ELSEVIER

Journal of Pharmaceutical and Biomedical Analysis
30 (2002) 1197–1206

JOURNAL OF
PHARMACEUTICAL
AND BIOMEDICAL
ANALYSIS

www.elsevier.com/locate/jpba

Derivation of system suitability test limits from a robustness test on an LC assay with complex antibiotic samples

Edelgard Hund, Yvan Vander Heyden, D. Luc Massart, Johanna Smeyers-Verbeke *

ChemoAC, Vrije Universiteit Brussel, Farmaceutisch Instituut, Laarbeeklaan 103, B-1090 Brussels, Belgium

Received 25 March 2002; received in revised form 3 July 2002; accepted 20 July 2002

Abstract

A System Suitability Test (SST) is a test to verify the adequate working of the equipment used for analytical measurements. In pharmaceutical analysis, SSTs are performed at least at the beginning of a series of routine analyses. The most generally applied SST considers the precision of the analysis, i.e. the repeatability standard deviation must not exceed a predefined value. Additionally, a SST can also consider responses indicative for the quality of the technique used, e.g. resolutions between peaks or peak asymmetry in high performance liquid chromatography. The system is then only declared suitable if the response is within given limits. However, it is not always evident how to define the SST limits to be fulfilled for a newly developed method. Robustness tests have been proposed as a starting point in a strategy to deduce these limits. Here, it is examined how such a strategy can be applied for complex samples of microbial origin.
© 2002 Elsevier Science B.V. All rights reserved.

Keywords: System suitability test; System suitability limit; Robustness test; Antibiotic sample

1. Introduction

A System Suitability Test (SST) is a test to control the adequate functioning of an equipment used for analytical measurements. It is performed at the beginning of and sometimes also during a series of routine analyses [1,2]. SSTs are more widespread in the pharmaceutical field than in other branches of analytical chemistry due to stricter regulations. According to the ICH (International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use), 'the SST is an integral part of many analytical procedures' [3].

The most frequently applied SST consists in testing a maximum allowable repeatability standard deviation [4]. To evaluate the short-time stability of the system and to test the precision of the method, the European Pharmacopoeia (EP) defines upper limits for the relative repeatability standard deviation (RSD_{max}) of a peak area, based on three to six injections [5], while the United States Pharmacopoeia (USP) prescribes five injections [6]. In high performance liquid chromato-

* Corresponding author. Fax: +32-2-477-47-35
E-mail address: asmeyers@fabi.vub.ac.be (J. Smeyers-Verbeke).

0731-7085/02/$ - see front matter © 2002 Elsevier Science B.V. All rights reserved.
PII: S0731-7085(02)00470-3

graphy (HPLC), the SST often also comprises limits for resolution [7–9], peak asymmetry [5,7,8], column efficiency [9] and/or retention factors [7,8]. The pharmacopoeias [5,6] in their monographs also specify some SST limits, e.g. a minimum peak resolution. However, for a newly developed HPLC method, the SST criteria have to be defined before the method can be used in analytical routine. The ICH [3] suggests deriving the SST limits from the results of a robustness test. As a part of method validation, robustness tests examine the susceptibility of analytical procedures to small but deliberately introduced changes in the method parameters (factors), which simulate the changes expected when transferring the method between laboratories, operators or instruments [10]. Typical factors examined in HPLC are the pH of a mobile phase buffer, the flow rate, the composition of the mobile phase, the column temperature and the detection wavelength [10]. These factors are studied in an experimental design approach within intervals that at least represent the changes expected during a method transfer.

