ELSEVIER

Journal of Chromatography B, 708 (1998) 131–143

JOURNAL OF CHROMATOGRAPHY B

How to better define the characteristics of dispersion of results in liquid chromatographic analyses through an interlaboratory study Example of collaborative studies on ketoprofen and spiramycin

Jérôme Vial[a], Isabelle Ménier[a], Alain Jardy[a,*], Pascal Amger[b], André Brun[b], Laurent Burbaud[b]

[a]*Laboratoire Environnement et Chimie Analytique, Ecole Supérieure de Physique et Chimie Industrielles de la Ville de Paris, 10 Rue Vauquelin, 75005 Paris, France*

[b]*Centre de Recherche de Vitry Alfortville, Rhône-Poulenc Rorer, 13 Quai Jules Guesde, 94400 Vitry sur Seine, France*

Received 4 August 1997; received in revised form 11 December 1997; accepted 16 December 1997

Abstract

The aim of this study was to use statistical tools, especially the analysis of variance (ANOVA), to improve knowledge of the characteristics of the dispersion of results in high-performance liquid chromatography (HPLC) methods for quantitative analysis. It is in this regard that two interlaboratory studies have been carried out in collaboration with Rhône-Poulenc Rorer. The first concerned the analysis of a single drug product (ketoprofen) and was typically a ''simple analysis''. The second one involved a complex mixture of drug products and related substances (spiramycin), requiring far more constraining analysis conditions. Preliminary studies of the analyses were carried out to develop an optimized protocol. Statistical exploitation of the data for ketoprofen showed that there was no significant influence of the factors ''laboratory'' and ''preparation'', under the conditions of the study. On the other hand, in the case of spiramycin, a significant influence of the factors ''laboratory'' and ''preparation'' was observed under the conditions of the collaborative study, indicating that the latter factor must be taken into account to establish certified assays. Results of these two studies will help to determine the factors that have a significant influence, depending on the product and the chromatographic method used. By completing the statistical data base, interlaboratory studies will also contribute in the near future to the elaboration of more rigorous protocols for analytical transfers. © 1998 Elsevier Science B.V.

Keywords: Interlaboratory study; ANOVA; Ketoprofen; Spiramycin

1. Introduction

Liquid chromatography is a technique for quantitative analysis that is widely used in many industrial fields, especially the pharmaceutical industry. In spite of technical progress, which has enabled full automation of the whole analysis sequence, differences between repeatability and reproducibility (characteristic of the dispersion of the results obtained for the same analysis carried out in different laboratories, with different operators, on different equipment) were observed [1]. Good repeatability (characterized by a relative standard deviation that is often less than 1%) may lead to the apparently

*Corresponding author.

0378-4347/98/$19.00
PII S0378-4347(97)00668-3

paradoxical situation that two laboratories using the same chromatographic method for the analysis of the same product produce significantly different results. Thus, the problem of analytical transfer arises. The only solution that enables one to have a global overview of the problem for each analytical method consists of running a collaborative study involving at least six laboratories, which are expected to be representative of the population of laboratories that can carry out the analysis. Then, a statistical analysis of the data must be carried out (in collaboration with a statistician, if necessary). Such an approach, as previously recommended [2–4], not only gives the typical repeatability of the method, but can also determine whether factors like ''laboratory'' or ''preparation technique of solutions'' introduce significant dispersion in the results. It is possible to give an estimate of the dispersion generated by each factor that has a significant influence. In this article, two interlaboratory studies, carried out in collaboration with Rhône-Poulenc Rorer, are described and discussed. The first concerns the analysis of a product including a single drug product (ketoprofen) and is characteristic of a ''simple analysis'', while the second concerns a complex mixture of drug substances and related substances (spiramycin), requiring far more constraining analysis conditions. We tried to develop the methodology as much as the results themselves, which is an approach followed in various ways in articles dealing with collaborative studies [5–15]. Recently, precise recommendations dealing with the nomenclature of interlaboratory studies were published by IUPAC [16].

2. Experimental

For both collaborative studies, the products involved were well known, and quality control chromatographic methods had been previously optimized, validated and fully tested [17]. The protocol that was given to each laboratory included a detailed description of the method, preliminary tests to check that the system was able to run such analyses with acceptable results, methodology to prepare solutions, the injection sequence that was to be used (number of injections for standards and batches to analyze, order of injection), charts and floppy disks for storing the experimental results.

Each laboratory was free to use its own device. An exception was made for the choice of columns: the type was specified; moreover, we furnished the columns to avoid discrepancies for spiramycin. The laboratories sent us back raw data (peak areas and mass weighed), and we made the necessary calculations for the assays.

2.1. Ketoprofen collaborative study

2.1.1. Laboratories involved

What we call a laboratory is not necessarily a geographical unit, but may consist of a combination of an operator and a chromatographic instrument. Seven laboratories in the Rhône-Poulenc Rorer Company (denoted as one to seven) agreed to take part in this study. We consider that they constituted a representative sample of the laboratory population concerned with this analysis since they are involved in different steps of drug development.

2.1.2. Equipment

Table 1 lists the equipment used by each laboratory.

2.1.3. Chromatographic method

Analyses were carried out by reversed-phase partition chromatography under isocratic conditions. All of the laboratories were required to use a Nucleosil C_{18} 5 μm 150×4.6 mm column. Since it was commonly available, the choice of supplier was left to each participant.

