ELSEVIER

Journal of Pharmaceutical and Biomedical Analysis
15 (1996) 329–338

JOURNAL OF
PHARMACEUTICAL
AND BIOMEDICAL
ANALYSIS

Partial least-squares regression for the quantitation of pharmaceutical dosages in control analyses

M. Blanco*, J. Coello, M. Elaamrani, H. Iturriaga, S. Maspoch

Departamento de Química, Unidad de Química Analítica, Universidad Autónoma de Barcelona, E-08193 Bellaterra, Barcelona, Spain

Received for review 6 February 1996

Abstract

A spectrophotometric method for the simultaneous determination of the active principle and a flavouring agent in syrups containing additional excipients is proposed. The calibration matrix must include all the variability expected in the samples and this is achieved using laboratory-made mixtures and production samples in order to ensure correct results. The optimum number of principal components for the regression model was selected by using various procedures. The proposed method was used to quantify samples from different production batches. The results are compared with those provided by HPLC.

Keywords: Anethole; Malate of clebopride; Partial least-squares calibration; Simultaneous determination; UV–visible

1. Introduction

UV–visible spectrophotometry is a rapid, inexpensive analytical technique and as such is highly suitable for control analyses of pharmaceutical preparations, the components of many of which absorb in the ultraviolet region. However, the lack of specificity of UV–visible absorption hinders application in the presence of overlap between bands of different components. Pharmaceutical preparations are usually mixtures of the active principle and various excipients that absorb in the same region as the component of interest, thereby resulting in band overlap and impeding the use of the technique with simple calibration methods. The inception of microcomputers and spectrophotometers that allow absorbance spectra to be expeditiously recorded at many wavelengths has enabled the development of analytical methods based on the mathematical resolution of multivariate signals for the rapid quantitation of mixtures of analytes in control analyses [1,2]. The most common choice for multicomponent determinations of this type, multiple linear regression (MLR), allows the easy resolution of complex mixtures provided the contribution of each mixture component to the overall spectrum can be determined from spectra of the pure components used for calibration [3,4]. While rather limited in

* Corresponding author. Tel.: (+34)3-5811367; Fax: (+34)3-5812477; e-mail: iqan1@cc.uab.es

PII S0731-7085(96)01859-6

scope [5], the method provides good results if background noise is fairly low and all the mixture components are known, absorb appreciably, and do not interact with one another. When one or more of these conditions is not met and some analyte absorbs weakly at a high concentration (e.g. some excipients in pharmaceutical preparations) of the analyte contributions to the mixture absorbance are rather disparate, a different mathematical procedure is required. Foremost among these is partial least-squares regression (PLSR) [6–8], which is also used with other analytical techniques where interactions between analytes are quite strong (e.g. in IR and NIR spectroscopy).

Implementation of this calibration technique entails constructing the calibration model with samples identical to those that are to be subsequently determined in routine control analyses. This may be troublesome since pharmaceuticals are manufactured to stringent standards as regards content variability, so that obtaining production samples encompassing a wide enough range for correct calibration is rather difficult. One other difficulty is determining the number of principal components (PCs) required to correctly define the calibration model. While commercially available software can automatically calculate the optimum number of PCs, it usually relies on highly conservative selection criteria as it has been designed for use in systems with relatively low reproducibility of measurements; consequently, the results cannot be directly extrapolated to UV–visible spectroscopy, where reproducibility is usually very high. Therefore, underfitting may arise using the automatically proposed model.

This paper reports an analytical method for the determination of the active principle in Flatoril®, a pharmaceutical marketed in suspended form, and one of its excipients, using first-derivative UV–visible spectroscopy, and of hydrogen malate of clebopride for calibration purposes. The use of synthetic samples to expand the concentration range for calibration was investigated with a view to the quantitation of production samples. Various procedures for selection of the optimum number of PCs to be used in the calibration model were tested and the results compared.

2. Theory

2.1. PLSR

The principles behind PLSR have been thoroughly described in the literature [6,7]. For a system of m samples, PLSR simultaneously resolves matrices $\boldsymbol{X}(m \times k)$ and $\boldsymbol{Y}(m \times p)$, which contain the absorbances of spectra recorded at k wavelengths and the concentrations of the p analytes to be quantified respectively.

Each of these matrices is resolved into the product of two smaller ones (the scores and loadings matrices), which include all relevant information from $\boldsymbol{X}$ and $\boldsymbol{Y}$. The loadings for matrix $\boldsymbol{X}$ are calculated from the scores of the concentration matrix, $\boldsymbol{Y}$, whereas those for matrix $\boldsymbol{Y}$ are obtained from the scores of the spectroscopic data matrix, $\boldsymbol{X}$.

