ELSEVIER

Journal of Chromatography A, 830 (1999) 3–28

JOURNAL OF CHROMATOGRAPHY A

Tutorial

Interlaboratory studies on two high-performance liquid chromatographic assays for tylosin (tartrate)

Y. Vander Heyden[a], J. Saevels[b], E. Roets[b], J. Hoogmartens[b], D. Decolin[c], M.G. Quaglia[d], W. Van den Bossche[e], R. Leemans[f], O. Smeets[g], F. Van de Vaart[g], B. Mason[h], G.C. Taylor[h], W. Underberg[i], A. Bult[i], P. Chiap[j], J. Crommen[j], J. De Beer[k], S.H. Hansen[l], D.L. Massart[a,*]

[a]*ChemoAC, Pharmaceutical Institute, Vrije Universiteit Brussel, Laarbeeklaan 103, B-1090 Brussels, Belgium*
[b]*Katholieke Universiteit Leuven, Leuven, Belgium*
[c]*Université de Nancy, Nancy, France*
[d]*Universita La Sapienza, Rome, Italy*
[e]*Universiteit van Gent, Ghent, Belgium*
[f]*Dopharma, Raamsdonkveer, The Netherlands*
[g]*Laboratorium der Nederlandse Apothekers, The Hague, The Netherlands*
[h]*Dista Products Ltd., Liverpool, UK*
[i]*Universiteit Utrecht, Utrecht, The Netherlands*
[j]*Université de Liège, Liege, Belgium*
[k]*Wetenschappelijk Instituut Volksgezondheid – Louis Pasteur, Brussels, Belgium*
[l]*Royal Danish School of Pharmacy, Copenhagen, Denmark*

Received 3 February 1998; received in revised form 28 July 1998; accepted 13 October 1998

Abstract

An interlaboratory study was performed on two high-performance liquid chromatographic methods to determine tylosin. The first method is a reversed-phase HPLC on a C_{18} column, while the second is a method using a polymeric stationary phase. The first method is described in several pharmacopoeia monographs on tylosin, to determine the composition of a tylosin mixture, while the second method is recently proposed to determine both the composition and the contents in such a mixture. The interlaboratory studies were set-up and interpreted according to ISO guidelines. This paper is written as a tutorial type of article explaining the principles and methods of these guidelines. The results of both methods were compared. Both were found to have disadvantages but in general the old method is still preferred, both for composition determination and to assay the components.

Keywords: Tutorial; Interlaboratory study; ISO guidelines; Tylosin

1. Introduction

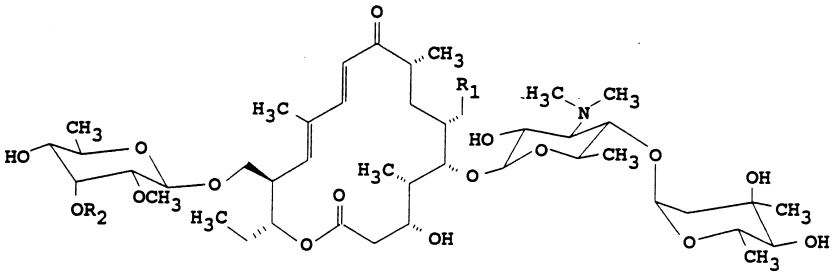
Tylosin is a mixture of macrolide antibiotics produced by fermentation of *Streptomyces* species. Tylosin A (TA) is the main substance in this mixture. During fermentation several related components can be formed among which tylosin B (desmycosin, TB), tylosin C (macrocin, TC), tylosin D (relomycin, TD) and demycinosyltylosin (DMT)

*Corresponding author. Tel.: +32-2-4774-737, fax: +32-2-4774-735, e-mail: fabi@vub.vub.ac.be

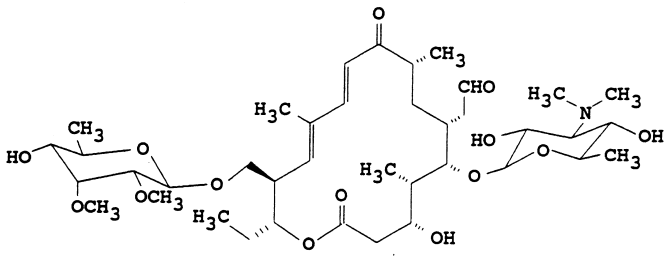
0021-9673/99/$ – see front matter
PII: S0021-9673(98)00840-1

for instance [1]. Their structures are given in Fig. 1. Tylosin is used in veterinary practice as a feed additive and as a therapeutic substance in the treatment of mycoplasmosis in poultry and livestock.

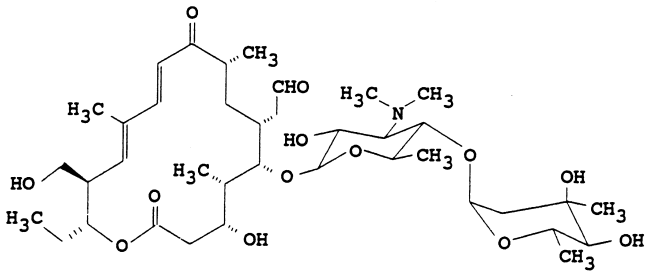
In a new proposal of a monograph for tylosin tartrate [2] two high-performance liquid chromatography (HPLC) methods are included. One method (method II), described in Ref. [3], is new and is



	R_1	R_2
Tylosin A	CHO	CH_3
Tylosin D	CH_2OH	CH_3
Tylosin C	CHO	H



Tylosin B



Demycinosyltylosin

Fig. 1. Structures of the different tylosins.

meant to assay the contents of tylosins A, B, C, D and other related or unknown substances. The other method (method I) is already given in several pharmacopoeias [4–6], and is now included in the annex of the new monograph of Ref. [2]. It is used to verify the composition of the tylosin mixture, and is followed by a microbiological assay.

The question we would like to answer is which of the two methods is most suitable to quantitatively determine the composition of the tylosin samples.

An interlaboratory study was conducted for both methods. Twelve laboratories participated in the study. The studies were organized and the protocols were set up according to the ISO 5725-2 standard to determine the repeatability and reproducibility of a measurement method [7]. Because of the column dependent selectivity reported for method I [1], the resolution between pairs of a number of solutes are also data to be considered.

At first, a training round was organised to allow the laboratories to become acquainted with the method and to verify if the protocols describing the experimental set-ups did not contain errors or ambiguities. The results of the final study were treated as described in the ISO 5725-2 guidelines.

This paper is written as a tutorial type of article describing the principles and methods used in the ISO 5725-2 guidelines. Interlaboratory studies are namely an important method validation topic with a potential interest for chemists working with analytical methods. The general principles of the ISO 5725-2 guidelines are given and applied on the case studies for tylosin. A more detailed discussion about the practical set-up and analysis of an interlaboratory study can be found in the guidelines.

2. Experimental

2.1. Solvents and substances

All samples and standards were provided by the laboratory of Professor Hoogmartens (Katholieke Universiteit Leuven, Leuven, Belgium) who performed the duties of the executive officer [7].

2.1.1. Method I

Acetonitrile was of liquid chromatography grade. In the preparation of solutions, distilled, deionized or Milli-Q water was used. To adjust the pH of the sodium perchlorate solution in the mobile phase, 1 *M* hydrochloric acid was prepared by diluting the concentrated acid (35.0–39.0%, m/m HCl) 10 times with water.

2.1.2. Method II

Tetrahydrofuran (THF) was required to be of liquid chromatography grade. In the preparation of solutions distilled, deionized or Milli-Q water was used. The phosphate buffer solution, pH 7.0 consists of 250.0 ml 0.2 *M* potassium dihydrogenphosphate solution and of 148.2 ml 0.8% (m/v) sodium hydroxide solution, per litre of buffer. The solution(s) to adjust the pH of the buffer are not specified. Dilute phosphoric acid contains 9.5–10.5% (m/m) H_3PO_4 and was prepared by mixing 115 g concentrated phosphoric acid and 885 g water. It is used to adjust the pH of the dipotassium hydrogenphosphate solution in the mobile phase.

2.2. Test and reference solutions

The following test and reference solutions were defined, (1) test solution containing 1.0 mg/ml sample in solvent, (2) low level test solution and high level test solution containing 0.8 and 1.1 mg/ml sample in solvent, (3) reference solution (a) containing 1.0 mg/ml tylosin A CRS in solvent (=100% solution), (4) reference solution (b) which is reference solution (a) diluted 25 times with solvent (=4.0% solution), (5) reference solution (c) which is reference solution (a) diluted 400 times with solvent (=0.25% solution), (6) reference solution (de) which contains 140 mg l^{-1} tylosin A CRS, 140 mg l^{-1} tylosin D CRS and 100 mg l^{-1} tylosin B CRS in solvent, (7) reference solution (f) which contains 140 mg l^{-1} tylosin C CRS and 100 mg l^{-1} of demycinosyltylosine (DMT) in solvent, (8) blank solution (=solvent) which was for method I a mixture of acetonitrile–water (1:1, v/v) and for method II a phosphate buffer solution, pH 7.0.

2.3. Chromatographic procedures

2.3.1. Method I

A stainless steel column (250×4.6 mm I.D.) packed with octadecylsilyl silica gel (5 μm) was

used under the following conditions: mobile phase: acetonitrile–sodium perchlorate solution, 22.5% (m/v) pH 2.5 (40:60, v/v); flow-rate 1.0 ml/min; UV detection at 290 nm; 20 μl loop; column temperature 35°C. The columns used in the different laboratories are shown in Table 1.

2.3.2. Method II

A column (250×4.6 mm I.D.) packed with styrene–divinylbenzene copolymer (8 to 10 μm particles with a pore size of 100 nm) was used under the following conditions: mobile phase: THF–di-potassium hydrogenphosphate solution 3.5% (m/v) pH 9.0–water (200:50:750, v/v/v); flow-rate 1.0 ml/min; UV detection at 280 nm; 20 μl loop; column temperature 60°C. The columns used in the different laboratories are shown in Table 1.

3. Theory

3.1. Generalities concerning the set-up and analysis of interlaboratory studies

The ISO 5725-2 standard provides guidance to design, perform and analyse a collaborative inter-laboratory experiment to estimate the precision of measurement methods which yield quantitative results on a continuous scale. The procedure allows to determine repeatability and reproducibility standard

Table 1
Columns used in the different laboratories, (a) for method I and (b) for method II

Laboratory No.	Columns used
(a) Method I	
1	Hypersil ODS 5 μm, 250×4.6 mm, Alltech
2	Hypersil ODS 5 μm, 250×4.6 mm, Alltech
3	Nucleosil C_{18} 5 μm, 250×4 mm, Macherey Nagel
5	Spherisorb S5 ODS1, 5 μm, Waters
6	Nucleosil C_{18} 100A-5 μm, 250×4.6 mm, Chrompack
7	Nucleosil ODS 100A-5 μm, 250×4.6 mm, Phase Separations
8	Nucleosil C_{18} 100A-5 μm, 250×4.6 mm, Chrompack (pre-column: pellicular RP 10×3 mm Chrompack)
9	Nucleosil C_{18} 5 μm, 250×4 mm, HiChrom
10	Nucleosil C_{18} 100A-5 μm, 250×4.6 mm, Phenomenex
11	LiChrospher 100 RP-18 endcapped 5 μm, 250×4.6 mm, Merck
12	Ultrasphere ODS, 5 μm, 250×4.6 mm, Beckman
13	ODS-AQ, S-5 μm 120A, 250×4.6 mm, YMC-Pack
(b) Method II	Column: PLRP-S 1000A 8 μm, 250×4.6 mm Origin:
1	Home packed
2	Polymer Laboratories
3	Polymer Laboratories
5	Polymer Laboratories
6	Polymer Laboratories
7	Polymer Laboratories
8	Rochrom
9	Polymer Laboratories
10	Polymer Laboratories
11	Home packed
12	Home packed in Laboratory 1
13	Home packed in Laboratory 1

In the end laboratory No. 4 did not participate in methods I or II.

deviations as precision estimates. The guidelines are meant to be applied on measurement methods that have been standardised and that in general are in use in a number of laboratories (standard measurement method).