If the quantitative determination is found to be robust, the ranges observed in the robustness test for the chromatographic responses can be expected to represent a suitable system. Vander Heyden et al. [7] state that the SSTs often fail in laboratories other than the developing one when the criteria are arbitrarily defined, because they do not account for the changes related to the method transfer. Consequently, their strategy to derive SST criteria is based on a robustness test [8]. The example given in Refs. [7,8] considers the separation of three substances from a synthetic drug sample. For that case study the chromatogram always has a similar shape, i.e. the number of impurities remains constant and the application of the strategy is straightforward. However, with more complex samples of microbial origin, such as antibiotics, one is often faced with a considerable variation in composition between samples and thus with changing chromatograms. In this study, SST limits are derived for the RP-HPLC (reversed-phase HPLC) method for tylosin for veterinary use from the EP monograph [11]. Tylosin for veterinary use consists of a mixture of macrolide antibiotics. Besides the main component tylosin A (TA), the related substances tylosin B (TB), tylosin C (TC) and tylosin D (TD) also contribute to the potency of the mixture. The samples generally contain several impurities, of which only some are identified [11,12]. The derivation of the SST limits in this case study started from a robustness test using a Plackett–Burman design [13]. It was evaluated how the strategy of [7,8] can be used for such complex samples.

2. Theory

Information on two-level Plackett–Burman designs can be found in most textbooks on experimental design. A good overview is also given in Ref. [14].

2.1. Calculation of the effects

In a two-level design, the effect of a factor X is calculated according to:

$$E_X = \frac{\sum Y_+}{n_+} - \frac{\sum Y_-}{n_-} \quad (1)$$

with ΣY_+ and ΣY_- the sums of the responses where factor X is at high and low level, respectively; n_+ and n_- the number of runs with factor X at high and low level, respectively.

2.2. Tests for significance of effects

Two methods to determine the significance of the effects were combined: (i) the half-normal probability plot [15], in which the significant effects can be identified as they deviate from the straight line formed by the non-significant ones, and (ii) a statistical method, which derives a critical effect from the distribution of the effects themselves [16]. A significance level α of 0.05 was applied [7,8,16,17].

2.3. System suitability test limits

If the content determination (quantitative aspect of the method) is robust, the most extreme results ('worst cases') that can be expected for the

chromatographic responses (e.g. resolution, asymmetry factor) within the experimental domain of the robustness test should still represent a suitable system [8]. By inversion, if these responses during routine analysis do not exceed the corresponding worst cases, the system should be considered appropriate. Therefore, the SST limits for these chromatographic responses are defined as the worst case results. They are derived from the effects significant at $\alpha = 0.1$, as proposed in Ref. [8]. The upper (lower) worst case $Y_{\text{up/low}}$ is estimated from the linear model:

$$Y_{\text{up/low}} = b_0 + \frac{E_{X_1}}{2} X_1 + \frac{E_{X_2}}{2} X_2 + \cdots + \frac{E_{X_K}}{2} X_K \quad (2)$$

with b_0 the average design result obtained for the response, E_{X_i} the effect of a factor X_i with $E_{X_i} = 0$ for the non-significant effects, and X_i the level (-1 or 1) leading to the worst case result. The robustness guideline [8] proposes either to directly use the predicted results as SST limits or to perform repeated measurements at the worst case conditions and to derive the limits from the respective confidence intervals. Here, the predicted SST limits were determined and evaluated.

If no significant effects are found, Ref. [8] proposes to use the confidence intervals around b_0 as lower and upper worst cases. The standard deviation s of n experiments repeated at least under time-different intermediate precision conditions is then used in the calculation of the SST limits:

$$Y_{\text{low}} = b_0 - t_{0.05,n-1} \frac{s}{\sqrt{n}}$$

$$Y_{\text{up}} = b_0 + t_{0.05,n-1} \frac{s}{\sqrt{n}} \quad (3)$$

where either the mean of the repeated experiments or the mean of the design experiments is used as b_0.

3. Experimental

3.1. Analyte

Three commercial tylosin samples (samples S_A, S_B and S_C) with different composition both in active components and impurities were considered. They were selected in such a way that they provide a good representation of possible compositions. Calibration was performed with an external standard of TA CRS. Samples and standard were provided by Prof. J. Hoogmartens (Katholieke Universiteit Leuven, Belgium). All solutions (standard and three samples) were prepared daily at a concentration of 1 mg/ml in acetonitrile/water 1:1 (v:v).