Each matrix is resolved into the sum of a principal components ($a \leqslant k$) in order to simultaneously calculate the following matrices:

$$\boldsymbol{X} = \boldsymbol{T}\boldsymbol{P}^{\mathrm{T}} + \boldsymbol{E} = \sum_{h=1}^{a} \boldsymbol{t}_h \boldsymbol{p}_h^{\mathrm{T}} + \boldsymbol{E}$$

$$\boldsymbol{Y} = \boldsymbol{U}\boldsymbol{Q}^{\mathrm{T}} + \boldsymbol{F} = \sum_{h=1}^{a} \boldsymbol{u}_h \boldsymbol{q}_h^{\mathrm{T}} + \boldsymbol{F}$$

where matrices $\boldsymbol{T}$ and $\boldsymbol{U}$, of dimensions $(m \times a)$, are the score matrices for $\boldsymbol{X}$ and $\boldsymbol{Y}$ respectively. Similarly, matrices $\boldsymbol{P}^{\mathrm{T}}$ and $\boldsymbol{Q}^{\mathrm{T}}$ [of dimensions $(a \times k)$ and $(a \times p)$ respectively] are the loading matrices for $\boldsymbol{X}$ and $\boldsymbol{Y}$, and $\boldsymbol{E}$ and $\boldsymbol{F}$ are the residual matrices for blocks $\boldsymbol{X}$ and $\boldsymbol{Y}$ when aPCs are used. Superscript T denotes transposed matrices. The two resolved matrices are not independent; in fact, there is an inner relationship between the scores of blocks $\boldsymbol{X}$and $\boldsymbol{Y}$, so that for each PC h

$$\hat{\boldsymbol{u}}_h = b_h \boldsymbol{t}_h$$

where the caret symbol ($\wedge$) denotes a calculated quantity and b_h is the regression coefficient for each PC.

Using the previous equation, $\boldsymbol{Y}$ can be calculated from $\hat{\boldsymbol{u}}_h$:

$$\boldsymbol{Y} = \boldsymbol{T}\boldsymbol{B}\boldsymbol{Q}^{\mathrm{T}} + \boldsymbol{F}$$

where $\boldsymbol{B}$ is the matrix of regressors b_h, of dimensions $(a \times a)$.

Once the scores and loadings matrices for blocks $\boldsymbol{X}$ and $\boldsymbol{Y}$ have been obtained, the concentrations of unknown samples can readily be calculated from the scores for the unknown samples, $\boldsymbol{t}^*$

$$\boldsymbol{t}_h^* = \boldsymbol{E}_h^* \; {}_1\boldsymbol{w}_h$$

$$\boldsymbol{E}_h^* = \boldsymbol{E}_h^* \; {}_1 - \boldsymbol{t}_h^* \boldsymbol{p}_h^{\mathrm{T}}$$

where $\boldsymbol{w}_h$ denotes the weights for the data in calibration $\boldsymbol{X}$, calculated using the PLSR algorithm. Such weights are not classical weighting factors, but intermediates in the determination of the loadings for block $\boldsymbol{X}$ in the calibration $(\boldsymbol{p}_h^{\mathrm{T}})$.

The $\boldsymbol{E}_0^*$ values used to calculate the first PC $(h = 1)$ are original data. The asterisk denotes new samples used to derive the predictions.

Finally, $\boldsymbol{Y}$ is obtained from

$$\hat{\boldsymbol{Y}} = \sum_{h=1}^{a} b_h \boldsymbol{t}_h \boldsymbol{q}_h^{\mathrm{T}}$$

With a single dependent variable $(p = 1)$, the algorithm is called PLS1. If there are p variables in matrix $\boldsymbol{Y}$, PLS1 can be repeated or, alternatively, an algorithm called PLS2 can be used to find one set of PCs yielding good predictions of all variables in $\boldsymbol{Y}$ simultaneously.