In the set-up, the execution and the analysis of an interlaboratory study different types of personnel are defined, such as the panel, the executive officer, the statistical officer, the supervisors and the operators. The panel consists of experts familiar with the method and its application and it decides on the planning and co-ordination of the study. The executive officer is responsible for the actual organisation of the experiment while the statistical officer analyses the data and reports about the results. The supervisors are staff members in each of the participating laboratories that are responsible for organising and supervising the actual performance of the measurements and for reporting the test results. They do not take part in performing the experiments. Their tasks consist among those already mentioned in (a) ensuring that the operator selected is one of those who would normally carry out such measurements in routine operations, (b) collecting the test results recorded to an agreed number of decimal places, including any anomalies and difficulties experienced, and comments made by the operators. The supervisor of each laboratory should write a full report. To ensure uniformity in reporting it can be recommended that the panel creates a standardised report form.

The operators perform the measurements according to the operating procedure of the method. In each laboratory the measurements are carried out by only one operator. Because the object of the experiment is to determine the precision obtainable by the general population of operators working from the standard measurement method, in general the operators should not be given amplifications to the protocol for the measurement method. However, it should be pointed out to the operators that the purpose of the exercise is to discover the extent to which results can vary in practice. Therefore they should not discard or rework results that they feel to be inconsistent. It is better to report a mistake than to adjust the test results, because one or two missing test results will not spoil the experiment and many indicate a deficiency in the standard (protocol). They are encouraged to comment on the protocol and, in particular, to state whether the instructions contained in it are sufficiently unambiguous and clear.

Some preliminary experiments (training round) can be organised for familiarisation with the measurement method and to evaluate the protocol for errors or ambiguities.

3.2. Layout of the interlaboratory study

The layout of the interlaboratory study is called a balanced uniform-level design. In this design q samples representing q different test levels are analysed in p different laboratories which each measure exactly n replicate test results under repeatability conditions for each of the q samples (see Table 2). A laboratory is defined as a combination of the operator, the equipment and the test site. One test site can deliver several laboratories if it can provide several operators with independent sets of equipment and working situations.

Any preliminary checking of the equipment shall be as specified in the operating procedure. The n tests under repeatability conditions are performed independently as if it were different samples. Preferably the samples are delivered in such a way the operator is not aware which ones are replicates of each other. However this approach is only possible if all $q \cdot n$ measurements could be done within a short interval of time (under repeatability conditions).

Table 2
Balanced uniform-level design for p laboratories, q samples (levels) and n replicates per sample

Laboratory	Sample (level)								
	1	2	..	..	j	..	..	$q-1$	q
1									
2									
..									
..					y_{ij1}				
..					..				
i					y_{ijk}				
..					..				
..					y_{ijn}				
..									
p									

3.3. Statistical analysis of the results

The analysis of the data involves three successive stages, (1) the critical examination of the data to identify outliers, (2) the calculation of the means, *m*, and of preliminary precision estimates for each level (sample) separately, and (3) the establishment of final precision estimates after determining relationships between precision and the means, *m*, when the data indicate that such a relationship may exist. Item (2) above involves the estimation for each level of the repeatability variance (s_{rj}^2), the between-laboratory variance (s_{Lj}^2) and the reproducibility variance (s_{Rj}^2).

3.3.1. Scrutiny of results for consistency and outliers

The presence of individual laboratories or values that appear to be inconsistent with the other laboratories or values, are detected by two approaches, namely (i) graphical consistency techniques and (ii) numerical outlier tests. Two types of consistency of data are evaluated namely those of the within-laboratory variation and of the between-laboratory variation.

In the examination of the within-laboratory consistency it is studied if the variance of the results for a sample, obtained within a laboratory is not considerably larger than in the other participating laboratories. This is done graphically by means of plotting the Mandel's *k* statistic (see Fig. 2) while numerically the Cochran's test is applied. The Mandel's *k* statistic is calculated as

$$k_{ij} = \frac{s_{ij}\sqrt{p_j}}{\sqrt{\sum s_{ij}^2}} \quad (1)$$

In this statistic the standard deviation within one cell (s_{ij}), i.e., from the *n* results for one sample within one laboratory, is compared to the pooled standard deviation obtained for that sample in the different laboratories. The different k_{ij} values are plotted. They can be grouped per laboratory (Fig. 2) or per sample. On these plots lines can be drawn corresponding to indicator values for the Mandel's *k* statistic at different significance levels (see Fig. 2). These values can be found in tables created by Mandel and published in the ISO 5725-2 guidelines [7]. These plots are used to evaluate if certain laboratories exhibit patterns that are markedly different from those from the others. If for instance one laboratory stands out, it indicates that it has a poorer

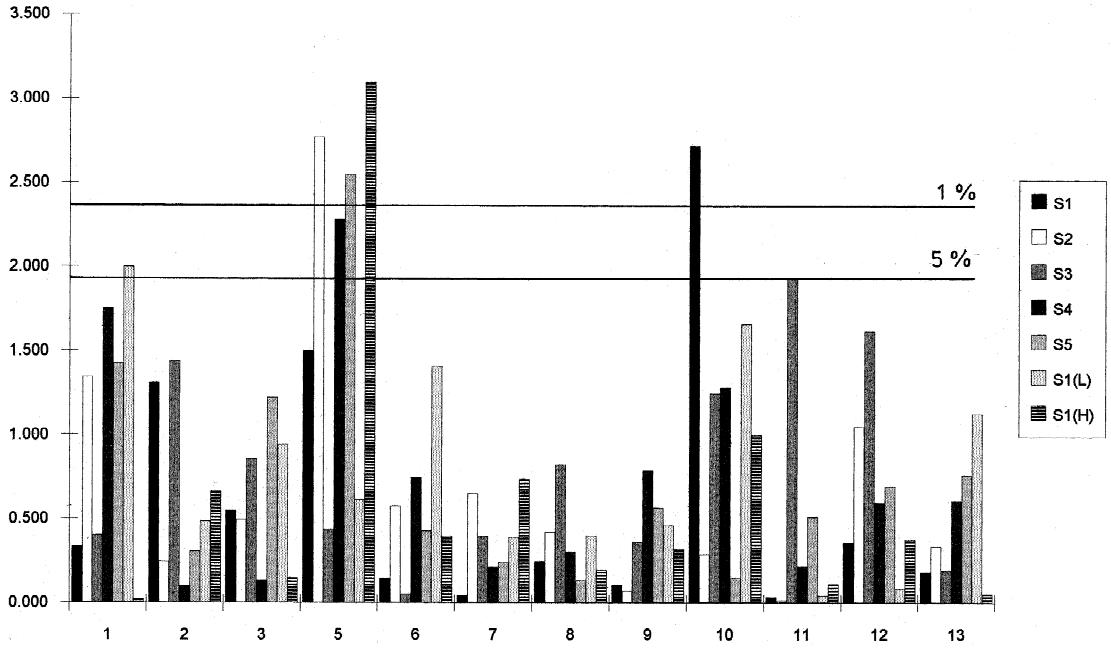


Fig. 2. Mandel's *k* statistics plotted for the results of tylosin A obtained with method I and grouped per laboratory.

repeatability than the other laboratories. If on the other hand a laboratory gives rise to consistently small k values, it could indicate that either an insensitive measurement scale was used or that an excessive rounding in the reported data was applied. To avoid these latter kind of situations, also in the calculation of other statistics, the number of required digits in the report should be indicated in the protocol of the study and the statistical officer should apply for intermediate calculation results a number of digits that is at least the double of this of the original data.

In the Cochran test, the test statistic, C, for a given set of p standard deviations s_i all computed from the same number (n) of replicate results, is calculated as

$$C = \frac{s_{\max}^2}{\sum_{i=1}^{p} s_i^2} \tag{2}$$

where $s_{\max}$ is the highest standard deviation estimated for one sample. A standard deviation (variance) is considered to be an outlier when C is larger than the 1% critical value and is called a straggler when it is smaller than the 1% value but larger than the 5% one. A table with the critical values can be found in Refs. [7,8]. If the highest standard deviation is classified as an outlier this value is omitted and the Cochran test is repeated on the remaining data set. The procedure can be repeated several times but could lead to an excessive number of rejections when the underlying assumption of normality is not fulfilled and therefore conclusions from such a repetitive procedure have to be drawn with great caution. When one laboratory shows several outliers and/or stragglers for the different samples analysed, this could indicate that the repeatability within this laboratory is unacceptably high.

In the examination of the between-laboratory consistency it is studied if there are laboratories with deviating results compared to those of the other laboratories. This is done graphically by means of plotting the Mandel's h statistic (see Fig. 3) while numerically Grubbs' test is applied. The Mandel's h statistic is calculated as

$$h_{ij} = \frac{\bar{y}_{ij} - \bar{\bar{y}}_j}{\sqrt{\frac{1}{p_j - 1}\sum_{i=1}^{p_j}\left(\bar{y}_{ij} - \bar{\bar{y}}_j\right)^2}} \tag{3}$$

in which $\bar{y}_{ij}$ represents a cell mean and $\bar{\bar{y}}_j$ the

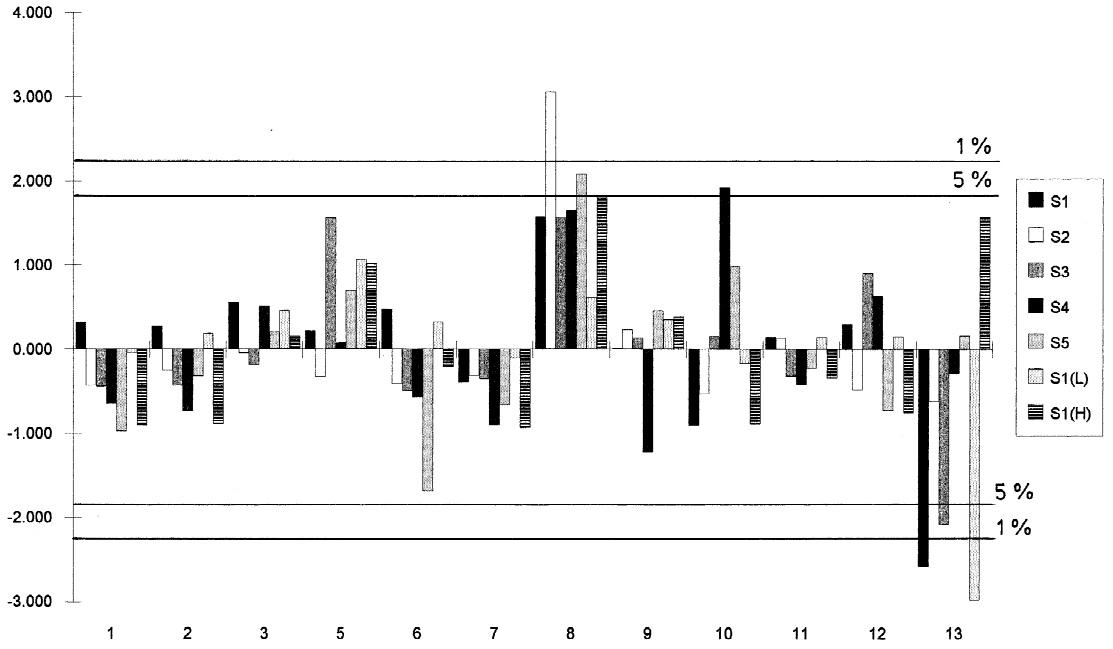


Fig. 3. Mandel's h statistics plotted for the results of tylosin A obtained with method I and grouped per laboratory.