The mobile phase was a mixture of acetonitrile, water and 0.5 mol l^{-1} phosphate buffer (60:38:2, v/v/v). The buffer pH was adjusted to 3.5 so that ketoprofen was in the molecular form. The chemicals could be obtained from any supplier but they had to be of high-performance liquid chromatography (HPLC) grade. The flow-rate was set at 1.5 ml/min. The injected volume was 5 μl. The detection wavelength was set at 254 nm and, if possible, the column temperature should be 25°C. Under these conditions, the analysis time was about 10 min. A typical chromatogram is shown in Fig. 1.

Table 1
Chromatographic equipment used for the ketoprofen study

Laboratory	Pump	Automatic injector	UV detector	Integrator
1	Gilson 305	Gilson 231	Spectromonitor 3200 LDC	Acquisition station P-E 7700+P-E LCI 100
2	Gilson 305	Waters 712	Spectromonitor 3200 LDC	Acquisition station Waters, Millenium 2.0.
3	Hewlett-Packard 1050	Waters 717 plus spectroflow 757	Kratos-LCI 100	P-E 7700+P-E
4	Gilson 305	Waters 717 plus	Spectromonitor 3100 LDC	Acquisition station P-E Turbochrom III
5	Gilson 305	Waters 717 plus	Spectromonitor 3200 LDC	P-E LCI 100
6	Shimadzu LC 9A	Spark Marathon	Spectromonitor 3200 LDC	P-E LCI 100
7	Gilson 305	Gilson 231	Spectromonitor 3100 LDC	P-E LCI 100

2.1.4. Preparation of solutions

Along with the detailed protocol, each laboratory received four vials containing 1 g of each product to analyze (these were selected from four different batches) and one vial containing 1 g of the reference standard. The preparation technique consisted of

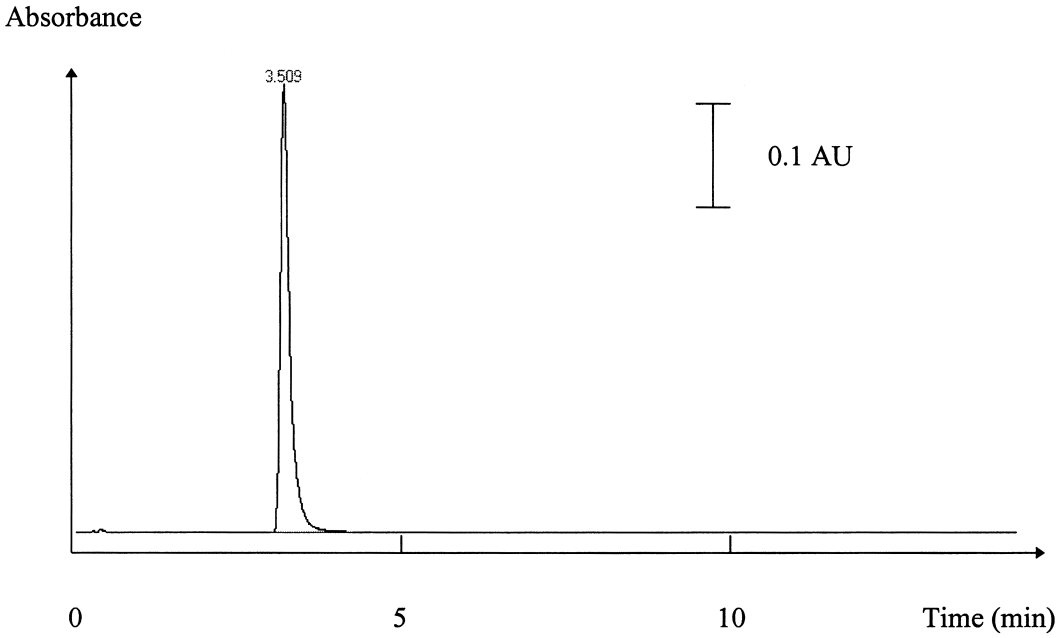

Fig. 1. Typical ketoprofen chromatogram.

accurately weighing about 50 mg of the product and diluting it in 100 ml of the mobile phase. For each product that was to be analyzed, two preparations per vial were made and each preparation was injected twice. For the reference standard, four preparations were made and they had to be injected fifteen times. Amber vials were recommended to avoid degradation of the product. A preliminary study showed that no significant degradation was observed with amber vials over a period of time covering the whole injection sequence.

2.1.5. Injection sequence

This sequence was designed to minimize the effects of a possible drift of the system. If we call R*i* the i^{th} preparation of the reference standard and B*jk* the k^{th} preparation of batch *j*, then the sequence can be written as follows:

R1/R1/R1/R1/R1/R1/R1/B11/B12/R2/B21/B22/R3/B31/B32/R4/B41/B42...

.../R1/B31/B32/R2/B41/B42/R3/B11/B12/R4/B21/B22/R1

2.1.6. Calculation of assay results

For this study, the choice of integration parameters was left to each laboratory, the only restriction being the compulsory use of areas for assay determination. As explained previously, each participant furnished us with raw data. First, the peak areas were normalized by the corresponding mass weighed. Then, for each injection of a product (four per batch), the result was calculated using the average of the two injections of the reference standard that were the nearest to each other in the injection sequence. Such a technique is called ''bracketing'' and is often used in industry.

2.2. The spiramycin collaborative study

This second interlaboratory study differed slightly from the first one. Differences were mainly in the injection sequence and in the method of calculating the results of the assays.

2.2.1. Laboratories involved

Seven laboratories in Rhône-Poulenc Rorer agreed to take part in this study. Again, we believe that they embodied a representative sample of the population of laboratories carrying out such analyses.

2.2.2. Equipment

Table 2 lists the equipment used by each laboratory.

