Available online at www.sciencedirect.com
SCIENCE DIRECT®
Talanta 64 (2004) 597–602
Talanta
www.elsevier.com/locate/talanta

Strategies for constructing the calibration set for a near infrared spectroscopic quantitation method

M. Blanco*, M.A. Romero, M. Alcalà

Department of Chemistry, Analytical Chemistry Unit, Faculty of Science, Autonomous University of Barcelona, E-08193 Bellaterra, Barcelona, Spain

Received 28 November 2003; received in revised form 18 February 2004; accepted 22 March 2004
Available online 20 May 2004

Abstract

Three strategies for the construction of calibration sets have been tried, with the objective to develop and to validate a NIR quantitation method.

The first two approaches consist of the use of two types of samples, named: samples of laboratory obtained by mixing the ingredients that compose the drug, and doped samples obtained by under- and over-dosed production samples. In order to improve the prediction results, production samples have been added to each calibration model. The ensuing models were validated with a view to determine their fitness for purpose. However, spectral differences between the laboratory samples and doped samples resulted in spurious predictions in quantifying samples of one type using the model developed from samples of the other.

Such differences were studied in depth and a third procedure has been proposed, based on a calibration model constructed with an unique type of sample (laboratory sample) for later to correct it with a few doped samples. This corrected model has a good predictive ability on production samples.

Keywords: Near infrared spectroscopy; NIR; Calibration; Pharmaceutical analysis

1. Introduction

The last two decades of the 20th century have no doubt seen the greatest expansion of near infrared spectroscopy (NIRS). The simplicity, precision, and expeditiousness of this technique, in addition to an improved knowledge of the chemometric tools required to apply NIRS-based methodologies, have extended its use to virtually all industrial areas.

The pharmaceutical industry has shown special interest in the NIRS technique on account not only of its expeditiousness and non-destructive character, but also, especially, of its flexibility for both qualitative analysis (e.g. in the identification of raw materials and finished products [1,2], reaction monitoring in process control operations [3,4], monitoring of blending processes [5], control of film coating procedures [6]) and quantitative analysis (e.g. in the determination of active principles in commercially available preparations [7,8], moisture [9] or even polymorphs [10]).

Proper development of an analytical methodology entails using samples representative of that to be analyzed and spanning an adequate concentration range. This is especially important in the NIRS technique, where spectra depend not only on the chemical properties of the sample, but also on physical properties of its matrix including particle size, shape and distribution, or degree of compaction, all of which significantly affect the spectroscopic signal. Consequently, the calibration samples used should be representative of chemical variability (in the concentrations of the active principle and excipients) and physical variability (associated with the manufacturing process and arising from particle size, the degree of compaction, etc.). Meeting both requirements in constructing a calibration model is usually difficult as the active principle and excipient concentrations are very close to the nominal value in virtually all samples of the pharmaceutical preparation. A number of procedures

* Corresponding author. Tel.: +34-93-581-1367; fax: +34-93-581-1367.
E-mail address: marcel.blanco@uab.es (M. Blanco).

0039-9140/$ – see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.talanta.2004.03.027

have been developed with a view to overcome this problem, however.

One approach involves preparing laboratory samples by mixing accurately weighed amounts of the active principle and excipients in appropriate proportions in order to expand the concentration range spanned to the desired bounds. This is probably the most simple and convenient choice; also, it allows one to design a sample preparation approach that minimizes correlation, facilitates the development of robust models and provides highly reliable reference values (weighings) for the analyte. However, laboratory samples are not obtained using the same procedure as production samples, so some of the physical variability in the manufacturing process is not included in the calibration process.

One other approach involves preparing synthetic samples at a pilot plant reproducing the operations of the production plant. This method is much more labour-intensive and expensive; also, its feasibility depends on the particular type of sample and on the concentration of the active principle in the preparation, and the ensuing model may incorporate some source of variability not present in the production samples. This method is usually employed in the analysis of tablets using transmission measurements.

A third choice involves under- and over-dosing production samples with small amounts of the excipients and active principle, respectively, in order to extend the original concentration range. This method is somewhat more laborious than preparing laboratory samples, but undoubtedly more expeditious than the pilot plant method. The procedure has the advantage that differences between doped (under- or over-dosed) samples and production samples are smaller than with laboratory samples as the addition of small amounts of the excipients or active principle does not alter matrix effects, so the ensuing calibration models are usually more simple. Correct doping requires that samples be in powdered or granular form, which may entail applying some sample pretreatment.