general mean for sample j. The denominator represents the standard deviation of the cell means for one sample. As for the k_{ij} values, the plots for the h_{ij} values can be grouped per laboratory (Fig. 3) or per sample. On these plots lines can be drawn corresponding to indicator values for the Mandel's h statistic at different significance levels (see Fig. 3). These values can be found in tables created by Mandel and published in the ISO 5725-2 guidelines [7]. These plots are used to evaluate if some laboratories have consistently high or low cell means compared to the average results in the different laboratories.

The Grubb's test is an outlier test used in this statistical analysis to evaluate the between-laboratory variability. It consists of four different subtests. In a first instance the occurrence of one outlying observation, either the smallest or the largest, is evaluated. Secondly the data set is examined for two outlying observations both situated at either of the extreme values of the data set. It also has to be remarked that in the calculation of the Grubbs' statistics only the data retained after the within-laboratory consistency evaluation are used.

To determine if the largest observation ($\bar{y}_p$ after sorting the data in ascending order) is an outlier, the test statistic used is

$$G_p = \frac{\bar{y}_p - \bar{\bar{y}}_j}{\sqrt{\frac{1}{p_j - 1}\sum_{i=1}^{p_j}\left(\bar{y}_{ij} - \bar{\bar{y}}_j\right)^2}} \quad (4)$$

To test the significance of the smallest observation ($\bar{y}_1$) the following statistic is computed

$$G_1 = \frac{\bar{\bar{y}}_j - \bar{y}_1}{\sqrt{\frac{1}{p_j - 1}\sum_{i=1}^{p_j}\left(\bar{y}_{ij} - \bar{\bar{y}}_j\right)^2}} \quad (5)$$

The tested value, $\bar{y}_p$ or $\bar{y}_1$, is considered to be an outlier when its statistic is larger than the 1% critical value and is called a straggler when it is smaller than the 1% value but larger than the 5% one. A table with the critical values can again be found in Refs. [7,8].

To examine whether the two largest observations are outliers calculate the Grubbs' test statistic $G_{p-1,p}$

$$G_{p-1,p} = \frac{\sum_{i=1}^{p-2}\left(\bar{y}_{ij} - \bar{\bar{y}}_{p-1,p}\right)^2}{\sum_{i=1}^{p_j}\left(\bar{y}_{ij} - \bar{\bar{y}}_j\right)^2} \quad (6)$$

where $\bar{\bar{y}}_{p-1,p}$ is the average of the $p-2$ smallest cell means in the data set. Alternatively, to test the two smallest observations the Grubbs' test statistic $G_{1,2}$ is computed

$$G_{1,2} = \frac{\sum_{i=3}^{p}\left(\bar{y}_{ij} - \bar{\bar{y}}_{1,2}\right)^2}{\sum_{i=1}^{p_j}\left(\bar{y}_{ij} - \bar{\bar{y}}_j\right)^2} \quad (7)$$

where $\bar{\bar{y}}_{1,2}$ is the average of the $p-2$ largest cell means in the data set. In Eqs. (6) and (7) one is dividing a smaller sum of squares by a larger one. Therefore the rules to decide on significance are different from those previously applied. For the Grubbs' tests for two outlying observations, outliers and stragglers give rise to test statistic values which are smaller than the tabulated 1% and 5% critical values, respectively [7,8].

The different Grubbs' tests are applied as follows. First a test for one outlying observation is executed. When that cell mean is found to be an outlier, it is excluded and the test at the other extreme levels is done, e.g., if the lowest result is an outlier then evaluate the highest with the lowest one excluded. When a value is found to be an outlier, the test for two outlying observations is no longer applied.

It can be remarked that a Grubb's test also could be applied to indicate and eliminate outliers in the results within one cell when more than two replicated results per cell are measured and when the previously performed Cochran test indicated the cell standard deviation as being suspect.

3.3.2. Calculation of the general mean and variances

In this part of the analysis an estimation for m, the real content in a sample, is made as well as for the precision at this sample level. The estimated mean content for a sample is symbolised as $\hat{m}_j$ and is equal to the $\bar{\bar{y}}_j$ value obtained from the dataset after elimination of the outliers for the sample considered.

Three variances are calculated for each sample. They are the repeatability variance, the between-laboratory variance and the reproducibility variance. The repeatability variance is the pooled variance of the within-cell variances

$$s_{rj}^2 = \frac{\sum_{i=1}^{p}(n_{ij}-1)s_{ij}^2}{\sum_{i=1}^{p}(n_{ij}-1)} \quad (8)$$

The between-laboratory variance is given by

$$s_{Lj}^2 = \frac{s_{dj}^2 - s_{rj}^2}{\bar{\bar{n}}_j} \quad (9)$$

where

$$s_{dj}^2 = \frac{1}{p-1}\sum_{i=1}^{p} n_{ij}\left(\bar{y}_{ij} - \bar{\bar{y}}_j\right)^2 \quad (10)$$

and

$$\bar{\bar{n}}_j = \frac{1}{p-1}\left(\sum_{i=1}^{p} n_{ij} - \frac{\sum_{i=1}^{p} n_{ij}^2}{\sum_{i=1}^{p} n_{ij}}\right) \quad (11)$$

For the particular case where all $n_{ij}=n=2$, which was the case in this study, simpler formulas can be used both for the repeatability and the between-laboratory variances

$$s_{rj}^2 = \frac{1}{2p}\sum_{i=1}^{p}(y_{ij1}-y_{ij2})^2 \quad (12)$$

and

$$s_{Lj}^2 = \frac{1}{p-1}\sum_{i=1}^{p}\left(\bar{y}_{ij} - \bar{\bar{y}}_j\right)^2 - \frac{s_{rj}^2}{2} \quad (13)$$

When due to random effects a negative value for s_{Lj}^2 is obtained, this value is assumed to be zero. The reproducibility variance is defined as

$$s_{Rj}^2 = s_{rj}^2 + s_{Lj}^2 \quad (14)$$

3.3.3. Establishing a functional relationship between precision values and the mean contents

Subsequently, it is investigated whether the precision depends upon m and, if so, the functional relationship is determined. The different approaches applied are discussed in more detail in Section 4.

4. Results and discussion

4.1. About the standard measurement method and the experimental set-up

Method I was slightly adapted in the sense that the operator was allowed to change the ratio acetonitrile–aqueous phase in the mobile phase when the resolution between tylosin A and tylosin D was below the limit value required in Refs. [2,4–6], i.e., below 2.0 (see further). A similar change, in the THF–aqueous phase ratio, is also allowed and described in Ref. [2] for method II, when the resolution TD–TA is below 4.0. The concentrations of samples and standards injected were the same for both methods, only differing in the solvent used. The concentrations described for method II in Ref. [2] were taken. They are described in more detail in Section 2.1. The concentrations and amounts used did not overload any column to be used and allow to observe by-products more clearly.

During the training round, the participating laboratories received one sample and the standards of tylosin A, B and D. This allowed one, for instance to conclude that, (i) in method II one should not use THF GR grade (pro analysis) because of precipitation problems with the antioxidant 2,6-di-*tert*.-butyl-4-methylphenol (BHT), in the mobile phase, (ii) THF should be free of peroxides because they can affect the peak shapes, therefore relatively old THF should not be used, (iii) for all solutions storage in the dark is more important than storage in the cold at 5°C as required in Ref. [2].

In the final study each laboratory ($p=12$) had to analyse in duplicate ($n=2$) and under repeatable conditions five bulk samples (labelled S1 to S5) by both methods. In addition, for one sample (S1) a high and a low concentration sample was simulated by weighing 110% and 80% of the nominal amount and considering the results as originating from a normal (100%) sample ($q=7$). In this way the applicability of the method for different sample levels (low and high tylosin A content) is investigated even when all five bulk samples have similar

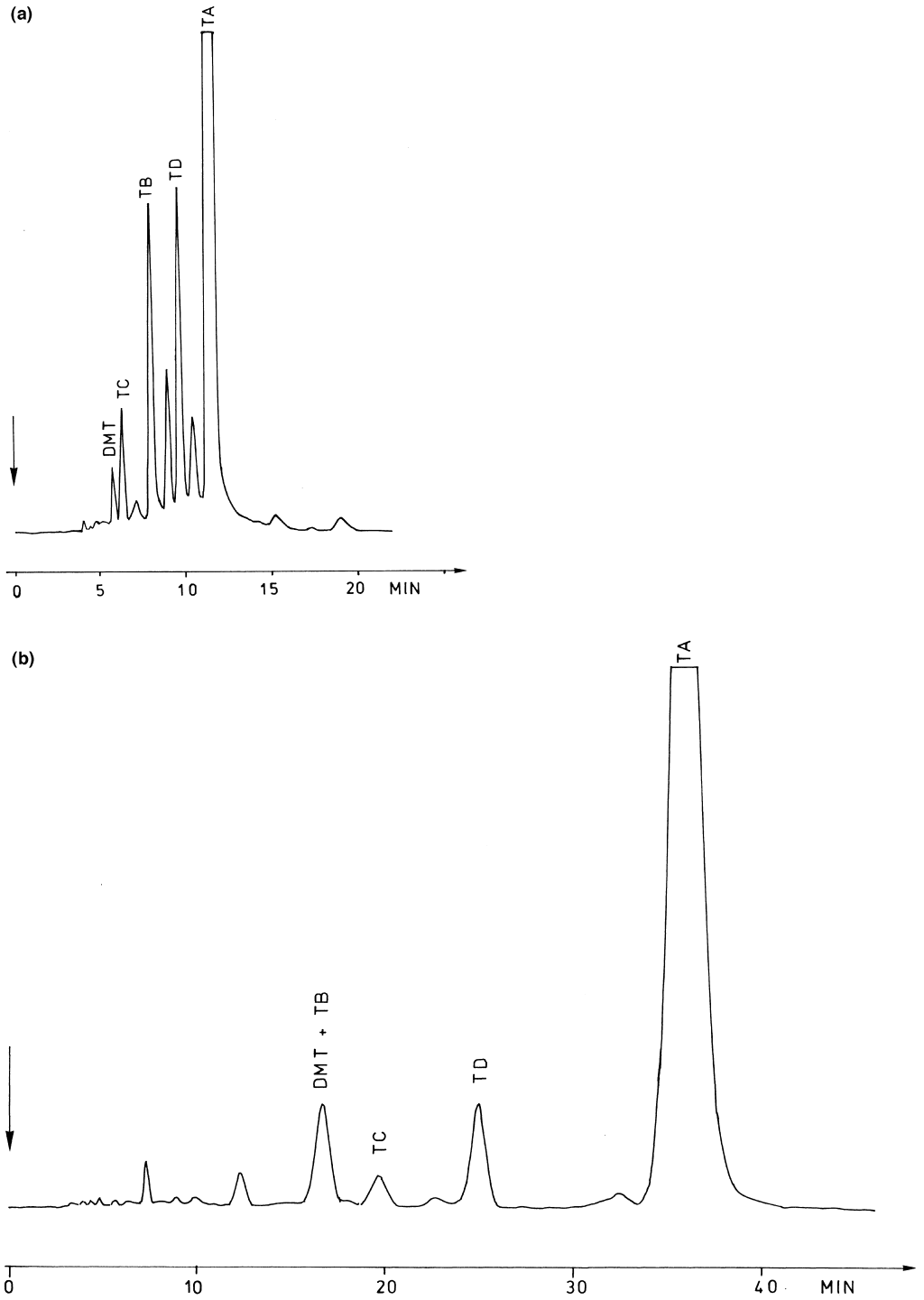


Fig. 4. Chromatogram from sample S3, (a) obtained with method I, and (b) obtained with method II. AUFS = 0.064.