2.2. Selection of the number of PCs for the model

The selection of the number of PCs used to construct the model is the most critical step in implementing multivariate calibration. Malinowski [9] has reviewed reported criteria for choosing the optimum number of PCs to be included in a PLS model with cross-validation. Wold [10] suggested using the absolute minimum of a plot of the prediction error sum of squares (PRESS) obtained in the cross-validation against the number of PCs; this procedure is widely used but can result in overfitting, as pointed out by some authors. The commercial software package Unscrambler [11] uses a much more conservative criterion which has also been adopted by other authors [12]; it employs the number of PCs resulting in the first local minimum in the PRESS vs. PC plot. Haaland and Thomas [13] performed an F-test to compare the number of PCs in the model yielding the minimum PRESS, n^*, with all the models with a smaller number of PCs $(n < n^*)$ and adopted as optimal the model with the fewest PCs whose PRESS is not significantly greater than the minimum value. Schematically, the algorithm is applied as follows: for a system of m samples containing p analytes, several models including a different number of principal components $(n = 1, 2 \ldots n^*$, where n is the number of PCs for the absolute minimum) are constructed and an $F(n)$ value is calculated from

$$F(n) = \frac{\text{PRESS}(n)\,(\text{model with } n \text{ PCs})}{\text{PRESS}(n^*)\,(\text{model with } n^* \text{ PCs})}$$

and compared with the tabulated value of $F_{(mp,mp,\alpha)}$. A value of $\alpha = 0.25$ is recommended based on a purely empirical criterion.

The number of PCs for the model can also be selected by using an F-test [12] to check whether introducing an additional principal component has a significant effect. To this end, an $F(\text{exp})$ value is calculated from

$$F(\text{exp}) = \frac{\dfrac{\text{PRESS}(n) - \text{PRESS}(n+1)}{k}}{\dfrac{\text{PRESS}(n+1)}{mk - (n+1)k}}$$

and compared with the tabulated value of $F_{[k,mk-(n+1)k,\alpha]}$, where k is the number of variables and $\alpha = 0.05$.

This method is similar to that employed by Eastment and Krzanowski [14], who selected the number of PCs from the change in PRESS as additional components were included in the model. They calculated a parameter W from

$$W = \frac{\dfrac{\text{PRESS}(n-1) - PRESS(n)}{D_M}}{\dfrac{\text{PRESS}(n)}{D_R}}$$

$$D_M = m + k - 2n$$

$$D_R = k(m-1) - \sum_{r=1}^{i=n} (m + k - 2i)$$

that they compared with unity.

Even though the last three methods are conceptually similar (they determine the significant number of PCs by using significance criteria), the Haaland and Thomas method establishes a comparison with the absolute minimum, whereas those of Osten and Eastment and Krzanowski compare differences between models constructed from consecutive numbers of PCs, so they may be less efficient if PRESS decreases very little in each step.

In this work, the above-described selection methods were used to determine the optimum number of PCs for constructing models with a view to the quantitation of synthetic and production samples of pharmaceuticals. Both the PLS1 and PLS2 algorithms were tested in order to ascertain which led to the smaller errors in the quantitation of samples other than those included in the calibration matrix.

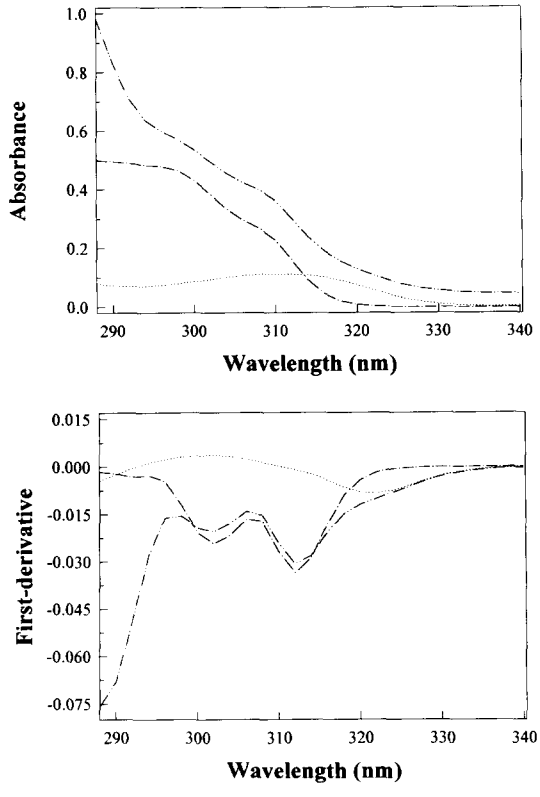

Fig. 1. Absorbance and first-derivative spectra for the pharmaceutical and the products analysed. Flatoril (– · · – · ·) was treated as described in Section 3 and diluted to 30 mg ml^{-1}. The concentrations of anethole (– · – · –) and hydrogen malate of clebopride (· · ·) (30.4 and 4.13 $\mu g\ ml^{-1}$ respectively) were the nominal values in the preparation after dilution.