3.2. Chromatographic conditions and experimental design

The nominal conditions of the RP-HPLC assay are shown in Table 1. In the robustness test, the quantitative factors examined were the flow rate, the detection wavelength (λ), the column temperature, the concentration of $NaClO_4 \cdot H_2O$ (representing the ionic strength μ), the pH of the aqueous fraction of the mobile phase and the ratio of organic and aqueous compounds in the mobile phase. They were studied at two levels symmetrically situated around the nominal one (Table 2). The chromatographic column (a qualitative factor) was also examined at two levels. The robust-

Table 1
Nominal conditions as defined in Ref. [11]

Factor	Nominal level
Chromatographic column	Stainless steel, packed with octadecyl silica gel R, 5 μm diameter of the particles, 200 mm length, 4.6 mm i.d.
Flow rate	1 ml/min
Detection wavelength (λ)	290 nm
Column temperature	35 °C
Mobile phase A	Acetonitrile (ACN)
Mobile phase B	200 g/l $NaClO_4$ in water, pH set to 2.5 with 1 M HCl
Ratio A:B	40:60 (v/v)
Injection volume	20 μl

Table 2
Levels of the quantitative factors examined

Factor code	Factor description	Level	
		−1	1
B	Flow rate (ml/min)	0.9	1.1
C	λ (nm)	287	293
D	Temperature (°C)	32	38
E	A:B	37:63	43:57
F	μ (concentration $NaClO_4$ g/l)	190	210
G	pH (aqueous phase)	2.3	2.7

ness test was based on the Plackett–Burman design given in Table 3.

3.3. Reagents and equipment

Analytical grade $NaClO_4 \cdot H_2O$ (Fluka, Buchs, Switzerland) and milli-Q-purified water (Millipore, Bedford, MA) were used to prepare the aqueous fraction of the mobile phase. The pH was adjusted with 1 M HCl (Merck, Darmstadt, Germany). The organic modifier was HiPerSolv acetonitrile (BDH, Poole, UK).

The EP-method prescribes a column of 200 mm length [11], which is no common column length. As the EP allows a deviation up to ±70% [5], columns of 250 mm were used, namely two Hypersil C18-columns (Alltech, Deerfield, IL) with 5 μm particle size and 4.6 mm internal diameter from different batches (indicated as columns −1 and 1 in Table 3).

A chromatograph from Merck-Hitachi (Tokyo, Japan) was used. It comprised a L-6200 pump, a L-4000 UV detector, a T-6300 column oven, a D-2500 integrator and a Rheodyne (Cotati, CA) 7125 injection system. The D-7000 HPLC integration program (Merck–Hitachi) was used to reintegrate the chromatograms. An Ankersmit A 520 pH meter (Orion, Boston, MA), calibrated with buffer solutions (Merck) at pH 7 and 4, was used for pH measurements. The mobile phase was degassed with a Branson ultrasonic bath (Soest, The Netherlands).

3.4. Sequence of measurements

Each experiment consisted in the sequence of injections shown in Table 4. A bracket-calibration was used. The experiments within a day always started with a complete sequence of injections at nominal conditions, which allows estimating the time-different intermediate precision. They finished with one injection of the standard solution

Table 4
Sequence of injections for each experiment

Injection	Solution	Number of injections
1	Solvent (blank)	1
2	Standard	1
3–4	Sample A	2
5–6	Sample B	2
7–8	Sample C	2
9	Standard	1

Table 3
Two-level Plackett–Burman design for 7 factors requiring 8 experiments

Exp.	Factor						
	A Column	B Flow rate	C λ	D Temp.	E A:B	F μ	G pH
1	1	1	1	−1	1	−1	−1
2	−1	1	1	1	−1	1	−1
3	−1	−1	1	1	1	−1	1
4	1	−1	−1	1	1	1	−1
5	−1	1	−1	−1	1	1	1
6	1	−1	1	−1	−1	1	1
7	1	1	−1	1	−1	−1	1
8	−1	−1	−1	−1	−1	−1	−1

at nominal conditions to check for drift [10]. Duplicate injections of the samples were performed.

A SST for repeatability was performed when a column was used for the first time: the standard solution was injected 5 times at nominal conditions.