contents. A chromatogram from one of the samples (S3) obtained by both methods is given in Fig. 4.

4.2. Preliminary checking of the equipment

Prior to the analysis of the samples, the equipment was checked as required in Ref. [2]. Therefore reference solutions (a), (c), (de) and (f), and the blank solution were injected. Reference solutions (de) and (f) are used to identify the peaks of Tylosin A, B, C, D and of DMT, and to determine the resolutions (see Appendix A) between the peaks. The results of reference solution (c) and of the blank solution are used to calculate the signal-to-noise (S/N) ratio (see Appendix A). Reference solution (a) is injected six times under repeatability conditions and the relative standard deviation (R.S.D.) of the peak area for Tylosin A is reported. The results of these preliminary experiments are shown in Table 3. The requirements, defined in Ref. [2], are the following, (i) the resolution between Tylosin A–D should be greater than 2.0 for method I and greater than 4.0 for method II, (ii) the S/N ratio should be greater than 10.0, and (iii) the repeatability of injection of reference solution (a) should be less than 1.0%. For method I all these requirements were fulfilled. However, a remarkable fact is reported by one laboratory (** in Table 3) in which the sequence of DMT and TC is inversed. This was confirmed in the concerned laboratory by addition of the substances to one of the samples. It does not seem to be caused by the fact that the standard vials contain the wrong substances

Table 3
Results of the preliminary experiments, (a) from method I, and (b) from method II

Laboratory No.	Mobile phase (% ACN)	Retention time (t_R)	Resolution				S/N	Repeatability of injection (RSD, %)
			DMT–TC	TC–TB	TB–TD	TD–TA		
(a) Method I								
1	35.0	21.59	3.81	4.55	5.98	3.80	54.0	0.48
2	40.0	11.36	1.67	4.58	3.87	3.96	56.8	0.48
3	40.0	15.16	1.81	4.10	5.04	3.92	61.2	0.17
5	40.0	14.66	3.05	0.92	7.78	2.11	94.0	0.98
6	40.0	15.55	1.84	3.34	4.55	3.44	141.0	0.52
7	40.0	20.13	2.79	3.65	5.86	3.53	15.5	0.26
8	40.0	15.38	2.90	3.02	5.21	3.03	16.5	0.21
9	40.0	13.25	1.55	3.52	4.61	2.90	125.5	0.59
10	50.0	9.91	1.64	3.75	3.87	2.49	94.5	0.43
11	40.0	11.13	1.34	3.91	3.68	3.70	121.1	0.41
12	40.0	10.83	1.38	4.35	3.78	3.93	56.3	0.33
13	40.0	11.31	1.26**	1.37	2.75	6.22	87.3	0.98
(b) Method II	Mobile phase (% THF)							
1	20.0	24.65	1.05	1.15	3.38	5.34	14.0	0.21
2	20.0	35.44	0.45	2.14	3.78	5.61	10.8	0.54
3	19.0	49.99	−0.49*	2.10	2.29	3.09	9.3	0.23
5	20.0	32.19	0.62	1.86	4.20	6.55	7.3	0.88
6	20.0	29.71	0.64	2.06	3.96	5.70	34.0	0.06
7	20.0	35.80	0.88	2.18	4.18	6.12	6.2	0.56
8	20.0	47.02	1.01	3.14	4.67	6.00	3.4	0.05
9	20.0	37.18	−0.19*	2.13	3.03	4.20	19.5	0.25
10	25.0	18.66	0.26	0.62	2.93	5.46	20.1	0.83
11	20.0	22.84	1.07	0.92	3.43	4.91	47.3	0.38
12	20.0	27.66	0.28	1.48	2.50	4.04	2.9	0.62
13	22.5	20.46	−0.15*	1.05	2.00	3.80	15.0	0.13

TA, TB, TC and TD stand for Tylosin A, B, C and D, respectively.
* and **: Explained in Section 4.

Table 4
The percentage content of Tylosin A obtained in the different laboratories and for the different samples, (a) with method I and (b) with method II

Laboratory No.	Percentage content Tylosin A						
	S1	S2	S3	S4	S5	S1 (L)	S1 (H)
(a) Method I							
1	84.77	81.03	75.82	36.60	58.61	65.38	91.68
	85.50	82.98	76.23	38.15	59.61	67.59	91.72
2	86.37	82.56	76.77	37.24	60.27	67.93	92.40
	83.51	82.20	75.32	37.33	60.06	67.40	91.06
3	86.69	82.41	77.28	38.62	61.45	69.56	93.46
	85.50	83.14	76.42	38.51	60.59	68.53	93.76
5	86.36	84.23**	82.18	37.10	62.70	71.76	92.03**
	83.09	80.20**	82.62	39.12	60.90	72.44	98.29**
6	85.93	81.62	75.88	37.78	57.82	67.57	93.36
	85.62	82.46	75.83	37.12	58.12	69.12	92.56
7	82.25	82.71	76.11	37.21	59.53	65.95	92.39
	82.15	81.77	76.51	37.03	59.70	66.38	90.91
8	90.60	88.70(**)	82.00	39.59	64.08	70.05	96.38
	90.06	89.31(**)	82.83	39.86	63.99	69.61	96.77
9	83.98	83.30	77.70	37.14	61.23	68.24	94.34
	83.75	83.40	78.06	36.44	61.63	68.74	93.70
10	83.05**	82.02	77.30	40.57	62.32	66.75	90.72
	77.11**	81.60	78.56	39.43	62.21	64.93	92.73
11	84.39	83.16	75.44	37.51	60.14	67.49	92.82
	84.46	83.14	77.39	37.70	60.50	67.44	92.60
12	84.68	81.15	79.50	38.95	59.75	67.45	91.59
	85.45	82.67	81.12	38.42	59.27	67.54	92.35
13	72.97(**)	81.88	70.71	38.01	60.68	52.30(**)	96.20
	73.36(**)	81.39	70.91	37.48	61.22	51.06(**)	96.09
(b) Method II							
1	85.49	81.32	75.81	37.46**	61.14	67.36	92.44
	86.09	83.30	77.79	38.74**	62.27	69.56	93.72
2	82.92	80.29	77.43	36.38	60.81	68.64	92.73
	82.90	82.13	76.29	37.19	60.74	68.19	92.74
3	83.53	81.40	77.36	37.38	60.70	67.35	92.85
	84.39	81.31	77.30	37.31	61.12	67.51	92.25
5	85.25**	81.27	78.27	36.87	63.45	67.56	90.97
	82.24**	81.19	78.03	37.32	63.14	72.75	90.85
6	84.49	83.81	77.12	36.58	59.95	67.41	93.10
	84.56	84.52	77.15	36.47	61.41	67.74	93.45
7	83.60	81.88	77.09	37.02	60.84	66.90	91.95
	83.08	82.39	76.60	37.20	61.22	67.69	93.03
8	83.43	81.47	77.00	37.62	61.25	66.43	92.06
	83.40	81.36	76.73	37.56	61.41	66.65	91.61

Table 4. Continued

Laboratory No.	Percentage content Tylosin A						
	S1	S2	S3	S4	S5	S1 (L)	S1 (H)
9	88.89(**)	83.53	79.03	36.24	60.64	68.96	93.57
	88.40(**)	83.21	78.69	36.24	61.20	68.31	94.24
10	83.39	82.44	77.11	38.66	61.97	63.63	93.60
	83.49	82.36	78.51	38.42	61.46	67.63	93.73
11	83.12	83.38	79.23	37.58	61.19	66.54	92.41
	83.60	84.00	79.48	37.83	60.91	67.34	92.82
12	82.75	82.13	74.70	38.69	59.47	66.19	92.97
	83.55	84.63	75.14	38.47	59.16	67.08	91.85
13	84.02	83.84	75.54	37.84	63.08	67.05	93.24
	84.33	83.44	75.21	38.28	63.55	67.71	92.06

The design is a balanced uniform-level design with 12 laboratories, seven samples and two replicates per sample. Legend: ** = values deleted from dataset after within-laboratory consistency evaluation; (**) = values deleted from dataset after between-laboratory consistency evaluation.

(DMT and TC switched) since the contents reported for DMT and TC in the samples are in correspondence with the other laboratories. Differences in the elution sequence of a same set of substances on similar columns is not so rare and was for another case study also reported in Ref. [9].

For method II some TD–TA resolutions were less than four and some S/N ratios were less than ten. As the deviations from the first requirement were small, all laboratories were nevertheless maintained for the statistical analysis of the data. It was also observed that for method II the separation between DMT and TB is bad. The resolution, in most laboratories, is less than one. For three laboratories (marked with *), a negative resolution is indicated in Table 3 because after injection of reference solutions (de) and (f), the retention time for DMT was found to be longer than that for TB, while in a mixture DMT elutes just before TB. From Table 3 it can also be observed that some laboratories, both for methods I and II, had changed the mobile phase composition as was allowed by the protocols. However, some changed the composition so that the TD–TA resolution and the analysis time were reduced in order to make the analysis faster, which was not the idea of the allowable change. These laboratories were however also maintained in the study since they remained within the required limits after their change or they deviated only slightly from them.

4.3. Original data from the study

The five samples (S1–S5) were assayed. For S1, the low and the high concentration, S1 (L) and S1 (H), were also analysed. For each sample, the test solution was twice prepared independently and the following injection order was maintained: (1) first replicate of the test solution, (2) reference solution (a), (3) second replicate of the test solution and (4) reference solution (b). The run time for each experiment was at least two times the retention time of tylosin A. As required by the ISO guidelines, within each laboratory all samples were analysed by the same operator. The content of Tylosin A in the samples was determined relative to the Tylosin A peak in reference solution (a) and those of the related and unknown substances relative to the Tylosin A peak in reference solution (b) (see Appendix A).