2.2.3. Chromatographic method

The method employed isocratic elution reversed-phase chromatography. All of the laboratories used a

Table 2
Chromatographic equipment used for the spiramycin study

Laboratory	Pump	Automatic injector	UV detector	Integrator
1	Shimadzu, LC-6A	Spark Marathon 3200LDC	Spectromonitor D 2500	Hitachi, Merck,
2	Gilson 305	Gilson 231	Spectromonitor 3200 LDC	Acquisition station P-E Turbochrom IV
3	Waters 510	Waters 717 plus	Waters 486	Acquisition station Waters-Millenium
4	Milton Roy 3000 Constametric	Spark Marathon	Spectromonitor 3200 LDC	Hitachi, Merck, D 2500
5	Gilson 305	Waters 717	TSP LDC 3100	Acquisition station P-E Turbochrom V 4.0
6	Varian 9010	Kontron 465	Spectromonitor 3200 LDC	Acquisition station VG Multichrom
7	Varian 9012	Waters 715	Varian 2050	Shimadzu C-R4A

Nucleosil C_8 120 Å 3 μm 200×4.6 mm column. Preliminary studies showed a strong relation between the quality of the separation and the column manufacturer. To minimize discrepancies, the columns were furnished.

The mobile phase was a mixture of acetonitrile and phosphate buffer, pH 2.2 (30:70, v/v). The laboratories were not limited to any specific supplier. However, chemicals had to be of HPLC grade and have successfully passed the conformity test (no CN- in acetonitrile). The flow-rate was set at 0.8 ml/min. The injected volume was either 10 or 20 μl (depending on the available loop). The detection wavelength was set at 232 nm, and the column temperature had to be exactly 23°C. DNT (dinitro-3,4-toluene) was used as an internal standard. A solution of DNT was prepared by dissolving 750 mg of product in 500 ml of a mixture of acetonitrile–water (30:70, v/v). Under these conditions, the analysis took about 45 min. A typical chromatogram is shown in Fig. 2.

2.2.4. Preparation of solutions

Along with the detailed protocol, each laboratory received two vials containing 1 g of product to analyze (from two different batches) and one vial containing 1 g of the reference standard. The preparation technique consisted of accurately weighing about 50 mg of the product and diluting it in 20 ml of the internal standard solution, then adjusting the volume to 100 ml (for use with a 10-μl loop) or to 200 ml (for use with a 20-μl loop) with mobile phase. For each batch, three preparations were made and each preparation was injected twice. For the reference standard, three preparations were made and they had to be injected nine times. To avoid degradation of the product, solutions had to be kept at 4°C. A preliminary study showed that no significant degradation was observed under these conditions over a period covering the whole injection sequence.

2.2.5. Injection sequence

This sequence was designed so that the effects of a possible drift of the system were minimized. The use of three preparations enables us to detect a possible outlier. If we call Ri the i^{th} preparation of the reference and Bjk the k^{th} preparation of the batch j, then the sequence can be written as follows:

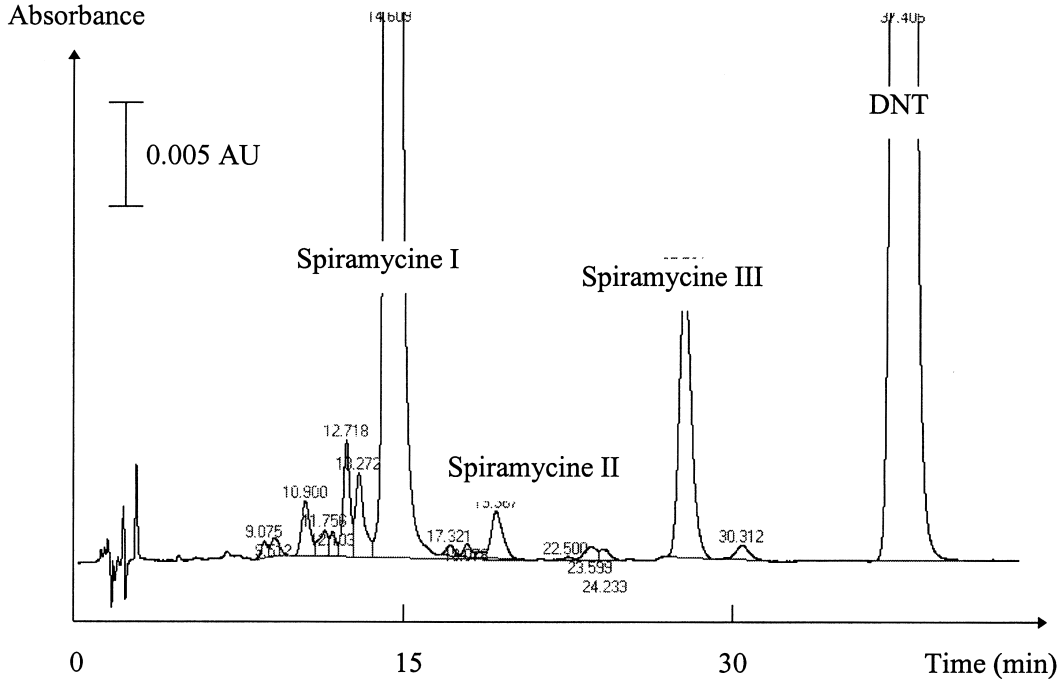

Fig. 2. Typical spiramycin chromatogram.

R1/R2/R3/B11/B12/B13/B11/B12/B13/R1/R2/R3...

