the differences between the matrix and full shelf-life were minimal, around 1 month in the worst case (see Table 3). The matrix shelf-life was lower than those for the full data set as a result of less precise estimates for the regression line. Note that the 95% confidence limits are increased in the matrix data

Table 3. Individual batch regression and shelf-life calculations for full testing and matrix approaches.

Batch	Intercept (%)	Slope (%/months)	S_{yx}[a]	Individual shelf-life (months)	Shelf-life[b] (months)
		Individual batch regression for full plan			
1	100.59	−0.7357	1.115	13.15	13.28 [12.40–14.16][c]
2	100.75	−0.7225	0.840	13.78	13.67 [12.91–14.43]
3	101.81	−0.7264	0.802	15.16	15.12 [14.37–15.88]
		Pooled batch regression with equivalent slope and intercept			
3	101.05	−0.7282	1.032		14.58 [14.14–15.02]
		Individual batch regression for matrix plan			
1	100.13	−0.7153	0.793	12.39	12.43 [11.32–13.54]
2	100.90	−0.7198	1.065	13.10	13.22 [11.94–14.50]
3	101.39	−0.7038	0.653	14.55	14.36 [13.36–15.36]
		Pooled batch regression with equivalent slope and intercept			
3	100.82	−0.7142	0.953		14.39 [13.83–14.95]

[a]Root mean square error.
[b]Shelf-life assuming a common slope model.
[c]95% confidence intervals.

Table 4. ANCOVA results of test for slope and intercept homogeneity for the model proposed by Ruberg and Stegeman.[3]

Source	d.f.[a]	Sum of squares	Mean square	*F*-value	*P*-value
		ANCOVA results for full plan			
Batch	2	2.772	1.386	1.595	0.230
Time	1	1603.9	1603.9	1845.7	<0.001
Time X batch	2	0.109	0.0545	0.0627	0.939
Error	18	15.646	0.869		
		ANCOVA results for matrix plan			
Batch	2	1.767	0.863	1.210	0.342
Time	1	1278.3	1278.3	1750.9	<0.001
Time X batch	2	0.109	0.0547	0.075	0.928
Error	9	6.571	0.730		

[a]d.f. Degree of freedom.

set due to the differences in the number of test points used, five vs. eight for the full test.

For full testing, the ANCOVA results indicate that the data can be pooled across batches since the *P*-value for the *time by batch* interaction (homogeneity of slopes) was higher than the prescribed significance level of 0.25 ($P = 0.939$) and a common estimated slope was obtained. Despite minor differences in the intercept estimates, ($\approx 1\%$), we must consider the separate intercept model for full testing since the *P*-value for intercepts ($P = 0.230$) was less than 0.25. Thus, the shelf-life for the insulin solution will correspond to the shortest period calculated from the individual batches studied, consequently, it will be that batch 1 of 13.28 months.

For the matrix design, the equality test for slopes and intercepts did not result in rejection for a significance level of 0.25, since the observed *P*-values were always higher than 0.25. The data from several batches can be combined, resulting in a common shelf-life of 14.4 months. Comparing the results for batch 1, the shelf-life is 12.4 months for the individual regression and 14.4 for the pooled regression, illustrating the advantages of pooling data across batches. These results could be due to lower power of the matrix design as a result of power measurements, five points vs. eight points for full design. Therefore, the chance to pool data from different batches will increase for poorly designed studies (small number of experimental points and the error variance is increased) than for well-designed studies.

The effect of matrix on shelf-life was also studied. Because the batches were arbitrarily designated as 1, 2, or 3 for calculating the matrix shelf-life, six permutations are possible for assigning the three batches to the three rows of the matrix plan mentioned earlier. The data points and fitted regression lines for the 123, 321, and 213 permutation corresponding to the matrix plan are plotted in Fig. 2. The null hypothesis for equality of slope was accepted for each of the six permutations, and the test for equality for zero time intercepts was accepted in four of the six combinations for a significance level of 0.25. These results indicate that the model used for estimating shelf-life was different. For the combined model (slopes and intercepts pooled), the average shelf-life was 14.4 months, whereas for the common slopes model (only slopes pooled), the average shelf-life was 12.4 months, since the shortest shelf-life must be taken.