The percentage contents of Tylosin A from both methods, obtained after a thorough check of the original results for erroneous data and after a verification of the reported contents by the statistical officer, are shown in Table 4. This table forms a balanced uniform-level design with 12 laboratories, seven samples and two replicates per cell. After the examination for erroneous results, the reported data were verified for outlying data. Then a further statistical analysis on the remaining data set was performed.

In the samples of method II where DMT co-elutes with tylosin B the result is reported as tylosin B since the latter substance occurs in a much higher amount.

4.4. Statistical analysis

4.4.1. Scrutiny of results for consistency and outliers

Firstly in the statistical analysis, the results were tested for consistency and outliers. Both within-laboratory and between-laboratory consistency were examined. This was in both cases done by means of the graphical Mandel's method and with numerical outlier tests.

It was verified if the within-laboratory variance of some laboratories was not considerably larger than in the other participating laboratories. This was done by plotting Mandel's k statistic (see Fig. 2) while the numerical outlier technique applied was the Cochran's test. The results from Mandel's method and from the Cochran test are shown in Table 5 for the percentage content of Tylosin A. On the Mandel's k plot, indicator lines at significance levels $\alpha = 0.01$ and 0.05 were drawn (see Fig. 2). From Fig. 2, for instance, it can be observed that laboratory 5 tends to show a rather high repeatability variance for the determination of the content of Tylosin A with method I.

Results were removed from the original data set (i) when they were found to be outliers with the numerical technique, which is the requirement of the ISO guideline, and (ii) when they were found to be stragglers with the numerical technique and they were exceeding the $\alpha = 0.01$ indicator line in the graphical technique. The values for the percentage content Tylosin A that were deleted from the original dataset are indicated with ** in Table 4. No data were deleted based on results from a second Cochran test (Table 5).

Is was also verified if there are laboratories with deviating results compared to those of the other laboratories, i.e., examining the between-laboratory consistency. This was done by plotting Mandel's h statistic (see Fig. 3) while the numerical outlier techniques applied were the Grubbs' tests for one or two outlying observations. The results from Mandel's method and from the Grubbs' tests are shown in Table 6 for the percentage content of Tylosin A.

On the Mandel's h plot, indicator lines at significance levels $\alpha = 0.01$ and 0.05 were drawn (see Fig. 3). From Fig. 3 it can be seen that laboratory 8 tends to report higher contents than the other laboratories,

Table 5
Analysis of the data for percentage content Tylosin A as reported in Table 4, for consistency and outliers in the within-laboratory variation, (a) method I and (b) method II

Laboratory No.	Mandel's k statistics						
	S1	S2	S3	S4	S5	S1 (L)	S1 (H)
(a) Method I							
1	0.335	1.342	0.401	1.752	1.424	1.998	0.020
2	1.309	0.247	1.438	0.098	0.304	0.484	0.663
3	0.547	0.496	0.853	0.130	1.220	0.941	0.145
5	1.496	2.769[x]	0.436	2.280	2.546[x]	0.614	3.093[x]
6	0.143	0.576	0.051	0.744	0.426	1.406	0.396
7	0.044	0.648	0.391	0.209	0.238	0.388	0.732
8	0.247	0.421	0.819	0.302	0.132	0.396	0.194
9	0.104	0.071	0.361	0.788	0.563	0.461	0.317
10	2.717[x]	0.287	1.241	1.281	0.146	1.656	0.996
11	0.035	0.013	1.929	0.214	0.510	0.043	0.111
12	0.355	1.046	1.612	0.594	0.691	0.084	0.372
13	0.180	0.336	0.194	0.602	0.759	1.124	0.053
	Indicator values for Mandel's k statistics ($p = 12$, $n = 2$)						
5% level	1.92	1.92	1.92	1.92	1.92	1.92	1.92
1% level	2.36	2.36	2.36	2.36	2.36	2.36	2.36

Table 5. Continued

Laboratory No.	Cochran's test for method I						
	S1	S2	S3	S4	S5	S1 (L)	S1 (H)
C	0.615*	0.639*	0.310	0.433	0.540	0.333	0.797**
	Critical values ($p=12$, $n=2$)						
5%	0.541	0.541	0.541	0.541	0.541	0.541	0.541
1%	0.653	0.653	0.653	0.653	0.653	0.653	0.653
	Second Cochran's test (after elimiation of outliers)						
C	nd	nd	nd	nd	nd	nd	0.407
	Critical values ($p=11$, $n=2$)						
5%							0.570
1%							0.684

(b) Method II

Laboratory No.	Mandel's *k* statistics						
	S1	S2	S3	S4	S5	S1 (L)	S1 (H)
1	0.614	1.773	2.422[x]	2.603[x]	1.791	1.070	1.702
2	0.021	1.647	1.391	1.637	0.118	0.219	0.012
3	0.873	0.082	0.066	0.142	0.658	0.080	0.796
5	3.051[x]	0.069	0.296	0.905	0.481	2.514[x]	0.158
6	0.065	0.633	0.026	0.222	2.296	0.158	0.463
7	0.527	0.451	0.596	0.366	0.599	0.383	1.442
8	0.032	0.095	0.330	0.111	0.257	0.110	0.605
9	0.504	0.288	0.416	0.001	0.875	0.316	0.884
10	0.103	0.069	1.712	0.494	0.807	1.940	0.168
11	0.485	0.552	0.302	0.510	0.455	0.385	0.540
12	0.817	2.236	0.539	0.446	0.484	0.431	1.498
13	0.316	0.358	0.397	0.898	0.749	0.318	1.563
	Indicator values for Mandel's *k* statistics ($p=12$, $n=2$)						
5% level	1.92	1.92	1.92	1.92	1.92	1.92	1.92
1% level	2.36	2.36	2.36	2.36	2.36	2.36	2.36
	Cochran's test for method II						
C	0.776**	0.417	0.489	0.565*	0.439	0.527	0.241
	Critical values ($p=12$, $n=2$)						
5%	0.541	0.541	0.541	0.541	0.541	0.541	0.541
1%	0.653	0.653	0.653	0.653	0.653	0.653	0.653
	Second Cochran's test (after elimiation of outliers)						
C	0.283						
	Critical values ($p=11$, $n=2$)						
5%	0.570						
1%	0.684						

Legend: x = above 1% level indicator value, * = straggler, ** = outlier, nd = not done.

for Tylosin A determined with method I, while laboratory 13 has a tendency towards low results.

If necessary, results were removed from the original data set according to the criteria mentioned above. The values for the percentage content Tylosin A that were deleted from the original dataset after the between-laboratory variation evaluation are indicated with (**) in Table 4. No data were deleted

Table 6
Analysis of the data for percentage content Tylosin A reported in Table 4 (after elimination of the within-laboratory variance outliers for the Grubbs' tests), for consistency and outliers in the between-laboratory variation, (a) method I and (b) method II

Laboratory No.	Mandel's h statistics						
	S1	S2	S3	S4	S5	S1 (L)	S1 (H)
(a) Method I							
1	0.320	−0.436	−0.444	−0.649	−0.978	−0.046	−0.909
2	0.273	−0.250	−0.438	−0.740	−0.322	0.188	−0.894
3	0.552	−0.051	−0.186	0.514	0.208	0.462	0.157
5	0.220	−0.333	1.562	0.072	0.696	1.068	1.020
6	0.474	−0.416	−0.498	−0.575	−1.685	0.324	−0.207
7	−0.391	−0.317	−0.354	−0.902	−0.663	−0.110	−0.936
8	1.579	x3.064	1.566	1.657	2.083	0.618	1.811
9	0.012	0.236	0.140	−1.225	0.464	0.352	0.386
10	−0.905	−0.533	0.155	1.925	0.985	−0.174	−0.895
11	0.147	0.137	−0.322	−0.423	−0.228	0.148	−0.346
12	0.302	−0.482	0.904	0.636	−0.729	0.155	−0.758
13	−2.582[x]	−0.620	−2.085	−0.289	0.168	−2.985[x]	1.571
	Indicator values for Mandel's h statistics ($p=12$)						
5% level	1.83	1.83	1.83	1.83	1.83	1.83	1.83
1% level	2.25	2.25	2.25	2.25	2.25	2.25	2.25
	Grubbs' tests (one outlying observation)						
G_1	2.650**	0.623	2.085*	1.225	1.685	2.985**	0.849
	Critical values for G_1						
p	11	11	12	12	12	12	11
5%	2.355	2.355	2.412	2.412	2.412	2.412	2.355
1%	2.564	2.564	2.636	2.636	2.636	2.636	2.564
G_p	2.434*	2.908**	1.566	1.925	2.083	2.226	1.917
	Critical values for G_p						
p	10	11	11	12	12	11	11
5%	2.290	2.355	2.355	2.412	2.412	2.355	2.355
1%	2.482	2.564	2.564	2.636	2.636	2.564	2.564
	Grubbs' tests (two outlying observations)						
	S1	S2	S3	S4	S5	S1 (L)	S1 (H)
$G_{1,2}$	nd	0.938	0.522	0.748	0.591	nd	0.848
	Critical values for $G_{1,2}$						
p		11	12	12	12		11
5%		0.211	0.254	0.254	0.254		0.211
1%		0.145	0.174	0.174	0.174		0.145
$G_{p-1,p}$	0.730	nd	0.466	0.297	0.432	0.836	0.214
	Critical values for $G_{p-1,p}$						
p	11		12	12	12	12	11
5%	0.211		0.254	0.254	0.254	0.254	0.211
1%	0.145		0.174	0.174	0.174	0.174	0.145

Table 6. Continued

Laboratory No.	Mandel's *h* statistics						
	S1	S2	S3	S4	S5	S1 (L)	S1 (H)
(b) Method II							
1	0.993	−0.197	−0.308	0.828	0.334	0.731	0.509
2	−0.811	−1.214	−0.259	−0.906	−0.509	0.696	0.069
3	−0.154	−1.075	0.108	−0.169	−0.385	−0.134	−0.154
5	−0.288	−1.193	0.756	−0.499	1.775	2.156	−2.222
6	0.201	1.512	−0.046	−1.244	−0.597	−0.014	0.753
7	−0.541	−0.361	−0.273	−0.476	−0.280	−0.251	−0.235
8	−0.495	−1.024	−0.258	0.153	−0.004	−0.884	−1.058
9	2.780[x]	0.777	1.313	−1.622	−0.379	0.881	1.543
10	−0.477	−0.113	0.486	1.402	0.343	−1.649	1.240
11	−0.530	1.075	1.702	0.301	−0.260	−0.547	−0.076
12	−0.660	0.785	−1.790	1.462	−1.832	−0.804	−0.334
13	−0.019	1.027	−1.432	0.770	1.793	−0.179	−0.035
	Indicator values for Mandel's *h* statistics ($p=12$)						
5% level	1.83	1.83	1.83	1.83	1.83	1.83	1.83
1% level	2.25	2.25	2.25	2.25	2.25	2.25	2.25
	Grubbs' tests (one outlying observation)						
G_1	0.802	1.214	1.790	1.528	1.832	1.649	2.222
	Critical values for G_1						
p	11	12	12	11	12	12	12
5%	2.355	2.412	2.412	2.355	2.412	2.412	2.412
1%	2.564	2.636	2.636	2.564	2.636	2.636	2.636
G_p	2.636**	1.512	1.702	1.518	1.793	2.156	1.543
	Critical values for G_p						
p	11	12	12	11	12	12	12
5%	2.355	2.412	2.412	2.355	2.412	2.412	2.412
1%	2.564	2.636	2.636	2.564	2.636	2.636	2.636
	Grubbs' tests (two outlying observations)						
	S1	S2	S3	S4	S5	S1 (L)	S1 (H)
$G_{1,2}$	0.889	0.684	0.428	0.566	0.609	0.623	0.352
	Critical values for $G_{1,2}$						
p	11	12	12	11	12	12	12
5%	0.221	0.254	0.254	0.221	0.254	0.254	0.254
1%	0.145	0.174	0.174	0.145	0.174	0.174	0.174
$G_{p-1,p}$	nd	0.626	0.497	0.468	0.306	0.423	0.574
	Critical values for $G_{p-1,p}$						
p		12	12	11	12	12	12
5%		0.254	0.254	0.221	0.254	0.254	0.254
1%		0.174	0.174	0.145	0.174	0.174	0.174

Legend: x = above 1% level indicator value, * = straggler, ** = outlier, nd = not done.

based on results from a two outlying observations Grubbs' test (Table 6).