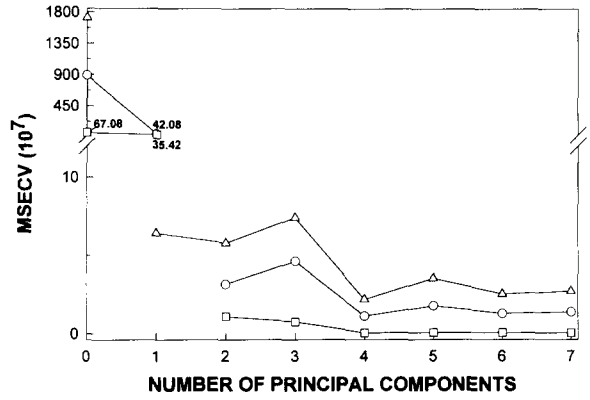

Fig. 2. Variation of MSECV with the number of PCs in calibration matrix I as applied to synthetic samples: (○) overall calibration with PLS2; (□) hydrogen malate of clebopride with PLS1; (△) anethole with PLS1.

3. Experimental

3.1. Sample and reagents

Ethanol, methanol, acetone, HCl and H_3PO_4(Panreac, pro-analysis grade) were used. The acids were employed at 1 M (HCl) and 0.01 M (H_3PO_4) concentrations. HPLC-grade methanol (Sharlau) and 1-hexanesulphonic acid sodium salt for liquid chromatography (Panreac) were used in HPLC analyses.

Samples contained the pharmaceutical Flatoril®, from Laboratorios Almirall (Barcelona, Spain), which consists of hydrogen malate of clebopride [hydrogenhydroxybutanedioate of 4-amino-5-chloro-2-methoxy-*N*-(1-phenylmethyl-4-piperidynyl)benzamide] as the active principle. The preparation also contains silicone, benzoic acid (preservative) and other excipients, including flavourings, sweeteners, thickeners, etc., and is commercially available as a suspension with a syrupy texture.

Standard solutions of hydrogen malate of clebopride (0.2273 $g\ l^{-1}$), anethole (0.3324 $g\ l^{-1}$) and the excipients were made by weighing and then dissolving in methanol.

All samples were supplied by Laboratorios Almirall and were assayed without further purification.

3.2. Apparatus and software

UV visible spectra were recorded on a Hewlett-Packard HP 8451A diode array spectrophotometer. Absorbance measurements were made at 2nm intervals in the 250–350 nm region using an integration time of 1 s.

Chromatographic analyses were carreid out on an assembly comprising Kontron HPLC System 600 pumps and a Hewlett-Packard 1040A HPLC UV–Vis diode array detector equipped with a 9152C data station, also from Hewlett-Packard.

The experimental set-up also included a reversed-phase Spherisorb ODS-2C_{18} column (15 cm long × 0.4 cm i.d., 5 μm particle size) and a C_{18} pre-column.

For PLSR calibration, recorded spectra were imported from the instrument into the program Unscrambler v. 5.0, developed by CAMO A/S.

3.3. Procedure

3.3.1. Synthetic mixtures of the pharmaceutical components

Overall, 28 synthetic samples, containing hydrogen malate of clebopride (3.6×10^{-3}–9.9×10^{-3} g l^{-1}), anethole (1.33×10^{-2}–4.65×10^{-2} g l^{-1}) and constant concentrations of the other excipients equal to that of the pharmaceutical, were prepared. Each sample was supplied with silicone from a solution containing 0.5 g in 10 ml of methanol, and centrifuged. Then 5 ml of the clear phase was withdrawn in order to ensure a similar composition to that of real samples at the time of analysis. Each sample was supplied with sufficient 1 M HCl and made up to volume in a 50 ml calibrated flask in order to obtain a solution that was 0.1 M HCl. The solutions were centrifuged to remove the precipitate formed and their absorbance measured against methanol–HCl (1 M) (90:10 v/v) as blank.