The only use of samples prepared by one of the processes commented above is not enough to obtain calibration models with an acceptable degree of accuracy in prediction of production samples. This fact is due to the physical differences between sample sets used in construction models and those production. The solution to build these models correctly usually involves expanding the calibration set, using, together with prepared samples, production samples in order to incorporate this physical variability. This procedure has been demonstrated to be effective and the models obtained are robust and accurate [11]. The ratio of prepared samples versus production samples to be used in the calibration set depends on the nature of the samples and also on the variability to cover. Generally, four or five production samples are enough to incorporate this variability source.

This paper compares the performance of three calibration procedures based on laboratory and doped samples in the development and validation of a method for the determination of an active principle in a pharmaceutical preparation.

2. Experimental

2.1. Apparatus and software

NIRS spectra were recorded on a NIRSystems 6500 near infrared spectrophotometer from Foss NIRSystems (Raamsdonksveer, The Netherlands) equipped with a reflectance detector and a model AP6641ANO4P fiber-optic probe. The instrument was governed via a PC computer running the software Vision 2.22, also from Foss NIRSystems, for data acquisition. Laboratory samples were homogenized in a Turbula Type T2C Mixer from WAB (Basel, Switzerland).

Spectral pretreatments and multivariate calibration were both done using Unscrambler 7.5 from CAMO (Trondheim, Norway).

2.2. Samples

The pharmaceutical preparation studied was a granulate with anti-inflammatory action containing nimesulide as active principle and sucrose as major excipient. Laboratory samples were prepared by weighing the different components of the preparation in pure form and mixing them in variable proportions to span a concentration range ±50% around the nominal content in the active principle. Since the final aim of the models to be built was to predict an sole concentration (active), no experimental design had been used to prepare the different calibration sets and the proportion of the different amounts of ingredients (active and excipients) had been designed to minimize correlation between them. The mixtures were blended to homogeneity and their NIR spectra recorded.

Doped samples were obtained by supplying production samples of known concentration with also known amounts of the active principle (over-dosed samples) or a mixture of excipients (under-dosed samples). Following doping, the samples were homogenized and their NIR spectra recorded. As with the laboratory samples, an active principle concentration range ±50% around the nominal content was thus encompassed.

The spectra of the sample types (laboratory, doped, and production) and the active principle (nimesulide) are shown in Fig. 1. As can be seen, the doped and production spectrum are practically equal, whereas the laboratory and doped samples have substantial differences along the whole spectrum.

All samples (production specimens, the active principle, and excipients) were supplied by Laboratorios Menarini (Badalona, Spain).

Overall 29 production samples (from as many different batches), 21 laboratory samples and 28 doped samples were used.

2.3. UV reference method

The active principle (nimesulide) content in the production samples was determined by ultrasonicating

approximately 0.25 g of sample with 35 ml of MeOH for 10 min. The solution was then diluted to 50 ml with MeOH and a 5 ml aliquot was supplied with 5 ml of water and 2.5 ml of 1 M HCl, and made to 50 ml with 1:1 MeOH/H_2O. The nimesulide content in this solution was determined by applying multiple linear regression (MLR) to the first derivative spectrum in the 250–450 nm wavelength range using pure nimesulide as standard of calibration. This procedure suppresses the systematic error produced by base line displacement due to the presence of an excipient dispersion.

3. Results and discussion

3.1. Calibration models

All models were constructed using the PLS1 algorithm on the 1100–2200 nm wavelength range and the different wavelength pretreatments assayed were: standard normal variate (SNV), detrending, first and second derivatives (applying Savitzky–Golay algorithm). Although the active principle has some bands in the 2200–2498 nm spectral range (Fig. 1), it was despised due to the spectral noise associated with the fiber-optic probe. Samples were split between the calibration and prediction sets to construct the two models, based on laboratory and doped samples. The calibration models have been constructed by cross-validation (using the method leave-one-out) and the optimal number of PLS factors chosen like the minimum in the graph of residual variance versus the number of factors. Their predictive ability was assessed via the root mean square error of prediction (RMSEP). The models constructed with only laboratory samples or doped samples (the calibration set has an unique set of each type of samples) result in poor prediction abilities: the laboratory model present a RMSEP value of 19.8, predicting the complete set of production samples, and the doped samples model yields better results (RMSEP = 2.8) predicting the same set of production samples, and both models show systematics errors.