4.4.2. Calculation of the general mean and variances

For Tylosin A, B, C, D and for DMT, the general mean and the variances were calculated for each sample using the remaining data set. The variances determined are (i) the repeatability variance (s^2_{rj}), (ii) the between-laboratory variance (s^2_{Lj}) and (iii) the reproducibility variance (s^2_{Rj}). In Table 7 the results for tylosin A are shown.

4.4.3. Establishing a functional relationship between precision values and the mean contents

After the calculation of the mean and the variances it is studied whether or not there is a relationship between the mean contents on the one hand and (i) the repeatability or (ii) the reproducibility standard deviations on the other. If there is no relationship, the finally reported standard deviation is the average one estimated for the different samples. An example where no relationship is observed is given in Fig. 5. The reported standard deviations for tylosin A are shown in Table 8.

If a regular relationship exists, three types are considered by the ISO guidelines, (i) $s = b_1 m$ (straight line through the origin), (ii) $s = b_0 + b_1 m$ (straight line with a positive intercept), (iii) $\log s = b_0 + b_1 \log m$ (or $s = C \cdot m^{b_1}$, exponential relationship) where s stands for the repeatability or the reproducibility standard deviation, b_0 for the intercept, b_1 for the slope, m for the mean content and C for a constant related to b_0. From a plot of s_j as a function of $\hat{m}_j$ other relationships can be found and the statistical officer then should seek an appropriate solution. It is also evident that an obtained relationship only applies in the concentration interval covered by the samples examined in the interlaboratory study.

To estimate the coefficients of the straight lines of relationships (i) and (ii), (iterative) weighted regression methods are used because the standard error of s is proportional to the predicted value of s_j ($=\hat{s}_j$). From a statistical point of view, the fitting of a straight line is complicated because both $\hat{m}_j$ and $\hat{s}_j$ are estimated and subjected to error. Such situations require an errors-in-variables regression method [8]. However since the slope b_1 is usually small (of the order of 0.1 or less) [7] the errors in $\hat{m}_j$ have little influence and those in s predominate. Therefore a common least-squares regression method is applied. The weighting factors used are proportional to $1/\hat{s}_j^2$. However $\hat{s}_j$ depends on the parameters that are to be estimated (b_0 and b_1). To find estimates corresponding to the weighted least-squares of residuals method an iterative approach is used. The following procedure, which has proved to be satisfactory in practice, is recommended [7].

For relationship (i), b_1 can be estimated from the simplified expression

Table 7
The mean contents and the variances estimated for tylosin A

	S1	S2	S3	S4	S5	S1 (L)	S1 (H)
Method I							
Mean % content	85.3	82.3	77.4	38.0	60.7	68.1	93.2
Variances							
Repeatability	1.10	0.44	0.51	0.39	0.25	0.59	0.45
Between-laboratories	3.80	0.12	9.85	0.84	2.46	2.96	2.94
Reproducibility	4.89	0.56	10.36	1.24	2.71	3.55	3.40
Method II							
Mean % content	83.8	82.5	77.2	37.4	61.3	67.6	92.7
Variances							
Repeatability	0.12	0.63	0.34	0.06	0.20	2.13	0.28
Between-laboratories	0.67	0.86	1.44	0.56	1.12	0.35	0.49
Reproducibility	0.79	1.49	1.78	0.62	1.32	2.48	0.78

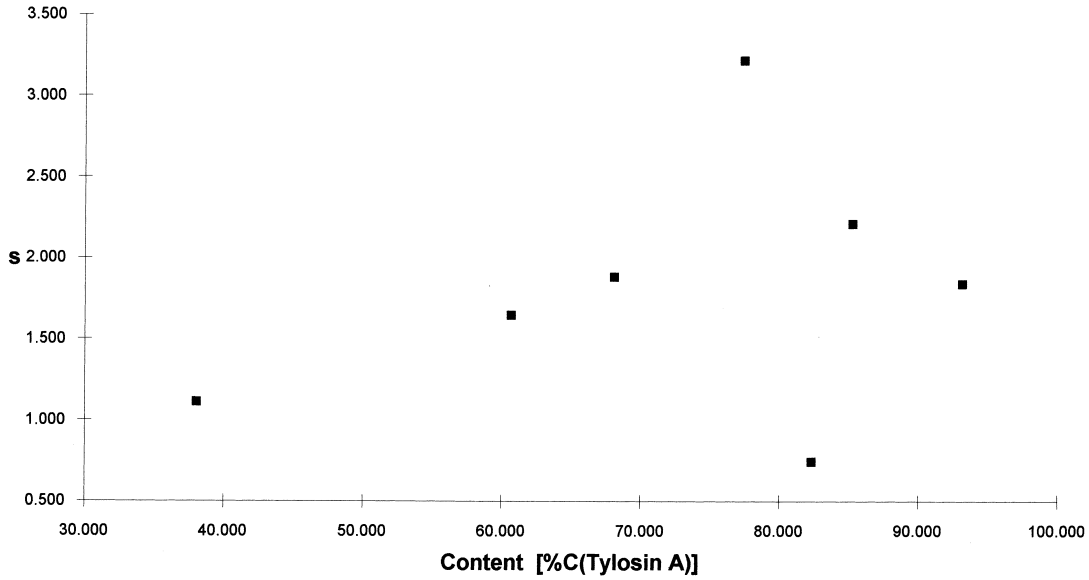


Fig. 5. Reproducibility standard deviation as a function of the content tylosin A, obtained for method I.

$$b_1 = \frac{\sum_j \left(\frac{S_j}{\hat{m}_j}\right)}{q} \qquad (15)$$

and no iteration is necessary.

For relationship (ii) an iterative approach is used. The initial weights $W_{0j} = 1/\hat{s}^2_{0j}$ with $j = 1, 2, \ldots, q$ and $\hat{s}^2_{0j}$ equal to the variances, s^2_{rj} or s^2_{Rj}, calculated for the different samples in the previous step, are used to estimate the coefficients b_{01} and b_{11} with weighted

Table 8
Reported standard deviations for tylosin A and comparison of the results

	Variances		F	$F_{critical}$	Means	
	Method I	Method II			Method I	Method II
Repeatability	0.51	0.39	1.29	$F(78,81)$ 1.49	85.3	83.8
Reproducibility	3.27	1.25	2.62	$F(91,102)$ 1.42	82.3	82.5
					77.4	77.2
	Standard deviations				38.0	37.4
	Method I	Method II			60.7	61.3
Repeatability	0.71	0.63			68.1	67.6
Reproducibility	1.81	1.12			93.2	92.7
	Measured variances (reproducibility)				Paired t-test	
Sample	Method I	Method II			$t = 1.394$	
S1	4.89	0.79	6.21	$F(11,10) = 2.94$	$t(6\ \mathrm{df}) = 2.44$	Two-tailed
S2	0.56	1.49	2.68	$F(17,18) = 2.24$		
S3	10.36	1.78	5.82	$F(12,13) = 2.60$		
S4	1.24	0.62	1.99	$F(15,11) = 2.72$		
S5	2.71	1.32	2.06	$F(12,13) = 2.60$		
S1 (L)	3.55	2.48	1.43	$F(12,22) = 2.23$		
S1 (H)	3.40	0.78	4.39	$F(11,16) = 2.46$		

least-squares regression. This leads to the equation $\hat{s}_{1j} = b_{01} + b_{11}\hat{m}_j$. The calculations are then repeated with $W_{1j} = 1/\hat{s}_{1j}^2$ to produce $\hat{s}_{2j} = b_{02} + b_{12}\hat{m}_j$. The same procedure could now be repeated but leads to unimportant changes. The step from W_{0j} to W_{1j} is effective in eliminating gross errors in the weights and the equation for $\hat{s}_{2j}$ can be considered as the final result. In this study all observed linear relationships