Table 1
Optimum number of PCs for the calibration matrix as determined from various criteria

Criterion	PLS1		PLS2 total
	Clebopride	Anethole	
Wold	4	4	4
Unscrambler	2	1	2
Haaland–Thomas	4	4	4
Osten	4	1	2
Eastment–Krzanowski	4	1	2

Table 2
RSEPP values of synthetic samples from the prediction set obtained by using calibration matrix I and models including a variable number of PCs

PC	PLS2			PLS1	
	Clebopride	Anethole	Total	Clebopride	Anethole
1	23.54	2.89	5.37	13.55	2.89
2	4.86	2.88	2.98	4.78	2.78
3	3.12	2.77	2.79	3.70	1.91
4	0.74	0.89	0.88	0.74	0.89
5	0.80	1.08	1.07	0.77	1.09
6	0.81	1.16	1.14	0.87	1.16
7	0.74	1.26	1.24	0.98	1.26

3.3.2. Determination of the commercial preparation

≈1.8 g of commercial formulation was placed in a 50 ml calibrated flask, supplied with about 25 ml of methanol–HCl (1 M) (90:10 v/v) and ultrasonicated for 2–5 min. Once cool, the solution was made up to volume with the same solvent. An aliquot of ≈10 ml was then centrifuged in a tube in order to remove the precipitate and the UV–Vis spectrum for the clear phase was recorded over the wavelength 250–350 nm against methanol–HCl (1 M) (90:10 v/v) as blank.

3.3.3. HPLC analyses

≈15 g of thoroughly homogenized sample was supplied with 25 ml of acetone and ultrasonicated for a few minutes. After cooling to room temperature, the solution was made up to 50 ml with acetone. About 10 ml of the suspension formed was centrifuged and the clear phase was passed through a Millipore filter of 0.5 μm pore size.

Table 3
Relative errors made in the quantitation of synthetic samples from the prediction set by using a model of PCs

Sample	Added ($\times 10^{-4}$ mg ml^{-1})		PLS1		PLS2	
	Clebopride	Anethole	Clebopride	Anethole	Clebopride	Anethole
1	90.90	265.92	0.17	0.23	0.18	0.23
2	36.36	265.92	2.57	0.39	2.57	0.39
3	54.55	132.96	0.31	−0.62	0.31	−0.62
4	54.55	265.92	0.42	−0.52	0.41	−0.52
5	54.55	332.40	0.72	−0.01	0.71	−0.01
6	59.09	199.44	−0.21	1.32	−0.21	−1.32
7	59.09	332.40	−0.37	−0.75	−0.37	−0.75
8	63.64	332.40	0.70	0.50	0.70	0.50
9	63.64	465.36	−1.13	−0.16	−1.14	0.17
10	74.54	371.52	−0.04	0.83	0.03	0.83
11	74.54	412.80	0.18	1.13	0.19	1.13
12	49.70	288.96	1.59	2.29	1.59	2.29
$RSEPP_{Analyte}$			0.74	0.89	0.74	0.89
$RSEPP_{Total}$			0.88		0.88	

Then, a 20 μl aliquot was injected into the HPLC system and eluted with a mobile phase consisting of methanol–H_3PO_4 (0.01 M) containing 0.01 M 1-hexanesulphonic acid sodium salt (75:25 v/v) at a flow rate of 0.8 ml min^{-1}. Three chromatograms per sample were obtained at 310 nm. The retention times were approximately 3 and 7.5 min for hydrogen malate of clebopride and anethole respectively. Peak areas were calculated and the results interpolated on a calibration curve previously run for each compound.

4. Results and discussion

Fig. 1 shows the absorbance and first-derivative spectra for a treated sample of the pharmaceutical that was diluted to 30 mg ml^{-1}, as well as those for anethole and hydrogen malate of celbopride at their nominal concentrations (with provision for the dilution performed).

4.1. Quantitation of synthetic samples

Hydrogen malate of clebopride and anethole were quantified in synthetic samples containing variable concentrations of both analytes in addition to the other excipients. For this purpose, first-derivative spectra were recorded over the wavelength range 288–330 nm and PLSR was used for calibration. The results were mean-centred and each analyte was quantified both individually (using the PLS1 algorithm) and jointly (with PLS2). Calibration matrix I was constructed from 16 synthetic samples that were validated using the cross-validation method and as many cancellation groups as samples were included in the calibration matrix (i.e. the leave-one-out method).

The variance in the prediction of samples by use of cross-validation (MSECV) and the leave-one-out method is defined as

$$\mathrm{MSECV} = \frac{\mathrm{PRESS}}{m} = \frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{p}(C_{ij} - \hat{C}_{ij})^2$$

where m is the number of samples in the calibration matrix, p is the number of analytes quantified in the sample and C_{ij} is the experimental concentration. The caret symbol denotes the calculated value.