3.5. Responses considered

A one-point external standard calibration (based on the bracket injections of the standard) is used to determine the content of TA in the samples. The content of the other active components is also calculated relative to the content of TA in the standard. The EP prescribes minimum contents for TA (80%) and for the sum of TA, TB, TC and TD (95%), which is referred to as total content in the following. Both the individual and the total contents were determined. Furthermore, the peak areas and heights of TA for the samples were taken into consideration. Perpendicular-drop integration was applied to peaks that are not baseline-separated. The retention factors k', the peak resolutions R_S, the peak asymmetry (A_S) for TA and the number of theoretical plates N (based on the peak of TA) are calculated as specified by the EP [5].

4. Results and discussion

4.1. General

For antibiotics, one is often faced with a large variability in the composition of the samples. To account for variations in composition, three samples were selected. Their chromatograms are given in Figs. 1–3. Besides the active components (TA, TB, TC and TD) and the identified impurities DMT and TAD, some unknown components elute, too. The selected samples differ in the amount of TA, TB, TC, TD, TAD and DMT as well as in the number and content of unknown impurities. The unknown substances X_1 and X_2 occurred in all three samples.

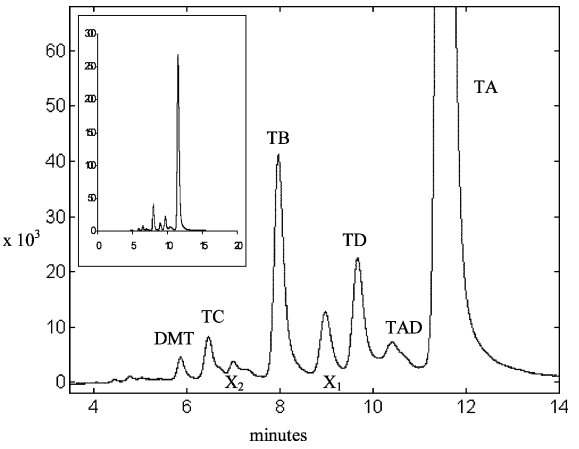

Fig. 1. Chromatogram of sample S_A at nominal conditions (column '−1'; DMT, desmycinosyltylosin; TC, tylosin C; X_2, unknown impurity; TB, tylosin B; X_1, unknown impurity; TD, tylosin D; TAD, tylosin A aldol; TA, tylosin A).

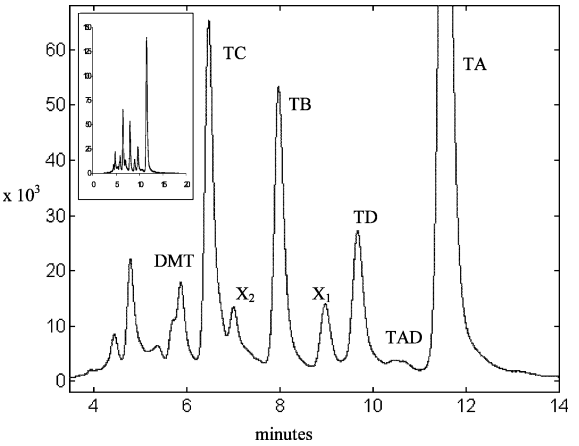

Fig. 2. Chromatogram of sample S_B at nominal conditions (column '−1'; DMT, desmycinosyltylosin; TC, tylosin C; X_2, unknown impurity; TB, tylosin B; X_1, unknown impurity; TD, tylosin D; TAD, tylosin A aldol; TA, tylosin A).

4.2. Robustness test

Prior to the robustness test, the columns were tested for repeatability at nominal conditions. The EP supplement 2001 [5] and Ref. [19] specify upper limits for the relative repeatability standard deviation, RSD_{max}. For HPLC methods, RSD_{max} is 0.59, 0.73 and 0.85% for four, five and six injections, respectively, [18]. Here, five injections of the standard were performed. The injection