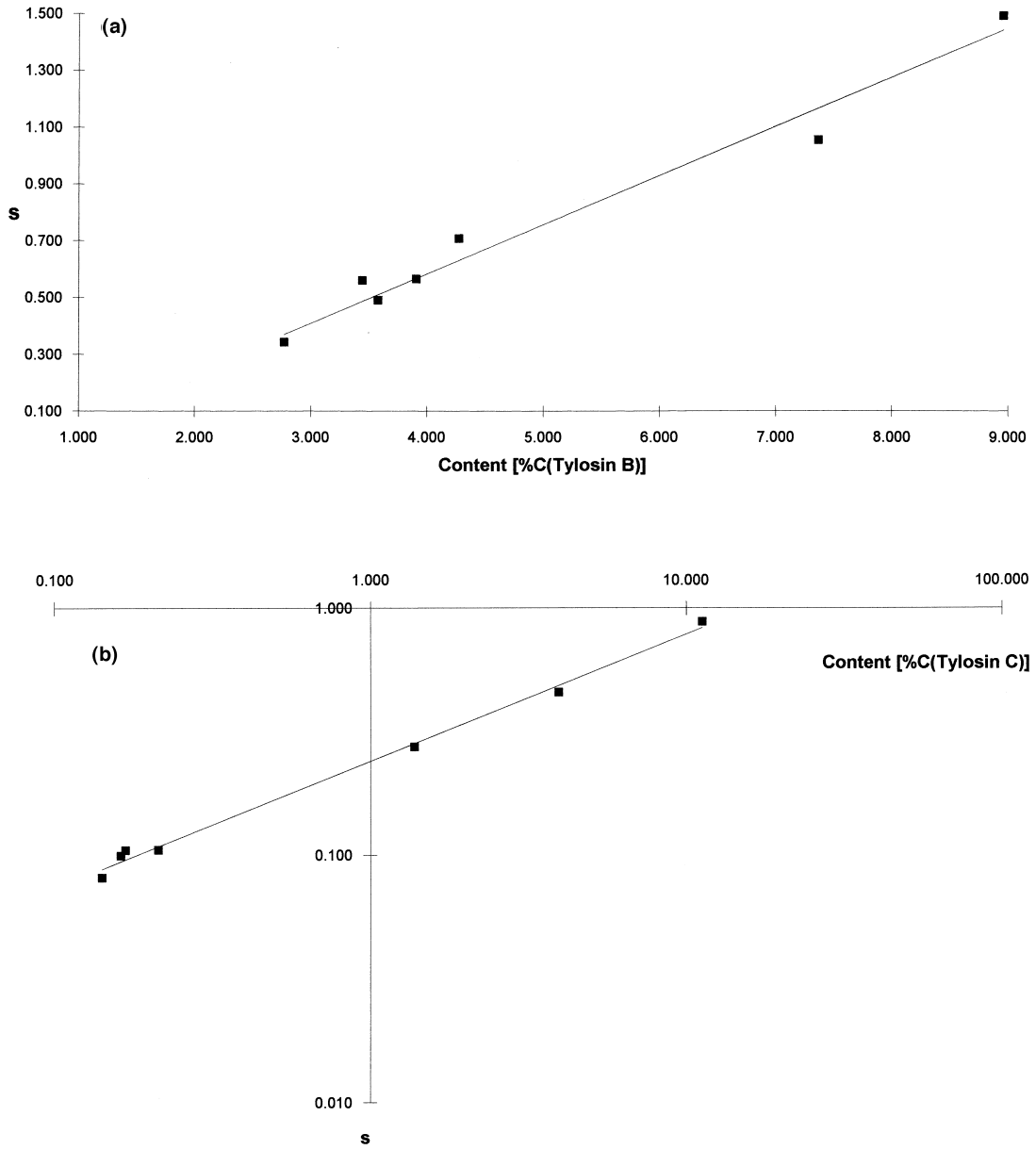


Fig. 6. (a) Linear relationship between the standard deviation and the mean content. The example shown is the reproducibility standard deviation of tylosin B in method II ($s = 0.173 \cdot C - 0.109$). (b) Exponential relationship between the standard deviation and the mean content. The example shown is the reproducibility standard deviation of tylosin C in method II [$\log(s) = 0.512 \cdot \log(C) - 0.619$].

Table 9
Comparison of the results for Tylosin B (variance dependent on concentration)

Sample	Estimated variances (reproducibility)		F		Means	
	Method I	Method II		$F_{critical}$	Method I	Method II
S1	0.21	0.26	1.25	$F(10,10)=2.98$	4.18	3.58
S2	0.20	0.24	1.19	$F(11,10)=2.94$	4.05	3.44
S3	0.23	0.40	1.73	$F(11,10)=2.94$	4.43	4.27
S4	0.72	2.08	2.87	$F(9,10)=3.02$	8.67	8.96
S5	0.62	1.36	2.19	$F(11,10)=2.94$	7.94	7.37
S1 (L)	0.14	0.14	1.04	$F(11,12)=2.72$	3.28	2.77
S1 (H)	0.24	0.32	1.35	$F(11,12)=2.72$	4.54	3.91
	Measured variances (reproducibility)				Paired t-test	
	Method I	Method II			$t=3.060$	
S1	0.32	0.24	1.34	$F(10,10)=2.98$	$t(6\ df)=2.447$	Two-tailed
S2	0.18	0.31	1.76	$F(11,10)=2.94$	$t(6\ df)=1.943$	One-tailed
S3	0.24	0.50	2.12	$F(11,10)=2.94$		
S4	0.70	2.22	3.17	$F(9,10)=3.02$		
S5	0.85	1.11	1.31	$F(11,10)=2.94$		
S1 (L)	0.19	0.12	1.61	$F(11,12)=2.72$		
S1 (H)	0.15	0.32	2.15	$F(11,12)=2.72$		

between s and m were treated according to the latter approach.

For relationship (iii) an ordinary least-squares method is used since the standard error of is independent of s. An example of a linear relationship between s and m is shown in Fig. 6a and of an exponential one in Fig. 6b. The estimated coefficients are then used to predict the standard deviations and the variances at a given concentration level. The variances presented in Tables 9 and 10 for instance were derived from such relationships. It was observed that when the ratio between the smallest and the largest content of a substance in the different samples is large enough, the standard deviation depended on the concentration. When this ratio is intermediate (e.g., tylosin B) the relationship is linear and when it is large (e.g., tylosin C and DMT) exponential. Now when it is small, no relationship can be detected, but this does not mean it does not exist (e.g., tylosin A and D).

4.5. Comparison of the two methods

Since studies were performed on both methods for the same samples, their results, i.e., the variances and the average content of the different substances, can be compared. Beside the quantitative results, other parameters were also compared, such as for instance differences in separation (resolutions between peaks and number of peaks detected for a sample), sensitivity, analysis time, applicability of a method, and specific problems reported by the participants.

The comparison of the variances and means obtained is not described in the guidelines of Ref. [7] since they are meant for evaluation of one method. It is however the part that allows to make conclusions about the performance of both methods. The reproducibility variances for tylosin A, B, C, D and DMT, both the measured ones and those estimated from the observed relationships between the standard deviations and the content, are shown in Tables 8–12. From Fig. 4 it can also be seen that besides the known substances, several unknowns can be detected. This is the case for each sample. The data of these substances were reported but not further analysed since this was not the object of the intercomparison and it is not obvious to identify these peaks in both methods.

Comparison of the measured variances with the predicted ones shows a good agreement between both, indicating that the models chosen to estimate

Table 10
Comparison of the results for Tylosin C (variance dependent on concentration)

Sample	Estimated variances (reproducibility)		F	$F_{critical}$	Means	
	Method I	Method II			Method I	Method II
S1	$3.6\cdot10^{-4}$	$8.9\cdot10^{-3}$	24.94	$F(3,4)=6.59$	0.04	0.16
S2	$1.6\cdot10^{-3}$	$1.2\cdot10^{-2}$	7.46	$F(7,9)=3.29$	0.12	0.21
S3	$6.3\cdot10^{-2}$	$8.0\cdot10^{-2}$	1.27	$F(11,11)=2.82$	1.49	1.38
S4	1.4	$6.9\cdot10^{-1}$	1.96	$F(10,12)=2.75$	11.80	11.25
S5	$2.8\cdot10^{-1}$	$2.4\cdot10^{-1}$	1.21	$F(10,11)=2.85$	4.11	3.94
S1 (L)	$3.0\cdot10^{-4}$	$7.7\cdot10^{-3}$	25.44	$F(4,5)=5.19$	0.04	0.14
S1 (H)	$5.1\cdot10^{-4}$	$9.3\cdot10^{-3}$	18.08	$F(5,6)=4.39$	0.06	0.17
	Measured variances (reproducibility)				Paired t-test	
	Method I	Method II			$t=0.640$	
S1	$4.3\cdot10^{-4}$	$1.0\cdot10^{-2}$	23.15	$F(3,4)=6.59$	$t(6\ df)=2.447$	Two-tailed
S2	$1.7\cdot10^{-3}$	$1.1\cdot10^{-2}$	6.69	$F(7,9)=3.29$		
S3	$4.8\cdot10^{-2}$	$7.6\cdot10^{-2}$	1.56	$F(11,11)=2.82$		
S4	1.7	$7.8\cdot10^{-1}$	2.13	$F(10,12)=2.75$		
S5	$2.6\cdot10^{-1}$	$2.1\cdot10^{-1}$	1.26	$F(10,11)=2.85$		
S1 (L)	$2.0\cdot10^{-4}$	$6.6\cdot10^{-3}$	32.93	$F(4,5)=5.19$		
S1 (H)	$7.1\cdot10^{-4}$	$1.1\cdot10^{-2}$	15.43	$F(5,6)=4.39$		

the relationships between standard deviations and substance contents were appropriately selected.

When, for a given sample, the reproducibility variances for the two methods were compared, the number of degrees of freedom applied to determine the critical F value are those obtained using a Satterthwaite approximation [10,11]. The following conclusions can be drawn for the analysed substances. For the main substance, tylosin A, the repeatability variances are equal while the repro-

Table 11
Comparison of the results for Tylosin D (variance independent on concentration)

	Variances		F	$F_{critical}$	Means	
	Method I	Method II			Method I	Method II
Repeatability	$4.4\cdot10^{-3}$	$3.0\cdot10^{-3}$	1.44	F(77,82) 1.47	3.14	2.55
Reproducibility	0.28	0.19	1.48	F(71,76) 1.49	4.19	3.36
					5.24	4.54
	Standard deviations				5.61	5.16
	Method I	Method II			5.53	4.85
Repeatability	$6.6\cdot10^{-2}$	$5.5\cdot10^{-2}$			2.50	2.02
Reproducibility	0.53	0.44			3.45	2.78
	Measured variances (reproducibility)				Paired t-test	
	Method I	Method II			$t=12.855$	
S1	0.16	0.14	1.14	F(11,11)=2.82	$t(6\ df)=2.447$	Two-tailed
S2	0.26	0.32	1.19	F(10,10)=2.98		
S3	0.76	0.10	7.74	F(10,10)=2.98		
S4	0.21	0.22	1.07	F(11,8)= 3.31		
S5	0.57	0.38	1.51	F(10,11)=2.85		
S1 (L)	0.11	0.10	1.07	F(10,12)=2.75		
S1 (H)	0.16	0.17	1.05	F(11,12)=2.72		

Table 12
Comparison of the results for DMT (variance dependent on concentration)

	Estimated variances (reproducibility)		F	$F_{critical}$	Means	
Sample	Method I	Method II			Method I	Method II
S1	$4.2\cdot10^{-4}$	$1.6\cdot10^{-3}$	3.74	$F(1,9)=5.12$	0.08	0.10
S2	$5.2\cdot10^{-4}$	$1.7\cdot10^{-3}$	3.29	$F(1,9)=5.12$	0.09	0.11
S3	$2.2\cdot10^{-2}$	$1.8\cdot10^{-2}$	1.18	$F(9,4)=6.00$	0.60	0.48
S4	$4.0\cdot10^{-1}$	$1.3\cdot10^{-1}$	3.21	$F(11,4)=5.93$	2.70	1.59
S5	$1.6\cdot10^{-2}$	$1.5\cdot10^{-2}$	1.06	$F(10,5)=4.74$	0.52	0.43
S1 (L)	$3.2\cdot10^{-4}$	$1.3\cdot10^{-3}$	3.90	$F(1,9)=5.12$	0.07	0.09
S1 (H)	$5.6\cdot10^{-4}$	$1.9\cdot10^{-3}$	3.40	$F(1,9)=5.12$	0.09	0.12
	Measured variances (reproducibility)				Paired t-test	
	Method I	Method II			$t=1.108$	
S1	$4.2\cdot10^{-4}$	$1.3\cdot10^{-3}$	3.03	$F(1,9)=5.12$	$t(6\ df)=2.447$	Two-tailed
S2	$2.8\cdot10^{-4}$	$1.3\cdot10^{-3}$	4.76	$F(1,9)=5.12$		
S3	$1.1\cdot10^{-2}$	$4.3\cdot10^{-2}$	3.78	$F(4,9)=3.63$		
S4	$6.4\cdot10^{-1}$	$5.8\cdot10^{-2}$	11.06	$F(11,4)=5.93$		
S5	$1.4\cdot10^{-2}$	$2.8\cdot10^{-2}$	1.95	$F(5,10)=3.33$		
S1 (L)	$4.6\cdot10^{-4}$	$1.1\cdot10^{-3}$	2.38	$F(1,9)=5.12$		
S1 (H)	$1.0\cdot10^{-3}$	$1.8\cdot10^{-3}$	1.77	$F(1,9)=5.12$		

ducibility variance of method I is found to be larger than the one of method II. For the related substances, it is observed that when the content of a substance is relatively high, the variances are comparable for both methods but when it becomes small (e.g., $<1\%$) then the variance (both repeatability and reproducibility) of method II is considerably higher than for method I. This is maybe less explicit for the results of DMT but only few laboratories were able to determine it with method II which makes the results less reliable. The difference in variance at low concentrations can be explained by the fact that method II is less sensitive than method I which can be observed from the difference in the S/N ratio (see Table 3).