Fig. 2 shows the variation of MSECV with the number of PCs used in the model. As can be seen,

Table 4
Concentrations obtained using HPLC and relative errors made in the quantitation of commercially available samples of the pharmaceutical using calibration matrix I, which consisted solely of synthetic samples

Batch	Found HPLC ($\times 10^{-4}$ mg ml^{-1})		Relative error (%)			
			PLS1		PLS2	
	Clebopride	Anethole	Clebopride	Anethole	Clebopride	Anethole
1a	43.10	351.80	8.59	−6.45	8.65	−6.44
1b	40.82	333.20	9.47	−6.52	9.53	−6.51
1c	38.57	314.80	9.53	−6.86	9.59	−6.85
2a	40.95	341.16	6.38	−6.15	6.42	−6.14
2b	43.95	366.14	4.50	−6.25	4.53	−6.25
2c	44.45	370.33	4.59	−6.28	4.62	−6.27
3a	40.14	278.74	12.77	−6.08	12.80	−6.08
3b	43.07	299.08	10.35	−5.94	10.38	−5.93
3c	44.93	312.03	10.22	−5.90	10.25	−5.90
4a	44.09	317.43	8.46	−6.28	8.50	−6.28
4b	44.40	319.65	7.35	−6.23	7.39	−6.23
4c	45.26	325.81	7.20	−6.49	7.24	−6.48
$RSEPP_{Analyte}$			8.51	6.30	8.54	6.30
$RSEPP_{Total}$			6.34		6.34	

MSECV in the calibration of clebopride using PLS1 decreased abruptly as the second PC was added and then more gradually up to the fourth PC. In contrast in the PLS1 calibration of anethole and in the PLS2 calibration, MSECV increased from the third and fifth PC respectively.

The optimum number of PCs obtained by using each of the above-described methods is given in Table 1. All provided four PCs for the quantitation of clebopride using PLS1, in contrast with only two PCs suggested by Unscrambler. The results obtained for the quantitation of anethole by PLS1 and those with PLS2 were of two types: on the one hand, Unscrambler and the Osten and Eastment–Krzanowski methods suggested the same number of PCs; on the other hand the Haaland–Thomas method suggested a number coinciding with that defining the absolute minimum.

For easier comparison and interpretation of the results in the quantitation of samples from the prediction set, they were expressed as the relative standard error of prediction (RSEPP), given by

$$\text{RSEPP (\%)} = \sqrt{\frac{\sum_{i=1}^{m}\sum_{j=1}^{p}(C_{ij}-\hat{C}_{ij})^2}{\sum_{i=1}^{m}\sum_{j=1}^{p}C_{ij}^2}} \times 100$$

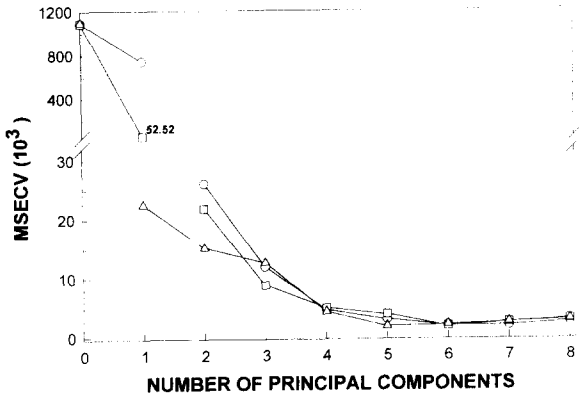

Fig. 3. Variation of MSECV with the number of PCs in calibration matrix II as applied to commercially available samples of the pharmaceutical: (○) overall calibration with PLS2; (□) hydrogen malate of clebopride with PLS1; (△) anethole with PLS1.

Table 5
Optimum number of PCs for the calibration matrix II as predicted from various criteria

Criterion	PLS1		PLS2 total
	Clebopride	Anethole	
Wold	6	5	7
Unscrambler	3	1	3
Haaland–Thomas	6	5	6
Osten	6	5	6
Eastment–Krzanowski	6	5	6

The expression to be used to determine RSEPP for each analyte, $RSEPP_{anal}$, was the result of making $p = 1$ in the previous equation.

In order to find which of the criteria used to select the number of PCs resulted in the best quantitation of samples other than those included in the calibration matrix, the 12 synthetic samples making up the prediction set were quantified. Table 2 shows the RSEPP values obtained using models involving a variable number of PCs. The lowest RSEPP values for both analytes with both PLS1 and PLS2 were obtained with the models including the number of PCs determined using the criteria of Haaland and Thomas or Wold.