Table 5 shows the estimated batch slopes for the matrix and full-testing approaches under the two alternative hypotheses described in Materials and Methods as well as the noncentrality parameter for

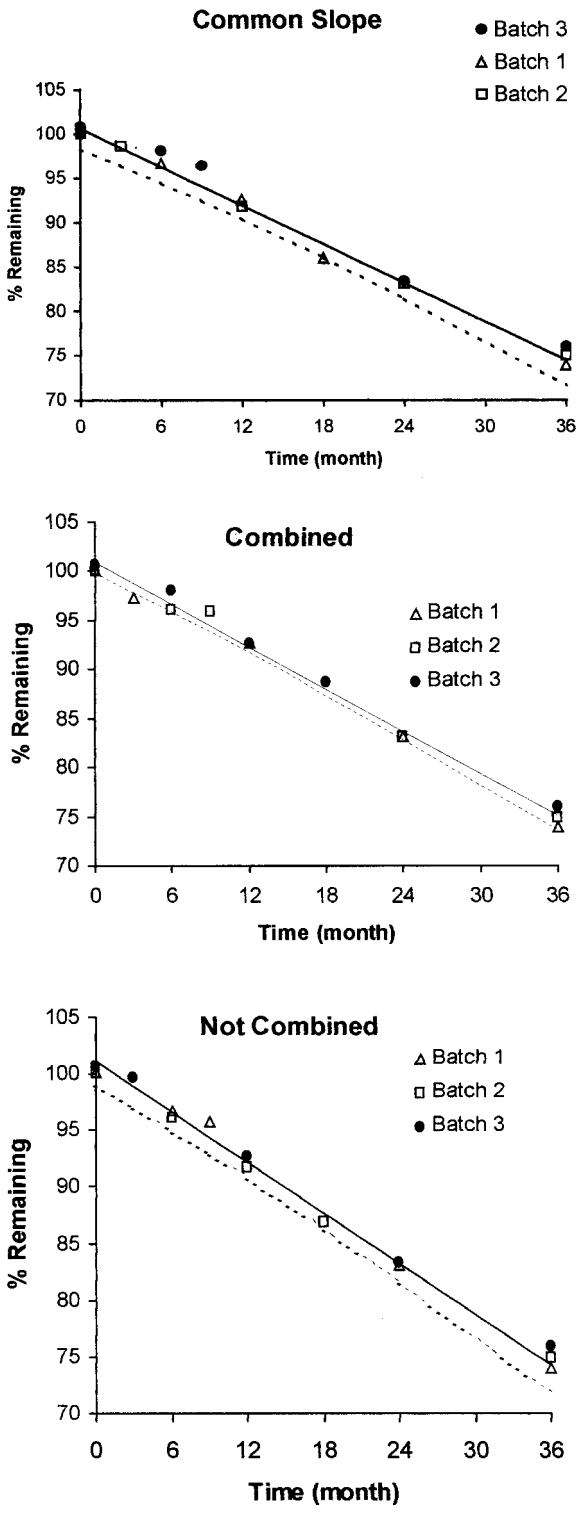

Figure 2. Shelf-life for matrix design proposed by Golden et al.[2] The shelf-life varied between 14.4 months for the combined model and 12.4 months for the noncombined model. In all cases, shelf-life was calculated from the lower 95% confidence limit on the regression line.

Table 5. Variance (σ^2) taken from the covariance analysis, critical significance level (α_c) for both alternative hypotheses, and *P*-value for the null hypothesis of slope homogeneity for the full and matrix stability test plan according to Ruberg and Stegeman.[3] The average shelf-life was 14.4 months.

						α_c			
	b_0	b_1	S.S.	d.f.	σ^2	Ha_1	Ha_2	*P*	$t_{10\%}$ (months)
					Full plan				
	100.6	−0.7363	15.646	18	0.869	0.367	0.469	0.939	14.6
	100.7	−0.7219							
	101.8	−0.7266							
Permutation					Matrix plan				
123	100.1	−0.7153	6.571	9	0.730	0.377	0.488	0.928	14.4
	100.9	−0.7198							
	101.4	−0.7038							
132	100.1	−0.7153	6.014	9	0.668	0.346	0.460	0.884	14.2
	101.8	−0.7253							
	100.2	−0.7050							
213	100.4	−0.7097	4.329	9	0.481	0.233	0.347	0.443	14.5
	101.1	−0.7455							
	101.4	−0.7038							
231	100.4	−0.7097	6.947	9	0.772	0.426	0.505	0.735	14.3
	101.8	−0.7253							
	100.6	−0.7442							
312	101.2	−0.7123	4.544	9	0.505	0.278	0.364	0.490	14.4
	101.1	−0.7455							
	100.2	−0.7050							
321	101.2	−0.7123	7.719	9	0.858	0.461	0.535	0.776	14.3
	100.9	−0.7198							
	100.6	−0.7442							