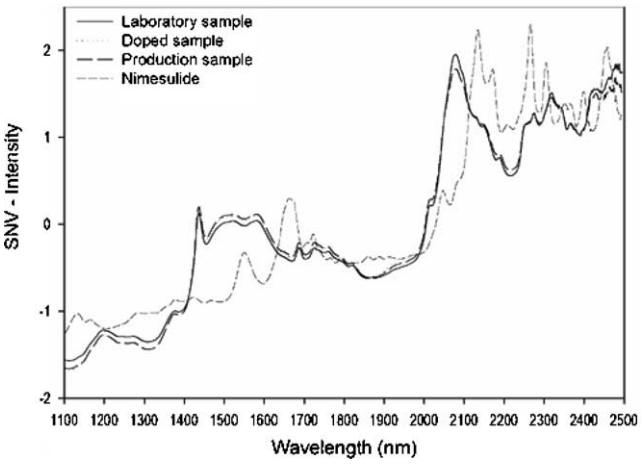

Fig. 1. Spectra for different sample types and active principle. The active principle concentration in the different sample types is the nominal concentration of the drug (±5%).

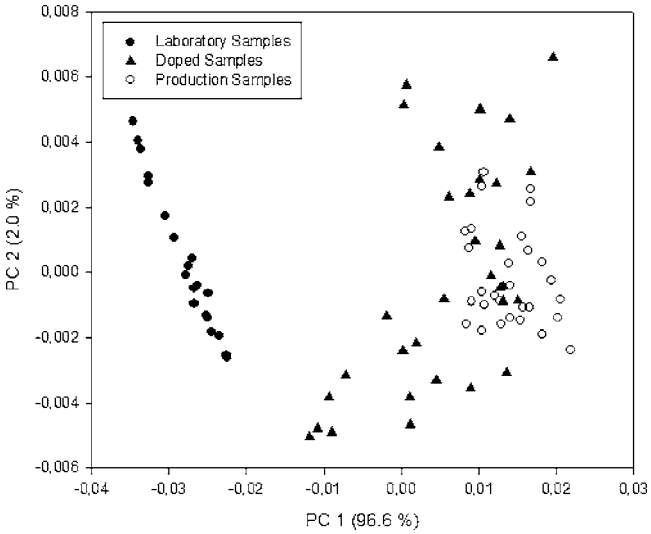

Fig. 2. First and second principal components of the PCA conducted on laboratory samples (●), doped samples (▲), and production samples (○). First-derivative spectra in the 1100–2200 nm range has been used.

The differences between sample origins can be observed in a plot of PC2 versus PC1 in a principal component analysis (Fig. 2) of production, laboratory, and doped samples. As can be seen, the spectral differences between laboratory samples and doped/production samples are significant. In order to improve the predictive ability of the models, each set was expanded with a given number of production samples that were the same for both calibration sets, as well as with those added to both prediction sets. Table 1 shows the characteristics of the best models for each type of sample, as well as the results they provided.

The results show that the accuracy has been improved considerably. This fact means that there is a source of variability in the production samples not covered by the laboratory and doped samples.

The most salient parameters of both models were very similar. In fact, both were constructed from first-derivative spectra, had the same number of PLS components and even similar predictive abilities; however, the model based on doped samples performed slightly better. In principle, both

Table 1
Characteristics of the models based on laboratory and production samples

	Laboratory model		Doped model	
	Calibration	Prediction	Calibration	Prediction
Samples	10 lab. 6 prod.	4 lab. 8 prod.	12 dop. 6 prod.	6 dop. 6 prod.
RMSEP	1.4	1.7	0.9	1.0

Both models were constructed using the wavelength range 1100–2200 nm, 1st derivative spectral preteatment and four PLS factors. Lab., laboratory; dop., doped; prod., production.

Table 2
Correlation coefficients obtained in the identification of unknown samples using the spectral library

Library classes	Samples analyzed		
	Pharmaceutical preparation	Nimesulide	Sucrose
Pharmaceutical preparation	0.999–1.000	−0.142 to −0.120	0.957–0.979
Nimesulide		0.997–0.998	−0.229 to −0.152
Sucrose			0.996–0.999

Correlation ranges are the extreme values of five samples from each class identified in the library.

models provided accurate results and were suitable for determining the active principle. However, their actual suitability was ascertained by validation.

3.2. Validation of the methods

Validating an analytical method entails determining whether it fulfills its intended purpose (i.e. its "fitness for purpose"). The two NIR methods corresponding to the previous calibration models were validated in order to determine whether they would allow the accurate quantitation of the active principle in the pharmaceutical preparation. To this end, their selectivity, accuracy, repeatability, intermediate precision, linearity, and robustness were determined, following the ICH guidelines [12].

3.2.1. Selectivity

The selectivity of a NIR method cannot be assessed as in other analytical methods. The proposed procedure involves identifying the pharmaceutical preparation in a library including various classes corresponding to the preparation and its pure components. As this identification step does not require the use of the quantitation models, we used the same validation procedure with both.