Table 3 shows the relative errors for each analyte in each individual sample provided by the model with four PCs. As can be seen, the errors were very small and RSEPP values differed very little between PLS1 and PLS2.

Table 6
RSEPP values for the prediction set consisting of commercially available samples obtained by using calibration matrix II and models with a variable number of PCs

PC	PLS2			PLS1	
	Clebopride	Anethole	Total	Clebopride	Anethole
1	49.53	4.30	7.70	6.79	4.32
2	8.21	4.98	5.05	5.62	3.26
3	3.91	2.45	2.48	3.91	2.32
4	4.24	1.80	1.86	4.22	1.80
5	3.21	0.99	1.06	2.76	0.91
6	1.01	0.87	0.87	1.00	1.63
7	1.29	1.63	1.63	1.66	1.45
8	1.46	1.41	1.41	1.34	1.41

4.2. Quantitation of production samples

Overall, 12 commercially available samples of the pharmaceutical from four different batches were treated as described in Section 3 and analysed by HPLC in order to obtain the reference concentrations; simultaneously, their UV–Visible spectra were recorded and measured using calibration matrix I, which was constructed from synthetic samples and the model with four PCs. Table 4 gives the relative error for each analyte in each sample, together with %RSEPP.

Unlike the synthetic samples, the results were very poor, with large systematic errors (overestimated clebopride concentrations and underestimated anethole concentrations). This indicates that the matrix used for the quantitation of laboratory samples does not account for all the variability present in production samples. Two effects may explain the differences. In the first place, as their concentrations were kept constant, the calibtration matrix did not mode the influence on the spectra of some excipients which present a slight absorption at the lower wavelengths of the working range. The second effect is more important, and is related to the incomplete precipitation of silicone by using methanol. In this way, any slight change in the composition of the excipients in the pharmaceutical was modelled as a change in the concentration of the analytes and calibration matrix I was not suited for the quantitation of production samples.

Constructing a calibration matrix from production samples alone was impossible since they did not encompass a wide enough range for correct calibration. Therefore, the variability in calibration matrix I was increased by introducing four production samples (one per batch). This new calibration matrix, which comprised synthetic and production samples, was named matrix II. Data were autoscaled and the cross-validation and leave-one-out methods were applied for validation. Both the PLS1 and PLS2 algorithms were used.

Table 7
Relative errors obtained in the quantitation of commercially available samples of the pharmaceutical using calibration matrix II

Batch	Found HPLC ($\times 10^{-4}$ mg ml^{-1})		Relative error (%)			
			PLS1		PLS2	
	Clebopride	Anethole	Clebopride	Anethole	Clebopride	Anethole
1a	43.10	351.80	−1.57	0.22	−1.57	0.13
1b	40.82	333.20	−0.30	−0.18	−0.29	−0.26
1c	38.57	314.80	−1.05	−1.35	−1.05	−1.46
2a	40.95	341.16	1.36	−0.84	1.37	−0.79
2b	43.95	366.14	−0.26	−0.61	−0.26	−0.56
2c	44.45	370.33	0.58	−0.96	0.60	−0.91
3a	40.14	278.74	1.71	0.24	1.72	0.00
3b	43.07	299.08	0.28	0.56	0.29	0.37
3c	44.93	312.03	−0.32	1.03	−0.32	0.82
4a	44.09	317.43	0.64	1.34	0.63	1.31
4b	44.40	319.65	−0.92	1.06	−0.92	0.99
4c	45.26	325.81	−1.47	1.32	−1.47	1.26
$RSEPP_{Analyte}$			1.00	0.91	1.01	0.87
$RSEPP_{Total}$			0.91		0.87	

Fig. 3 shows the MSECV values obtained as a function of the number of PCs for the new calibration matrix. As can be seen, the absolute minimum with PLS2 appeared at seven PCs, but the difference between the MSECV for six PCs was very small. PLS1 produced a sharp minimum at five PCs for anethole and six PCs for clebopride.

The optimum number of factors determined according to each criterion is given in Table 5.

The number of PCs needed to describe synthetic and production samples jointly was greater than in the previous case since variability between sample types must also be accounted for. In this case, Unscrambler provided a number of PCs that differed markedly from the rest. All other criteria suggested a large number of PCs for modelling synthetic and production samples in combination: even those of Osten and Eastment–Krzanowski, which proved very conservative in modelling the synthetic samples.