.../B21/B22/B23/B21/B22/B23/R1/R2/R3

2.2.6. Calculation of assay results

For this study, the integration parameters were precisely defined in the protocol. As explained previously, each participant provided us with raw data. First, the peak areas were normalized by the corresponding mass weighed and the area of the DNT peak. Then, for each injection of product (six per batch), the results were calculated using the average of all of the injections of the reference standard. Such a technique follows the Food and Drug Administration's (FDA) recommendation for the calculation of assay results.

3. Statistical technique used for data evaluation

The general technique used for data evaluation is the analysis of variance (ANOVA). A theoretical basis for this technique is detailed in Ref. [18]. The aim is to determine if a controlled factor had a significant influence on the dispersion of the results by comparing mean squares through a statistical test.

The response used was the assay result (defined as explained in Sections 2.1.6 and 2.2.6). The data population must be Gaussian, which is a necessary condition for the ANOVA to be applicable. A Shapiro-Wilk's test showed that, for both studies, the normality assumption could not be rejected.

The theoretical statistical analysis was the same for both collaborative studies, the controlled factors were identical, with only the number of modalities being different.

3.1.1. Controlled factors

The controlled factors were all of the same type, i.e. random. This means that modalities considered for each factor in the collaborative study only stood for a very small sample of the whole population of modalities.

3.1.2. Homogeneity of variances

Before using an overall model for the ANOVA, including the influence of all the controlled factors, the homogeneity of variances must be checked. In practice, this means that, for all of the laboratories, the residual variances (characteristic of the laboratory's repeatability) and the ''preparation'' variances must be homogeneous. Such a criteria was tested using Cochran's test [18]. The estimates of the residual and ''preparation'' (if significant) variances per laboratory were obtained by an ANOVA carried out for each laboratory and using only the factors ''batch'' and ''preparation''. The significance test for the ''preparation'' factor consisted of comparing the ''preparation'' mean square of each laboratory to the pooled estimate (including the results from all of the laboratories) of the residual variance through a Fischer's test. In cases where the preparation factor had a significant influence, its estimated variance was computed using the pooled estimate of the residual variance.

3.1.3. Mathematical design

The calculation carried out with the ANOVA depended on the way the controlled factors were linked. Each laboratory analyzed the same batches as the others. Modalities for the factors ''laboratory'' and ''batch'' were independent and, for each factor, all of the modalities of the other factor were accessible. As a consequence, these two factors were crossed. The effect of a potential interaction could be investigated through the experimental results. The situation for the ''preparation'' factor was quite different. Each laboratory made its own preparations for each batch. The modalities of the ''preparation'' factor were dependent on the modalities of the factors, ''laboratory'' and ''batch''. Consequently, the ''preparation'' factor was nested with the ''laboratory'' and ''batch'' factors. Taking into account this information, the theoretical expression for the response is reported in Eq. (1):

$$x_{ijk\alpha} = m + a_i + b_j + ab_{ij} + c_{k(i,j)} + \epsilon_{ijk\alpha} \quad (1)$$

where $x_{ijk\alpha}$ stands for the result of a determination carried out in laboratory i on batch j with preparation k(i,j), m stands for the expected value of x (estimated by the grand mean), a_i stands for the effect of the ''laboratory'' factor at level i, b_j stands for the effect of the ''batch'' factor at level j, ab_{ij} stands for the effect of the interaction between the ''laboratory''

and the ''batch'' factors, respectively, at levels i and j, $c_{k(i,j)}$ stands for the effect of the ''preparation'' factor at level k, the ''laboratory'' factor being at level i and the ''batch'' factor being at level j, $\epsilon_{ijk\alpha}$ stands for the effect of the random error (residual error).

3.1.4. Theoretical ANOVA table

The theoretical ANOVA table corresponding to a mixed design, as was the case here, and taking into account the type of factors involved [18], is reported in Table 3.

3.1.5. Statistical tests

The influence of each controlled factor was tested with a Fisher–Snedecor's test (also called the *F*-test). This test consists of comparing two mean squares, which, in the case of a non-significant influence, are independent estimates of the same quantity, i.e. variance or a combination of variances. All of the *F*-tests used below were carried out with a significance level of 5% (which is the value that is generally used; [11,19–21]).

4. Results for the ketoprofen collaborative study

All of the calculations necessary for the ANOVA were achieved with the help of the Software MINITAB (Release 10) for Windows [22]. The *F*-tests used in our evaluation can, however, be different, since, in the software, the feedback regarding conclusions of previous tests is not taken into account, whereas they affect the discriminant functions to be used.

4.1. Outlier rejection

A Student's test resulted in the rejection of one preparation in laboratory five as an outlier. To avoid unbalancing the design, the concerned values were replaced by those of the other preparation of the same batch. No other outlier was detected.

4.2. Homogeneity of variances

The results of the ANOVA carried out in each laboratory to test the influence of the ''preparation'' factor and to get the estimates of the residual variances are reported in Table 4.