Five different samples from as many batches per product, belonging to each class defining the spectral library constructed for this purpose, were identified. The correlation ranges were established from the extreme values obtained in the identifications. The production samples were always identified as the pharmaceutical preparation, with a threshold of 0.98. Although the correlation coefficients of the production samples with sucrose were always high, none was confused with the excipient, so all were accurately identified (see Table 2).

3.2.2. Linearity

The linearity of a multivariate calibration model is evaluated by plotting the results for a series of samples spanning a given concentration range against their reference values. The linearity of the two calibration models was assessed by using samples of the same type in each calibration (i.e. doped samples with the doped model and laboratory samples with the laboratory model). The linearity results are shown in Table 3. As can be seen, both models were linear throughout the concentration range studied.

3.2.3. Accuracy

Fifteen production samples were used to compare the active principle concentrations provided by both models with the reference values. A paired *t*-test of differences was conducted to this end that revealed the NIR values not to be significantly different from the reference values. As can be seen from Table 3, both models provided accurate values.

3.2.4. Repeatability

Repeatability was evaluated by having the same operator replicate the determination of the active principle in the same sample at least six times on the same day. Table 3 shows the results, alongside their standard deviations and percent coefficients of variation (% CV).

3.2.5. Intermediate precision

Intermediate precision was determined to establish between-day and between-operator variability. To this end, two operators replicated the determination of the active content in the same sample on 3 different days. The data thus obtained were subjected to a statistical study to determine their standard deviation and % CV, as well as to a variance analysis intended to establish whether either effect was significant. Table 3 shows the results obtained for both models.

3.2.6. Robustness

Robustness was assessed by checking the results obtained in the determination of active principle with both models using samples collected over a period of 1.5 year. Table 3 shows the results of a test of differences between the results and the reference values. As can be seen, both models provided results consistent with the reference values, so both can be assumed to be robust.

From Table 3 it follows that both models are suitable for quantifying the active principle as the two provide acceptable results for each validation parameter. The model using doped samples provides slightly better results as regards repeatability and intermediate precision; both, however, are comparable in terms of accuracy and linearity.

3.3. Doped or laboratory samples?

The two models passed the validation tests, so both were deemed effective for determining the active principle in the

Table 3
Results obtained in the validation of both calibration models for the determination of the active principle nimesulide

Aspect	Procedure		Laboratory model	Doped model
Linearity	NIR $= a + b \cdot$ REF		7 laboratory samples	10 dop. samples
		Concentration range	35–65 mg g^{-1}	35–65 mg g^{-1}
		b	1.02 ± 0.06	1.01 ± 0.09
		a	−2.18 ± 2.83	−0.75 ± 4.81
		Correlation	0.999	0.994
Accuracy	Paired t-test of NIR values and REF values of production batches		15 samples	15 samples
		Avg. diff.	−0.77	−0.37
		S.D.	3.54	1.63
		t_{exp}	0.84	0.88
		t_{crit}	2.14	2.14
Repeatability	Production samples analyzed six times by the same operator	NIR average	50.57	50.54
		S.D.	1.41	0.37
		R.S.D.	2.82%	0.72%
Intermediate precision	Production samples analyzed on 3 days by two operators	NIR average	52.65	51.50
		S.D.	2.16	0.68
		R.S.D.	4.10%	1.33%
		Two factors ANOVA	No significance effect of day or operator	No significance effect of day or operator
Robustness	Paired t-test of NIR values and REF values of production batches analyzed over a period of 1.5 year		30 samples	30 samples
		Avg. diff.	0.20	0.15
		S.D.	2.80	1.39
		t_{exp}	0.39	0.59
		t_{crit}	2.04	2.04

pharmaceutical preparation. However, the determination of laboratory and doped samples with the two PLS models revealed that each type of sample produced a different regression line. By way of example, Fig. 3 shows the results of the quantitation of doped samples with the two models; while the concentrations obtained with the model for doped samples were consistent with the reference values (slope and intercept were not significantly different from 1 and 0, respectively), those provided by the model for laboratory samples were not. When the laboratory samples are predicted with both models, the resulting graph is the opposed one to the Fig. 3. Now the prediction with the laboratory model shows results not significantly different from the reference values, while the prediction with the doped model shows a linear regression with a slope different from the unit and an intercept different from zero. This behaviour was maintained after the application of various spectral pretreatments (derivatives, SNV, DT).