Abbreviations: b_0: intercept (%); b_1: slope (%/months); S.S.: sum of squares; d.f.: degree of freedom; Ha_1, Ha_2: alternative hypothesis.

a power of 0.80 and the critical significance level (α_c) needed to achieve that power according to Ruberg and Stegeman.[3] At this stage, the desired shelf-life (θ) was 14 months, and the maximum allowable difference (ΔC) in drug potency was 1%. For both the full testing and matrix designs, α_c was higher than 0.25, indicating that batches may be pooled, independently of the alternative hypothesis (see Table 5). In this way, poorly designed studies (small number of experimental points, large error variance, few sampling times, etc.) are penalized because the resulting critical significance levels will increase.

Llabrés et al.[8] have proposed an alternative statistical method for testing the equivalence between batches. This method is based on the statistical linear model, making use of a set of dummy variables to code the different batches. The main advantages of this method are the possibility of comparing several batches, to disclose the equivalence stability criteria from the statistical hypothesis about the equality between slopes, and the joint estimate of the residual variance whatever the decision to pool or not the data from different batches. Table 6 shows the model parameter estimates for full testing and matrix plan, using the method described by Llabrés et al.[8]

We also analyzed the data taking into consideration the stability equivalence criteria, for example, those defined by Ruberg and Stegeman,[3] where the maximum allowable difference in slopes among any two batches is equal to Δ. The data from two different batches are pooled if two conditions are fulfilled: first, the 95% confidence intervals for the difference in slopes are contained in the interval $[-\Delta, +\Delta]$, and second, that this interval contains the zero value. Thus, the first condition assures that the power of the test is sufficient to detect a difference among the batches considered relevant, while the second avoids the bias of pooling batches with different slopes. If we take a Δ value equal to 0.0714%/month (e.g., $\Delta = 1\%/14$ months), the 95% confidence intervals for the difference between slopes are outside the interval $[-0.0714, 0.0714]$ for the batches in both testing

Table 6. Model parameter estimates for full testing and matrix plan using the method described by Llabrés et al.[8]

Plan	Model parameter	Estimate	Standard deviation	95% confidence intervals	
				Lower	Upper
Full	b_1	−0.7363	0.0294	−0.7980	−0.6746
	a_3	0.0144	0.0415	−0.0728	0.1017
	a_4	0.0097	0.0415	−0.0776	0.0969
Matrix					
123	b_1	−0.7153	0.0284	−0.7798	−0.6509
	a_3	−0.0045	0.0407	−0.09652	0.0876
	a_4	0.0116	0.0420	−0.0835	0.1067
132	b_1	−0.7153	0.0272	−0.7770	−0.6537
	a_3	−0.0100	0.0389	−0.0981	0.0781
	a_4	0.0103	0.0402	−0.0807	0.1013
213	b_1	−0.7097	0.0231	−0.7620	−0.6574
	a_3	−0.0358	0.0330	−0.1105	0.0389
	a_4	0.0059	0.0341	−0.0713	0.0831
231	b_1	−0.7097	0.0293	−0.7759	−0.6434
	a_3	−0.0157	0.0418	−0.1103	0.0790
	a_4	−0.0345	0.0432	−0.1323	0.0633
312	b_1	−0.7123	0.0237	−0.7659	−0.6587
	a_3	−0.0331	0.0338	−0.1097	0.0434
	a_4	0.0073	0.0350	−0.0718	0.0864
321	b_1	−0.7123	0.0309	−0.7822	−0.6425
	a_3	−0.0074	0.0441	−0.1072	0.0923
	a_4	−0.0318	0.0456	−0.1349	0.0712

Coefficient b_1 corresponds to the slope for batch 1. Coefficients a_3 to a_4 correspond to the differences between the slope of batches 2 to 3 and the slope for batch 1. As can be seen, the 95% confidence intervals for the difference between slopes were outside the interval $[-0.0714, 0.0714]$ if the maximum allowable difference was 1% and the shelf-life was 14 months, indicating that the batches should not be pooled. However, the batches may be pooled if the maximum allowable difference was 3%, the shelf-life being of 14 months.

approaches, and therefore data from different batches should not be pooled. In contrast, if the maximum allowable difference (ΔC) was increased to 3% and the desired shelf-life (θ) was 14 months, then the 95% confidence intervals for the difference between slopes are inside the interval $[-0.214, 0.214]$ for both designs, and an average shelf-life should be estimated, 14.4 ± 0.14 months ($n = 6$) for the matrix plan, a value very close to that obtained for full testing, 14.6 months.