Table 3
Theoretical ANOVA table

Source of variation	Sum of squares (Q)	Degrees of freedom, ν	Quantity estimated by the mean square: $q=Q/\nu$
Laboratory factor	$Q_A = nrq\sum_{i=1}^{p}(\bar{x}_{i..}-\bar{x}_{...})^2$	$p-1$	$\sigma_r^2+n\sigma_C^2+nr\sigma_{AB}^2+nrq\sigma_A^2$
Batch factor	$Q_B = pnr\sum_{j=1}^{q}(\bar{x}_{.j.}-\bar{x}_{...})^2$	$q-1$	$\sigma_r^2+n\sigma_C^2+nr\sigma_{AB}^2+nrp\sigma_B^2$
Laboratory/batch interaction	$Q_{AB} = nr\sum_{i=1}^{p}\sum_{j=1}^{q}(\bar{x}_{ij.}-\bar{x}_{i..}-\bar{x}_{.j.}+\bar{x}_{...})^2$	$(p-1)\cdot(q-1)$	$\sigma_r^2+n\sigma_C^2+nr\sigma_{AB}^2$
Preparation (laboratory, batch) factor	$Q_C = n\sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{k=1}^{r}(\bar{x}_{ijk}-\bar{x}_{ij.})^2$	$pq(r-1)$	$\sigma_r^2+n\sigma_C^2$
Residual	$Q_r = \sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{k=1}^{r}\sum_{\alpha=1}^{n}(x_{ijk\alpha}-\bar{x}_{ijk})^2$	$pqr(n-1)$	σ_r^2
Total	$Q_T = \sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{k=1}^{r}\sum_{\alpha=1}^{n}(x_{ijk\alpha}-\bar{x}_{...})^2$	$pqrn-1$	σ_T^2

p = number of laboratories, q = number of batches, r = number of preparations for each batch, n = number of repetitions for each preparation, σ_r^2 = residual variance, σ_T^2 = total variance, σ_A^2 = laboratory variance, σ_B^2 = batch variance, σ_{AB}^2 = laboratory/batch interaction variance, σ_C^2 = preparation variance, and

$$\bar{x}_{ijk}=\frac{1}{n}\sum_{\alpha=1}^{n}x_{ijk\alpha},\quad \bar{x}_{ij.}=\frac{1}{nr}\sum_{k=1}^{r}\sum_{\alpha=1}^{n}x_{ijk\alpha},\quad \bar{x}_{i..}=\frac{1}{nrq}\sum_{j=1}^{q}\sum_{k=1}^{r}\sum_{\alpha=1}^{n}x_{ijk\alpha},\quad \bar{x}_{.j.}=\frac{1}{pnr}\sum_{i=1}^{p}\sum_{k=1}^{r}\sum_{\alpha=1}^{n}x_{ijk\alpha},\quad \bar{x}_{...}=\frac{1}{pnrq}\sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{k=1}^{r}\sum_{\alpha=1}^{n}x_{ijk\alpha}$$

Table 4
ANOVA results used for testing the homogeneity of variances

	q_{prep}	q_r	F_{exp}	Probability>F
Laboratory 1	1.97E−04	3.34E−04	1.57	19.5%
Laboratory 2	4.86E−05	6.52E−05	0.39	81.7%
Laboratory 3	2.94E−04	1.04E−04	2.34	6.6%
Laboratory 4	1.88E−05	2.89E−05	0.15	96.2%
Laboratory 5	6.68E−06	3.66E−05	0.05	99.5%
Laboratory 6	2.57E−04	1.93E−04	2.05	10.0%
Laboratory 7	1.45E−04	1.16E−04	1.16	34.0%
		pooled $\hat{\sigma}_r^2$=1.25E−04		

q_{prep} stands for the between-preparation mean square, obtained in the ANOVA per laboratory; q_r is the residual mean square, obtained in the ANOVA per laboratory (it is also the estimate of the residual variance σ_r^2 for each laboratory); F_{exp} is the ratio of q_{prep} divided by pooled $\hat{\sigma}_r^2$; Probability>F stands for the probability that a Snedecor's variable with four and 56 degrees of freedom leads to a value greater than the observed F value.

First, the homogeneity of the residual variances for the laboratories was checked. For this purpose, a Cochran's test was used [15,1,23]. The observed value for the discriminant function, 0.381, corresponded to a probability greater than the critical threshold of 1% (the value that is used most frequently for homogeneity of variances). As a consequence, the hypothesis of homogeneity of residual variances between laboratories could not be rejected.

If the type 1 error, α, was fixed at 5%, which is usual for F-tests, the hypothesis of non-influence of the ''preparation'' factor for all laboratories could not be rejected. More clearly, the ''preparation'' factor was not found to be significant in any laboratory. As a consequence, the homogeneity of the ''preparation'' variances was obtained de facto.

4.3. ANOVA table

Since the homogeneity of variances is acquired, it is possible to carry out the ANOVA. Results of the calculation are reported in Table 5. The significance level was chosen to be 5%. Each F-test took into account the result (significant or not) of the former one for the choice of the denominator. As the different estimates had sufficient degrees of freedom, no change in the model used for calculations was carried out following our conclusions. Indeed, possible pooling operations, which would have integrated the conclusions obtained previously, would not have given additional information.

With the significance level set at 5%, no controlled factor showed a significant influence on the response. This means that, under the conditions of the collaborative study, preparation of the solutions introduced no additional error (compared to the residual). No significant difference between batches could be detected, i.e., the process was under control. The assay did not depend upon the laboratory.

The standard deviation associated with the method can be extracted from Table 5. It corresponds to the square root of the pooled residual mean square. Numerically $\hat{\sigma}_r$=1.13%.