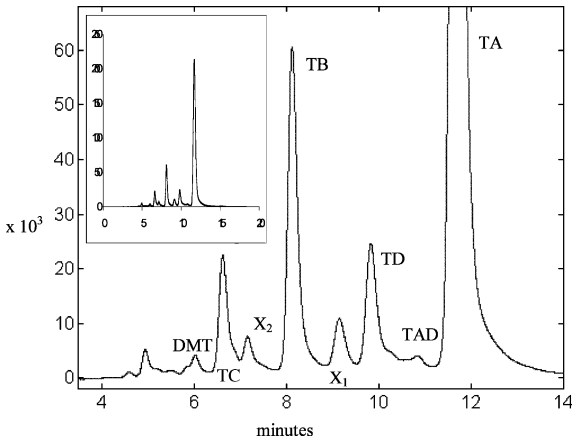

Fig. 3. Chromatogram of sample S_C at nominal conditions (column '−1'; DMT, desmycinosyltylosin; TC, tylosin C; X_2, unknown impurity; TB, tylosin B; X_1, unknown impurity; TD, tylosin D; TAD, tylosin A aldol; TA, tylosin A).

repeatability RSD_r for the peak areas on one column was better than on the other one, namely 0.75% (column 1) versus 1.15% (column −1). For the latter, the peak area obtained in the first injection considerably deviated from the remaining ones. Grubbs' test identified it as an outlier [20]. It is assumed that the system was not sufficiently equilibrated yet, and consequently, the remaining four injections are considered. The new RSD_r, 0.59%, fulfils the requirement. Column +1 is a borderline case for five injections but it was nevertheless used in the study. Thus, for injection repeatability, both systems can be considered as borderline suitable.

The EP [5] prescribes a SST limit for asymmetry factors, which should be between 0.8 and 1.5. The tylosin monograph [11] also contains a limit for a resolution, $R_S(TD-TA) \geqslant 2$. The results for $A_S(TA)$, $k'(TA)$ and $N(TA)$ from repeated injections of the standard are given in Table 5(a), while those for some other responses, estimated from the samples, are shown in Table 5(b). The EP requirements for A_S and $R_S(TD-TA)$ are fulfilled on both columns. Column −1 in general shows a slightly better performance, namely less tailing, a higher number of theoretical plates and a better resolution between the active components.

Table 6 indicates the significant effects observed in the robustness test. The factors column, λ, temperature and μ are not displayed since they had no significant effect on any response. For the chromatographic responses, the effects found significant at 10% level were also significant at 5%. In Table 6, the signs of the effects are given to facilitate chromatographic interpretation.

No significant effect is observed on the main response, the total content, nor on that of the active components TA and TB. Usually, in pharmaceutical analysis, it is expected that an effect is either significant or non-significant for all samples. However, here, some effects were significant only for one or two samples. This can be explained by the difference in composition between the samples. For instance, one observes that in sample S_B, the content of TC is affected by two factors. From the chromatogram (Fig. 2) one can see that the unknown impurity X_2 is eluting in the tail of the TC peak. If factors affect the separation between TC and X_2 (e.g. the ratio A:B), they can also have an influence on the content determination of TC. Moreover, the perpendicular drop integration between TC and X_2, which affects the peak area, could enhance the observed effects for S_B. In addition, a minor impurity elutes between TC and X_2 as can be seen in samples S_A and S_C (Figs. 1 and 3), which complicates the situation further and which can also contribute to the observed effects.

The flow rate was indicated to have a significant effect on the content of TD for sample S_B only. The relative effect of the flow rate on the average content of TD is 10.8, 7.2 and 13.4% for samples S_A, S_B and S_C, respectively, while the relative critical effects were 17.5, 4.9 and 17.2%, respectively. This means that the variability of the TD content determination is smaller for S_B than for the other samples, which explains the statistical significance. The reason for the difference in variability is not immediately clear, but could be related to the differences in composition, as minor components elution between TD and TAD again can affect the area determination.