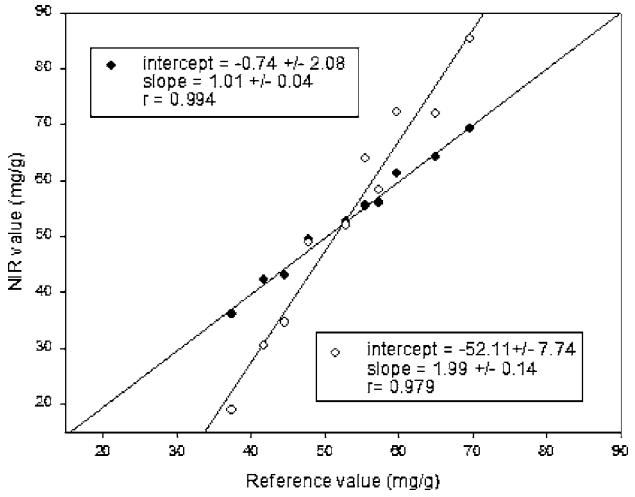

Fig. 3. Quantitation of doped samples using the two PLS models: doped model (●) and laboratory model (○).

In order to obtain a model capable of accurately predicting the concentrations of laboratory and doped samples, we tested a joint model involving both types of samples; this joint model is more complex (its require six factors) and exhibits a poorer predictive ability for production samples and the prediction quality was a function of the number of samples of each type used for calibration.

The obtainment of different regression lines depending on the origin of the sample is also commonplace in calibration transfers between similar instruments and could thus be approached similarly in theory. We thus constructed a calibration model with only one type of sample and then corrected with an small number of samples of the other type to correct the bias and slope of the calibration equation. Thus, the model was constructed from laboratory samples alone spanning the previous wavelength range (1100–2200 nm), but used only three factors; subsequently, bias and the slope were corrected by using four doped samples. The corrected model was found to accurately predict the concentrations of both the doped and the production

samples. A plot of the predicted values for 20 doped samples versus its reference values shows a slope of 1.05 ± 0.08 and an intercept of 1.90 ± 3.96 (correlation coefficient $= 0.989$). A paired *t*-test was used to compare the NIR and reference values for the production samples used to validate the robustness of the method; the test revealed the absence of significant differences ($t_{exp} = 1.91$, $t_{crit} = 2.08$, 95% confidence).

4. Conclusions

Different calibration sample sets have been tried to obtain suitable models of calibration for the prediction of production samples. The use of doped or laboratory samples is identically effective for constructing calibration methods allowing the accurate determination of the active principle in commercially available pharmaceutical preparations; both models are simple (four factors) and suitable for this purpose, although the doped samples model shows better statistical parameters. The validation of each developed method with these two models requires to use the same type as those employed to construct the model.

Another strategy is to construct calibration model using an unique type of sample and then adapt it to predict the other type correcting the bias and slope of the equation calibration; one obtains a model capable of accurately determining the values for samples of the other type without losing any predictive ability.

Acknowledgements

The authors are grateful to Spain's DGICyT for funding this research within the framework of Project BQU2000-0234. M.A. Romero acknowledges additional funding from Spain's Ministry of Education and Culture in the form of a researcher training grant. Finally, the authors wish to thank Laboratorios Menarini (Badalona, Spain) for kindly supplying the samples.

References

[1] C.I. Geräusser, K.A. Kovar, Appl. Spectrosc. 51 (10) (1997) 1504.
[2] W. Plugge, C. Van der Vlies, J. Pharm. Biomed. Anal. 11 (1993) 435.
[3] C. Coffey, B.E. Cooley Jr., D.S. Walker, Anal. Chim. Acta 395 (1999) 335.
[4] Z. Ge, B. Buchanan, J. Timmermans, D. DeTora, D. Ellison, J. Wyvratt, Process Control Qual. 11 (4) (1999) 277.
[5] S. Sekulic, J. Wakeman, P. Doherty, P.A. Hayley, J. Pharm. Biomed. Anal. 17 (1998) 1285.
[6] J.D. Kirsch, J.K. Drennen, J. Pharm. Biomed. Anal. 13 (1995) 1273.
[7] D. Trafford, R.D. Jee, A.C. Moffat, P. Graham, Analyst 124 (1999) 163.
[8] P. Corti, G. Ceramelli, E. Dreassi, S. Matii, Analyst 124 (1999) 755.
[9] I.R. Last, K.A. Prebble, J. Pharm. Biomed. Anal. 14 (1993) 1071.
[10] M. Blanco, A. Villar, Analyst 125 (2000) 2311.
[11] M. Blanco, J. Coello, H. Iturriaga, S. Maspoch, C. de la Pezuela, Anal. Chim. Acta 333 (1996) 147.
[12] ICH Q2B: Validation of Analytical Procedures: Methodology, Consensus Guideline, International Conference on Harmonisation (ICH), 1998.