4.4. Certified assays

Another reason for performing a collaborative study is to produce certified materials. In the present work, certified materials were in fact the batches used during the collaborative study. The assay, obtained by averaging the results from all of the

Table 5
Ketoprofen collaborative study, ANOVA table

Source of variation	Q	ν	q	F	Probability>F
Laboratory factor	0.000792	6	0.000132	1.06	39.6%
Batch factor	0.001003	3	0.000334	2.69	5.5%
Laboratory/batch interaction	0.002141	18	0.000119	0.96	51.8%
Preparation factor (batch, laboratory)	0.003860	28	0.000138	1.11	36.2%
Residual	0.006957	56	0.00012		
Total	0.014753	111			

laboratories, was given with a confidence interval, the formula of which depended upon the factors that had a significant influence on the response. In the ketoprofen study, as no factor was found to be significant, the confidence interval is only related to the pooled residual variance. The 95% confidence interval is given by Eq. (2). Such an assay, expressed along with the confidence interval from the collaborative study, is called a certified assay.

$$\left[\tau_i - u_{95\%}\frac{\hat{\sigma}_r}{\sqrt{n \cdot p \cdot r}};\ \tau_i + u_{95\%}\frac{\hat{\sigma}_r}{\sqrt{n \cdot p \cdot r}}\right] \quad (2)$$

τ_i stands for the average assay for batch i, $u_{95\%}$ stands for the value that a normal variable has 95 chances out of 100 not to exceed in module and $\hat{\sigma}_r$ stands for the estimate of the residual standard deviation. Due to the high number of degrees of freedom, the standard deviation can be considered to be known in order to determine the confidence interval. This explains the use of the critical value of a standard normal variable rather than the critical value of a Student's variable.

Numerically, Eq. (2) becomes Eq. (3).

$$[\tau_i - 0.42\%;\ \tau_i + 0.42\%] \quad (3)$$

5. Results for the spiramycin collaborative study

For the sake of simplicity and in order to easily compare the results to those obtained for ketoprofen, the only response chosen in this article is the assay for spiramycin I (the main component). All of the calculations necessary for the ANOVA were achieved with the help of the Software JMP (Version 3.1) for Windows [24]. *F*-tests used in our exploitation may, however, differ, for the same reasons as those given previously. This software gave exactly the same results as MINITAB [22] (recalculations for the ketoprofen study were carried out and numerical results were identical). The software was used here because it was recommended by Rhône-Poulenc Rorer.

5.1. Outlier rejection

No outlier was detected.

5.2. Homogeneity of variances

The results of the ANOVA carried out in each laboratory to test the influence of the ''preparation'' factor and to get estimates of the residual variances are reported in Table 6.

First, the homogeneity of the residual variances was tested through a Cochran's test. The observed value for the discriminant function, 0.365, corresponded to a probability greater than the critical threshold of 1%. As a consequence, the hypothesis of homogeneity of residual variances between laboratories could not be rejected.

If the first type error, α, is fixed at 5%, which is usual for *F*-tests, the hypothesis that the ''preparation'' factor had no influence could be rejected for all laboratories, except three and four. With the exceptions of laboratories three and four, the repeatability was good enough to show a significant influence of the ''preparation'' factor on the assays.

Consequently, Cochran's test was also used to test the homogeneity of the ''preparation'' variances. The

Table 6
ANOVA results used for testing the homogeneity of variances

	q_{prep}	q_r	F_{exp}	Probability$>F$	$\hat{\sigma}_C$ per laboratory
Laboratory 1	0.87186	0.03944	5.96	0.07%	0.36
Laboratory 2	0.51971	0.11197	3.55	1.39%	0.19
Laboratory 3	0.01859	0.01095	0.13	97.18%	0
Laboratory 4	0.35368	0.37286	2.42	6.36%	0
Laboratory 5	2.93196	0.07058	20.04	0.00%	1.39
Laboratory 6	2.05287	0.31232	14.03	0.00%	0.95
Laboratory 7	1.85603	0.10606	12.69	0.00%	0.85
		pooled $\hat{\sigma}_r^2 = 0.146$			

observed value for the discriminant function, 0.371, corresponded to a probability greater than the critical threshold of 1%. As a consequence, the hypothesis of homogeneity of ''preparation'' variances between laboratories could not be rejected.

5.2.1. ANOVA table

Since the homogeneity of variances was acquired, it is possible to carry out the ANOVA. Results of the calculations are reported in Table 7. The significance level chosen was 5%. Each *F*-test took into account the result (significant or not) of the former one for the choice of the denominator. As the different estimates had sufficient degrees of freedom, no change in the model used for calculations was carried out following our conclusions. Indeed, possible pooling operations, which would have integrated the conclusions obtained previously, would not have provided additional information.

The α level was fixed at 5%. The typical repeatability of the method was characterised by a residual standard deviation of 0.39%. Under these conditions, a significant influence of the ''preparation'' factor was observed. It was even possible to give an estimate of the associated standard deviation, 0.73%. No significant influence of the interaction could be observed. No significant differences between batches could be detected, i.e., the process was under control. The laboratory factor had a significant influence on the response, e.g. the results obtained depended upon the laboratory. The associated standard deviation was estimated to be 0.53%.

5.3. Certified assays

The methodology used to define certified assays is similar to that used for the ketoprofen study. In the spiramycin study, factors laboratory and batch were found to be significant, the confidence interval must take into account their influences. The 95% confidence interval for certified batches is given by Eq. (4).

$$\left[\tau_i - t_{95\%}\sqrt{\frac{\hat{\sigma}_A^2}{p} + \frac{\hat{\sigma}_C^2}{p\cdot r} + \frac{\hat{\sigma}_r^2}{n\cdot p\cdot r}};\ \tau_i + t_{95\%}\sqrt{\frac{\hat{\sigma}_A^2}{p} + \frac{\hat{\sigma}_C^2}{p\cdot r} + \frac{\hat{\sigma}_r^2}{n\cdot p\cdot r}}\right] \quad (4)$$

τ_i stands for the average assay for batch i, $t_{95\%}$ stands for the value that a Student's variable with six degrees of freedom and has 95 chances out of 100 not to exceed in module, $\hat{\sigma}_r$ is the estimate of the residual standard deviation, $\hat{\sigma}_A$ stands for the estimate of the standard deviation for a given laboratory and $\hat{\sigma}_C$ stands for the estimate of the preparation standard deviation.