Several factor effects were significant for all samples (Table 6). The negative effect of the flow rate on the peak area of TA results from a reduced remaining time in the detection cell. The positive effect of the solvent strength on the peak height

Table 5
Average values of some responses obtained at nominal level

Responses	Standard					
	Column					
	−1	1				
(a)						
k' TA	3.55	3.17				
A_S (TA)	1.27	1.50				
N (TA)	8824	6374				
(b)	Sample S_A		Sample S_B		Sample S_C	
	Column		Column		Column	
	−1	1	−1	1	−1	1
k' TC	1.50	1.39	1.51	1.36	1.45	1.33
α(TA/TC)	2.26	2.26	2.27	2.29	2.24	2.28
R_s TC−TB	4.39	4.07	4.22	3.96	4.15	3.98
R_s TB−TD	4.37	3.98	4.20	3.98	4.22	3.95
R_s TD−TA	3.86	3.67	3.86	3.81	3.84	3.64
R_s TC−X_2	1.86	1.83	1.68	1.81	1.66	1.75

(a), for the standard from the injection repeatability experiments ($n = 5$); (b), for the three samples from the nominal experiments performed between the design experiments ($n = 4$).

Table 6
Indication of significant effects for samples S_A, S_B and S_C (+, positive effect; −, negative effect)

Responses	Factors		
	Flow rate	A:B	PH
Total content			
Content TA			
Content TC	S_B (+)	S_B, S_C (+)	S_C (+)
Content TB			
Content TD	S_B (+)		
Area TA	S_A, S_B, S_C (−)		
Height TA		S_A, S_B, S_C (+)	
N (TA)			
As (TA)			
k' TC		S_A, S_B, S_C (−)	
k' TB		S_A, S_B, S_C (−)	
k' TD		S_A, S_B, S_C (−)	
k' TA		S_A, S_B, S_C (−)	
α(TA/TC)			
R_s TC−TB		S_A, S_B, S_C (−)	
R_s TB−TD		S_A, S_B, S_C (−)	
R_s TD−TA		S_A, S_B, S_C (−)	

Not shown: factors column, λ, temperature and μ (no significant effects).

(not on the peak area) of TA is caused by an accelerated elution, which leads to higher peaks. For the same reason, this factor also has an effect on the retention factors and the peak resolutions.

Besides the significance of the effects, the variability of the design results can also give an indication of the robustness of the method. Table 7 shows the %RSD for the content of the main component TA and for the total content both from nominal and design experiments. One expects the variability of the design experiments to be considerably larger than the one of the nominal experiments in case of non-robustness. From Table 7, it can be observed that the %RSD values are either comparable or slightly higher in the design experiments. However, the latter can be expected even for a robust method since the variability in the design conditions is larger than in the time-different experiments.

4.3. System suitability test limits

Although the significant effects observed on the contents of TC and of TD in sample S_B indicate

Table 7
%RSD of the content of TA and of the total content for the different samples

	Sample S_A	Sample S_B	Sample S_C
Nominal experiments			
Content TA			
Both columns	2.96	4.00	3.04
Column −1	2.03	5.22	3.48
Column 1	2.77	2.95	2.96
Total content			
Both columns	2.94	2.48	1.58
Column −1	1.53	3.52	0.29
Column 1	1.24	0.60	1.76
Design experiments			
Content TA			
Both columns	3.16	4.69	3.00
Column −1	1.14	5.98	2.84
Column 1	2.63	3.85	3.57
Total content			
Both columns	2.81	3.95	1.17
Column −1	1.55	6.24	0.24
Column 1	1.50	1.25	1.56

that their determination is not always robust, the system suitability criteria were derived according to [8]. For responses subject to significant effects, the predicted worst case results were used as SST limits. For those without significant effects, the standard deviation of the nominal experiments performed on four measurement days was used in Eq. (3). Since the columns are similar (same manufacturer) and have a comparable performance (Table 5), the between-day average on the 'worst' column was used as b_0.

Table 8 shows the limits individually deduced for the three samples. The responses are those typically considered in SSTs. For the retention factors k', only the limits for the first and the last active component were considered. The acceptable ranges for these responses overlap, which from a practical point of view does not give a useful indication of the suitability of the system. Therefore, the definition of a SST limit for the separation factor α(TA/TC) seems more appropriate.

Comparison with the EP limits shows that the SST limits for the asymmetry factor derived from the robustness test are similar to the EP limit: the limits derived range between 1.50 and 1.63, while EP prescribes an upper limit of 1.50. However, a considerably better separation between TD and TA is required by the new limits: the SST limit for resolution is at least 3.4, whereas EP only requires 2.0.