The differences in variance observed for tylosin A (Table 8) and for tylosin D (Table 11, for S3 individually determined) are therefore, in our opinion, not due to a less good precision of method I but to the complexity of the tylosin mixtures and to the differences in selectivity observed on the different columns. It can, for instance, be seen in Fig. 4a that an unknown substance is eluting between TD and TA. However in some laboratories no peak was observed between these two substances which involves that its peak area will be added to the one of TA or to the one of TD or partly to both.

When the average contents were compared by means of a paired t test [8,12], it was seen that for Tylosin A they were comparable for both methods. For Tylosins B, C, D and DMT (for contents above 1%, because lower ones were considered unreliable with method II) it is observed that the contents determined with method I are systematically higher than those with method II. An explanation for this is that in both methods the contents are determined relative to the Tylosin A peak of reference solution (b), but that the detection wavelength for both methods is different, namely 290 nm and 280 nm for methods I and II, respectively. This means that when the substances have absorption spectra with a different $\epsilon_{substance}/\epsilon_{TA}$ ratio (where ϵ is the molar absorption coefficient) at both wavelengths, this explains the differences in the estimated contents. It is namely so that the concentration estimated for a substance is directly related to the ratio between the area of the substance and the one of Tylosin A $[C_{substance} \div (\text{Area}_{substance}/\text{Area}_{TA})\cdot C_{TA}]$. Since the detection is spectrophotometric it is also directly related to the ratio between the absorptions of the substances and to the one between the absorption coefficients of the substances $[C_{substance} \div (\text{Area}_{substance}/\text{Area}_{TA}) \div (\epsilon_{substance}/\epsilon_{TA})]$. To have a correct estimation of the content of a substance the ratio $\epsilon_{substance}/\epsilon_{TA}$ should

be equal to one. Now, (i) when the ratio is different from one, incorrect estimates are made, and (ii) when it is different at one wavelength compared to another, different estimates are made. The latter is what probably has happened here.

It was also reported by a participant that the absorption maxima for all known substances was determined as 288 nm in both methods [determined with HPLC–diode array detection (DAD)] which makes 290 nm more recommendable than 280 nm since then one works at the wavelength with maximal absorption which involves that the measurement is more sensitive and more robust than when measuring in the slope of a spectrum peak which will be the case at 280 nm.

From the comparison of both methods a number of disadvantages for both could be formulated. The disadvantages of method II are that (i) only one type of stationary phase from one manufacturer can be used, (ii) the analysis time [see t_R (TA) in Table 3] is much longer, (iii) DMT and Tylosin B are not or badly separated; moreover the best separations between these substances were observed on the home packed columns (see Tables 1 and 3), (iv) small variations in the THF concentration in the mobile phase cause large changes in the retention times (see Table 3). The latter observation also explains the results, reported by two of the participants, that (a) different retention times can be observed between consecutively prepared mobile phases, and (b) the analysis time is extremely dependent on the degassing method used. It was reported that, on a same instrument, the retention times of Tylosin B, D and A were 20.1, 32.6 and 47.0 min, respectively when degassing the mobile phase by vacuum-filtration, and 13.8, 20.1 and 28.7 min when degassing by water-pump-filtration. Vacuum-filtration probably reduced the THF content slightly with an increase in retention times as a consequence.

The disadvantages of method I are that (i) the mobile phase is not ''equipment-friendly'', since it has a relatively low pH, an extremely high salt (sodium perchlorate) concentration which can cause crystallisation problems (reported by one laboratory) and contains the corrosive chloride ion, and (ii) the separation between the substances is not always optimal neither, i.e., for instance, not all unknown substances are separated from TD and/or TA. These latter observations confirm the column dependent selectivity towards minor impurities which was reported in Ref. [1]. The occurrence of ghost peaks and stability problems in reference solutions (de) and (f) were reported during the training round by two laboratories and repeated for the final study results. We were not able to find the reason for it. Injection of reference solutions (de) for more than 48 h did not indicate a degradation of the substances and certainly not the occurrence of new (ghost) peaks.

5. Conclusions

The reversed-phase method described in Refs. [4–6] in general performs better than the one described in Ref. [3]. The repeatability and reproducibility variances are smaller, especially for low concentrations. However, the requirement that the resolution between TD and TA should be at least two is probably too low. A higher limit could help to avoid that unknown substances co-elute with one of the major substances of the tylosin mixture. It would also allow a more exact determination of the composition, which in fact was the original aim of this particular method.

Acknowledgements

Parts of this work were funded by the Research Contract No. NO/03/003 of the Belgian government (Scientific Support for Standardisation Programme) and by the European Community Standards, Measurements and Testing research programme. The authors are grateful to V. Reynders and K. Decq for technical aid.

Appendix A

(1) The resolution between consecutive peaks was calculated using the formula

$$R_S = \frac{1.18(t_{R_b} - t_{R_a})}{W_{a(0.5)} + W_{b(0.5)}} \quad \text{(A1)}$$

where t_{R_a} and t_{R_b} are the retention times, $W_{a(0.5)}$ and

$W_{b(0.5)}$ are the peak widths at half height of the first and second peak.

(2) The signal-to-noise ratio is calculated from the equation

$$S/N = \frac{2H}{h_n} \quad \text{(A2)}$$

where H is the height of the tylosin A peak in the chromatogram obtained with reference solution (c) and h_n the absolute value of the largest noise fluctuation from the baseline in a chromatogram obtained after injection of the blank solution and observed over a distance equal to 20 times the width at half height obtained with reference solution (c) and situated equally around the retention time of tylosin A. Measurement of H and h_n is shown in Fig. 7.

(3) The percentage content of tylosin A (% C_{TA}) is calculated using the chromatograms obtained with reference solution (a) and the test solution using the following formula

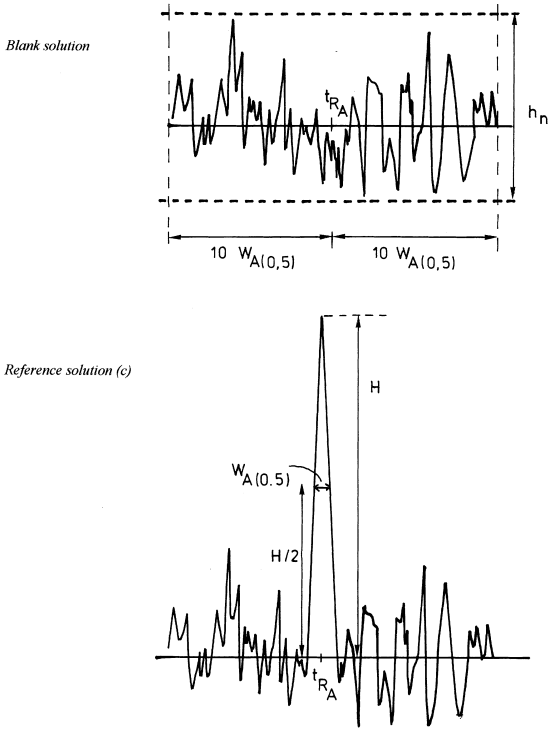


Fig. 7. Measurement of H and h_n for the calculation of the signal-to-noise ratio.

$$\%\, C_{TA} = \frac{A_{test}}{A_{ref}} \cdot \frac{m_{ref}}{m_{test}} \cdot 100 \quad \text{(A3)}$$

where A_{test} and A_{ref} are the peak areas of tylosin A in a test and in the reference solution, respectively, m_{test} and m_{ref} are the masses of sample and of tylosin A CRS, respectively, weighed to prepare identical volumes of test and of reference solution.

The percentages content of DMT (% C_{DMT}), tylosin B (% C_{TB}), tylosin C (% C_{TC}), tylosin D (% C_{TD}) and unknown substances are calculated using the chromatograms obtained with reference solution (b) and the test solution using the following formula

$$\%\, C_T = \frac{A'_{test}}{A'_{ref}} \cdot \frac{m_{ref}}{m_{test}} \cdot 4 \quad \text{(A4)}$$

where A'_{test} is the peak area of DMT, tylosin B, tylosin C, tylosin D or an unknown substance in the test solution chromatogram, A'_{ref} the peak area of tylosin A in the reference solution (b) chromatogram, m_{test} and m_{ref} are the masses of sample and of tylosin A CRS, respectively, weighed to prepare identical volumes of test and of reference solution (a).

For the contents in the low and the high level test solutions, Eqs. (A3) and (A4) are multiplied by 0.8 and 1.1, respectively.

References

[1] E. Roets, P. Beirinckx, I. Quintens, J. Hoogmartens, J. Chromatogr. 630 (1993) 159–166.
[2] European Pharmacopoeia Commission, Tylosini Tartras, PA/PH/Exp. 7/T (93) 41, ANP, February 1995, unpublished document.
[3] J. Paesen, P. Claeys, W. Cypers, E. Roets, J. Hoogmartens, J. Chromatogr. A 699 (1995) 93–97.
[4] Tylosine (tartate de) pour usage vétérinaire, Pharmacopée Française, X ed., 1990.
[5] British Veterinary Pharmacopoeia 1985, Tylosin tartrate, H.M. Stationary Office, London, 1985, p. 43.
[6] European Pharmacopoeia, Supplement 1998, Council of Europe, Strasbourg, 3rd ed., 1997, pp. 504–507.
[7] International Organisation for Standardisation (ISO); Accuracy (Trueness and Precision) of Measurement Methods and Results – Part 2: Basic Method for the Determination of Repeatability and Reproducibility of a Standard Measurement Method, International Standard ISO 5725-2:1994(E), 1st ed.

[8] D.L. Massart, B.G.M. Vandeginste, L.M.C. Buydens, S. De Jong, P.J. Lewi, J. Smeyers-Verbeke, Handbook of Chemometrics and Qualimetrics: Part A, Elsevier, Amsterdam, 1997.
[9] R.J. Steffeck, S.L. Woo, R.J. Weigand, J.M. Anderson, LC·GC 13 (1995) 720–726.
[10] F.E. Satterthwaite, Biom. Bull. 2 (1946) 110–114.
[11] G.T. Wernimont, in: W. Spendley (Ed.), Use of Statistics to Develop and Evaluate Analytical Methods, Association of Official Analytical Chemists, Arlington, VA, 1985, pp. 36–43.
[12] D.L. Massart, B.G.M. Vandeginste, S.N. Deming, Y. Michotte, L. Kaufman, Chemometrics – A Textbook, Elsevier, Amsterdam, 1988, p. 47.