Numerically, Eq. (4) becomes Eq. (5).

$$[\tau_i - 0.64\%;\ \tau_i + 0.64\%] \quad (5)$$

6. Comparison of the two studies

6.1. Summary of the results for the two studies

In order to facilitate the comparison, the main characteristics are summarized in Table 8.

Table 7
Spiramycin collaborative study, ANOVA table

Source of variation	*Q*	ν	*q*	*F*	Probability$>F$
Laboratory factor	27.48	6	4.579	3.75	0.8%
Batch factor	1.66	1	1.663	1.35	25.46%
Laboratory/batch interaction	14.17	6	2.362	1.92	11.22%
Preparation factor (batch, laboratory)	34.42	28	1.229	8.4	$<$0.01%
Residual	6.145	42	0.146		
Total	83.88	83			

Table 8
Comparison of the characteristics of the two studies

Characteristic	Ketoprofen study	Spiramycin study
Calibration mode	External standard	External standard + internal standard
Assay calculation	Closest standards	All the standards
Injections per batch	4 per laboratory	6 per laboratory
Chromatogram	A single peak	Several peaks
Method ruggedness	quite rugged	Sensitive to many parameters but columns provided
Laboratories involved	7	7
Repeatability ($\hat{\sigma}_r$)	1.13%	0.39%
Preparation factor	Non significant	0.73%
Interaction	Non significant	Non significant
Batch factor	Non significant	Non significant
Laboratory factor	Non significant	0.53%
Certified assays	±0.42%	±0.64%
IC single laboratory[a]	±1.11%	±1.70%

[a] The IC single laboratory is the confidence interval for the results of a single laboratory (with the hypothesis that its characteristics are not different from those of collaborative study laboratories). Obtained from Eqs. (2) and (4) with $p=1$.

6.2. Discussion

First of all, it should be mentioned that the experimental magnitudes observed for the relative standard deviation are of the same order as those reported in ref. [4] and are even slightly better. Using Table 8, it is clear that the ketoprofen method was simple compared to the spiramycin one. It consequently required less attention from the operators. The first point of comparison, based on statistical exploitation, was the repeatability. Under this item are gathered all of the effects of the uncontrolled factors that correspond to the intrinsic performances of the chromatographic devices. At first sight, it was surprising that the repeatability was far better for the spiramycin method, yet less rugged. However, this could be explained by the fact that, given the ''simplissim'' aspect of the ketoprofen analysis, the choice for the column and the integration parameters was left free, which was not the case for the spiramycin assay. The improvement in the monitoring of the performance of the chromatographic devices could also be the reason for the differences observed in repeatability: More than one year separated the two studies. Moreover, as is shown in Fig. 1, the single peak for ketoprofen was affected by end-tailing, which may have had an influence on the dispersion of the results when the integration method was not imposed. It is also possible that the difference could come from the use of an internal standard, however, if it improves the repeatability, the change is not important enough to be significant. Consequently, the internal standard could only explain a small part of the observed differences [15]. The same is true for the consequences of the differences in the calculation mode of the assays: Theoretically, results could differ slightly, but the deviations observed experimentally were too great to be ascribed only to the calculation mode. The ''preparation'' factor only had an effect for the spiramycin method, however, this may be because the ketoprofen method repeatability was too bad to reveal a ''preparation'' factor of this magnitude. What is more, the spiramycin study led to a greater number of degrees of freedom for the ''preparation'' factor, which increased the power of statistical tests. For both methods, the ''batch'' factor was not significant, which means that it was not possible to show significant differences between the batches involved in the studies in any case. The ''laboratory'' factor had an effect only for the spiramycin method, but it might be for the same reasons as for the ''preparation'' factor. Globally, in the ketoprofen method, the larger value for the residual variance might hide all other possible effects. One can also note that, for the certified assays, the CI (confidence

interval) widths were not so different. However, when one laboratory uses the method alone, the difference becomes bigger. Multiplying the injections and the number of preparation would never reduce the CI below a certain threshold (see Fig. 3), since the expression of the CI includes (cf. Eq. (4)) a term that depends only upon the number of laboratories involved. To conclude, results obtained from the spiramycin study seem to be more reliable than those coming from the ketoprofen study. The ketoprofen collaborative study was carried out first and should be considered as a first attempt to define an efficient protocol for further collaborative studies.

6.3. Usefulness of such collaborative studies

Such studies appear necessary for many reasons. First, they enable analysts to define and estimate rigorously the dispersion characteristics of a chromatographic method that could be used in many different laboratories. With this in mind, collaborative studies should be integrated in method validation. Second, it is the only rigorous way to validate an analytical transfer to a new laboratory that wishes to use the method. Indeed, only collaborative studies can furnish certified materials and statistical information about the influence of the various factors involved. The new laboratory that wishes to validate itself for the method can compare its dispersion of results characteristics to those obtained in the collaborative study and, thanks to the certified materials, test its accuracy (e.g. if the results observed are within the limits predicted from the collaborative study).