Table 6 gives the %RSEPP values calculated for the commercial samples of the prediction set using calibration matrix II. The lowest RSEPP values were obtained by using the number of PCs suggested by the Haaland–Thomas, Ostern and Eastment–Krzanowski criteria.

The relative quantitation error for each analyte in each sample is listed in Table 7. The results obtained with PLS1 and PLS2 were very similar; only two samples exceeded 1.5% for clebopride, and none for anethole.

Table 8 shows the results obtained in the analysis of four different batches of the commercially available pharmaceutical using the proposed spectrophotometric method with multivariate calibration. They were very similar to those provided by HPLC.

5. Conclusions

Control analyses of pharmaceutical preparations based on UV–Visible spectrophotometry and PLSR calibration have proved a competent alternative to HPLC. Besides covering a suitable range of concentrations of the analyzed components, the calibration matrix must include the expected variability of the excipients not quantified in order to correctly quantify production samples. In this case, the calibration matrix

Table 8
Mean concentrations for the two analytes, μg (g suspension)$^{-1}$, in different batches, analysed by HPLC and the proposed spectrophotometric method

Analyte	Batch 1			Batch 2			Batch 3			Batch 4		
	HPLC	PLS1	PLS2	HPLC	PLS1	PLS2	HPLC	PLS1	PLS2	HPLC	PLS1	PLS2
Malate acid clebopride	123.8	122.6	122.6	123.1	123.8	123.8	116.3	116.9	116.9	118.8	118.1	118.1
%RSD	2.61	0.64	0.65	3.49	0.80	0.81	2.54	1.04	1.05	2.27	1.09	1.04
Anethole	1010.4	1006.0	1005.1	1025.5	1017.30	1017.7	807.6	812.5	810.8	855.3	865.9	865.4
%RSD	1.13	0.81	0.83	2.35	0.18	0.18	4.20	0.40	0.41	0.21	0.15	0.17

was constructed from synthetic samples in addition to a few production samples.

Of the different criteria used to select the optimum number of PCs, only those of Haaland and Thomas led to the model that subsequently resulted in the lowest RSEPP values. The method using the first local minimum resulted in marked underfitting, whereas those of Osten and Eastment–Krzanowski performed similarly overall.

The results obtained in the analysis of the pharmaceutical using the proposed spectrophotometric method were consistent with those provided by HPLC.

Acknowledgements

The authors are grateful to the Spanish Dirección General de Investigación Científica Técnica (DGICyT) for financial support granted for the realization of this work as part of Project PB93-0899 and to Laboratorios Almirall for complimentary supplies of the pharmaceutical samples assayed. M. E. also thanks the Instituto de Cooperación con el Mundo Arabe for the award of a scholarship.

References

[1] W. Landvoigt, Mikrochim. Acta, Part I, (1985) 265–280.
[2] B.V. Fisher and R. Jones, J. Pharm. Biomed. Anal., 5 (1987) 455–467.
[3] G. Sala, S. Maspoch, H. Iturriaga, M. Blanco and V. Cerdà, J. Pharm. Biomed. Anal., 6 (1988) 765–772.
[4] M. Blanco, J. Coello, H. Iturriaga, S. Maspoch, T. Gómez-Cotín, S. Alaoui-Ismaili and E. Rovira, Fresenius', J. Anal. Chem., 351 (1995) 315–319.
[5] K.R. Beebe and B.R. Kowalski, Anal. Chem., 59 (1987) 1007A–1017A.
[6] P. Geladi and B.R. Kowalski, Anal. Chim. Acta, 185 (1986) 1–17.
[7] H. Martens and T. Naes, Multivariate Calibration, Wiley, Chichester, UK, 1989.
[8] M. Blanco, J. Coello, F. González, H. Iturriaga, S. Maspoch and X. Tomás, J. Pharm. Biomed. Anal., 12 (1994) 509–514.
[9] E.R. Malinowski, Factor Analysis in Chemistry, Wiley, New York, 1991, pp. 98–121.
[10] S. Wold, Technometrics, 20 (1978) 397–406.
[11] Unscrambler User's Guide, Computed Aided Modelling A/S, Trondheim, Norway, 1993.
[12] D.W. Osten, J. Chemomet., 2 (1988) 39–48.
[13] D.M. Haaland and E.V. Thomas, Anal. Chem., 60 (1988) 1193–1202.
[14] H.T. Eastment and W.J. Krzanowski, Technometrics, 24 (1982) 73–77.