In conclusion, the matrix plan should yield only one shelf-life for all the configurations; however, the data obtained in this study do not support this fact. The estimated shelf-life varied between 12.4 and 14.6 months in function of the permutation used. The equality test of slopes was accepted for all the configurations and the average shelf-life was 12.7 months considering the common slope model, against 13.2 months for the full plan, indicating the conservative nature of the matrix plan.

If we used the procedure proposed by Ruberg and Stegeman,[3] a common shelf-life for both alternative hypotheses could be estimated, independently of the design used, without meaning a loss of consumer protection against one batch that degrades more rapidly than the rest. In contrast, with the method proposed by Llabrés et al.,[8] a common shelf-life could only be calculated when the maximum allowable difference was increased to 3% and the desired shelf-life was 14 months.

Finally, the statistical models used in this study are an alternative to those proposed by USP and at the same time they warrant the same objectives. However, special attention to the definition of a significant difference in slopes among batches is required, which should be a priori stated, since in the multiple comparisons context, statistically significant differences in batches may be found although the differences may be of little practical importance.

On this point, what is the limit for this, and what should be the criteria to establish it? At first, the consumer's risk of receiving a product which does not fulfil the requirements must be controlled and the consumer must be adequately protected. This is a topic of further research on our part.

ACKNOWLEDGMENT

This work was financed by the Gobierno de la Comunidad Autónma de Canarias as part of Project PI-082/2000.

REFERENCES

1. The Stability Testing of New Drug Substances and Products—The Tripartite Guideline. US Pharmacopeia XXIII. US Pharmacopeial Convention: Rockville, MD, 1995; pp. 1959–1963.
2. Golden, M.H.; Cooper, D.C.; Riebe, M.T.; Carswell, K.E. A matrixed approach to long-term stability testing of pharmaceutical products. J. Pharm. Sci. **1996**, *85*, 240–245.
3. Ruberg, S.J.; Stegeman, J.W. Pooling data for stability studies: testing the equality of batch degradation slopes. Biometrics **1991**, *47*, 1059–1069.
4. Bancroft, T.A. Analysis and inference for incompletely specified models involving the use of preliminary test of significance. Biometrics **1964**, *20*, 427–442.
5. Norwood, T.E. Statistical analysis of pharmaceutical stability data. Drug. Dev. Ind. Pharm. **1986**, *12*, 553–560.
6. Langenbucher, F. Statistical analysis of stability data by means of general linear models (GLM). Drug. Dev. Ind. Pharm. **1991**, *17*, 165–181.
7. Chow, S.C.; Shao, J. Estimating drug shelf-life with random batches. Biometrics **1991**, *47*, 1071–1079.
8. Llabrés, M.; Oliva, A.; Fariña, J. Statistical assessment of between batch stability equivalence. Int J. Pharm. **2000**, *204*, 61–68.
9. Farid, N.A.; Atkins, L.M.; Becker, G.W.; Dinner, A.; Heiney, R.E.; Miner, D.J.; Riggin, R.M. Liquid chromatographic control of the identity, purity and potency of biomolecules used as drugs. J. Pharm. Biomed. Anal. **1989**, *7*, 185–188.
10. Hunter, W.G.; Lamboy, W.F. A Bayesian analysis of the linear calibration problem. Technometrics **1981**, *23*, 323–328.
11. Oliva, A.; Fariña, J.; Llabrés, M. Development of two high-performance liquid chromatographic methods for the analysis and characterization of insulin and its degradation products in pharmaceutical preparations. J. Chromatogr. B **2000**, *749*, 25–34.
12. Haynes, J.D. Worldwide virtual temperature for product stability testing. J. Pharm. Sci. **1971**, *60*, 927–929.
13. Oliva, A.; Fariña, J.; Llabrés, M. Influence of temperature and shaking on stability of insulin preparations: Degradation kinetics. Int. J. Pharm. **1996**, *143*, 163–170.
14. Wolfram, S. *Mathematica: A System for Doing Mathematics by Computer*; Wolfram Research, Inc. & Addison-Wesley Publishing Company, Inc., 1988.