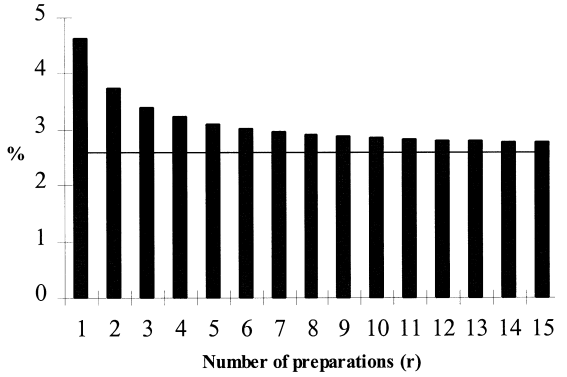

Fig. 3. Variation of the CI width for a single laboratory analyzing a batch of spiramycin as a function of the preparation number.

7. Conclusion

We have seen the different steps that are necessary within a collaborative study. The protocol must be as clear as possible, to reduce the possibility of misinterpretation and, consequently, the number of potential sources of variation. The use of ANOVA for data exploitation in collaborative studies is quite appropriate, but the greatest care must be taken of the model definition: it must correspond exactly to the actual situation. Not only the results for the two studies but also the methodology developed are of great interest, especially for better defining the characteristics of a method or further method transfer. Such studies must be followed by others on different products in order to improve our knowledge about chromatographic methods in general. The influence of the ''laboratory'' factor must also be regarded as a starting point for further investigations and experiments that will allow us to understand which are the factors (not controlled until now) that are hidden under this item.

Acknowledgements

We would like to thank the Rhône-Poulenc Rorer's laboratories of Chemical and Pharmaceutical Development that agreed to take part in these collaborative studies and to give us part of their precious time. They are Laboratories of Drug Safety Support in the Pharmaceutical Quality Analysis Department, Laboratories of the Analytical Chemistry Unit in the Process Chemistry Department and the Laboratory of Pharmaceutical Development in the Pharmaceutical Sciences Department.

References

[1] R.E. Pauls, R.W. McCoy, E.R. Ziegel, G.T. Fritz, D.M. Marmion, D.L. Krieger, J. Chromatogr. Sci. 26 (1988) 489.
[2] D.L. Massart, L. Perelgritz, Analusis 22 (1994) M13.

[3] E.A. Maier, Ph. Quevauviller, B. Griepink, Anal. Chim. Acta 283 (1993) 590.
[4] R.M. Venable, W. Horwitz, in I.W. Wainer (Editor), Liquid Chromatography in Pharmaceutical Development: An Introduction, Aster, Springfield, OR, 1985, pp. 399–408.
[5] A. Marchetto, R. Mosello, G.A. Tartari, H. Muntau, M. Bianchi, H. Geiss, G. Serrini, G.S. Lanza, J. Chromatogr. A 706 (1995) 13.
[6] P. Parvy, J. Bardet, D. Rabier, M. Gasquet, P. Kamoun, Clin. Chem. 39 (1993) 1831.
[7] J. Paesen, D.H. Calam, J.H.Mc.B. Miller, G. Raiola, A. Rozanski, B. Silver, J. Hoogmartens, J. Liq. Chromatogr. 16 (1993) 1529.
[8] K. Smets, E. Roets, J.McB. Miller, E. Porqueras, N. Berti, P. Vettori, F. Folliard, J. Hoogmartens, Chromatographia 37 (1993) 640.
[9] Ph. Quevauviller, D. Van Renterghem, B. Griepink, M. Valcarcel, M.D. Luque de Castro, J. Cosano, Anal. Chim. Acta 283 (1993) 600.
[10] H. Chung, M. Hsu, J. Chromatogr. 629 (1993) 277.
[11] L.M. Tsanaclis, J.F. Wilson, Clin. Chem. 39 (1993) 851.
[12] C. Hendrix, E. Roets, J. Crommen, J. De Beer, E. Porqueras, W. Van Den Bossche, J. Hoogmartens, J. Liq. Chromatogr. 16 (1993) 3321.
[13] R.W. McCoy, R.L. Aiken, R.E. Pauls, E.R. Ziegel, T. Wolf, G.T. Fritz, D.M. Marmion, J. Chromatogr. Sci. 22 (1984) 425.
[14] R.E. Pauls, R.W. McCoy, E.R. Ziegel, T. Wolf, G.T. Fritz, D.M. Marmion, J. Chromatogr. Sci. 24 (1986) 273.
[15] R.E. Pauls, R.W. McCoy, J. High Resolut. Chromatogr. 9 (1986) 600.
[16] W. Horwitz (Editor), Nomenclature of Interlaboratory Analytical Studies, Pure Appl. Chem., 66 (1994) 1903.
[17] French Pharmacopoeia, 10th ed., ADRAPHARM, Maisonneuve S.A., Paris, 1989.
[18] Commissariat à l'Energie Atomique, Statistique Appliquee à l'Exploitation des Mesures, I and II, Masson, Paris, 1978.
[19] Y. Vander Heyden, D.L. Massart, Y. Zhu, J. Hoogmartens, J. De Beer, J. Pharm. Biomed. Anal. 14 (1996) 1313.
[20] Y. Vander Heyden, K. Luypaert, C. Hartmann, D.L. Massart, J. Hoogmartens, J. De Beer, Anal. Chim. Acta 312 (1995) 245.
[21] Y. Vander Heyden, C. Hartmann, D.L. Massart, L. Michel, P. Kiechle, F. Erni, Anal. Chim. Acta 316 (1995) 15.
[22] Minitab Reference Manual and User's guide (Release 10 for Windows), Minitab Inc., State College, PA, USA, 1994.
[23] J. Bastin, M. Bomans, F. Dugain, C. Michaud, J.M. Pujade-Renaud, J. Van Audenhove, Analusis 9 (1981) 417.
[24] JMP Statistic and Graphics Guide and User's Guide (Version 3.1), SAS Institute, Inc., Cary, NC, USA, 1995.