Table 8
System suitability test limits

Response	Sample S_A	Sample S_B	Sample S_C
A_s (TA)[a]	1.54	1.50	1.63
N (TA)[b]	5545	5521	5821
k' TC[c]	[0.91; 2.54]	[0.92; 2.53]	[0.93; 2.55]
k' TA[c]	[2.07; 5.79]	[2.15; 5.78]	[2.09; 5.82]
α(TA/TC)[b]	2.25	2.25	2.23
R_s TC-TB[b]	3.77	3.74	3.77
R_s TB-TD[b]	2.90	2.93	2.87
R_s TD-TA[b]	3.54	3.59	3.48
R_s TC-X_2[b]	2.0[d]	2.0[d]	2.0[d]

[a] Upper limit.
[b] Lower limit.
[c] Upper and lower limit.
[d] SST limits defined according to FDA recommendations, independent from robustness test results.

To assess the practical relevance of the SST limits derived according to the strategy of Ref. [8], they should now be critically evaluated. For that purpose, the limits derived in Table 8 are compared with the most extreme results of the earlier performed SST and nominal experiments (Table 9). They should not violate the SST limits of Table 8 unless their definition was too strict. First, the most extreme results and the SST limits for the individual samples are considered. Usually, the most extreme results are within the SST specifications. However, some borderline cases can be observed, e.g. for R_S(TD−TA) and α(TA/TC). This suggests that the method applied [8] can lead to very strict SST limits, especially if no significant effects are observed and if the normal variation of a response is rather small. As no significant violations are observed, the derivation of the SST limits can be considered acceptable. A possible less strict alternative in those situations could be to use the most extreme design result observed. However, this is only allowed—as in fact is the case for the derivation of any SST limit—for

Table 9
Most extreme results in the SST and nominal experiments

Response	Standard (SST)		Sample S_A		Sample S_B		Sample S_C	
	Column		Column		Column		Column	
	−1	1	−1	1	−1	1	−1	1
A_s (TA)[a]	1.40	1.52	1.52	1.42	1.38	1.24	1.46	1.31
N (TA)[b]	8750	6126	7318	6030	7351	6009	6541	6243
k' TC[b]			1.61	1.48	1.61	1.47	1.61	1.47
k' TC[a]			1.61	1.50	1.61	1.48	1.67	1.48
k' TA[b]	3.44	3.15	3.63	3.36	3.64	3.34	3.64	3.34
k' TA[a]	3.69	3.21	3.64	3.37	3.65	3.41	3.70	3.39
α(TA/TC)[b]			2.26	2.24	2.27	2.28	2.22	2.27
R_s TC-TB[b]			4.34	4.02	4.21	3.95	4.13	3.97
R_s TB-TD[b]			4.30	3.66	4.20	3.63	4.22	3.64
R_s TD-TA[b]			3.84	3.51	3.86	3.66	3.69	3.59
R_s TC-X_2[b]			1.81	1.82	1.67	1.80	1.62	1.74

[a] Largest value.
[b] Smallest value.

responses related to peaks for which the content determination is not affected.

Secondly, it is also checked whether some of the most extreme results violate the limits derived with the other samples. In general, this is not the case, which could be expected since the SST limits derived from the different samples are rather comparable. As in the previous situation some borderline situations are encountered, e.g. for A_S and the responses already mentioned. To select a less strict SST limit, the same remark as above is valid. Thus, although the samples show considerable differences in their composition, the strategy proposed in Ref. [8] leads to SST limits that are rather independent of the sample considered. Therefore, the study of different samples does not seem to be required, even for these rather complex antibiotic samples with varying composition. A single representative sample seems to be sufficient. However, if one intends to perform two injections per experiment in the robustness test, it is certainly preferable to examine two different samples instead of replicated injections of one sample.

The resolution $R_S(TC-X_2)$ was also considered as SST response, since the observations for samples S_B and S_C led to the suspicion that an inadequate separation of these peaks could contribute to the problems in the content determination for TC. Notice however, that no significant effects on $R_S(TC-X_2)$ were found in the robustness test. Although the SST limits derived according to Ref. [8] were larger than 1.5 (i.e. around 1.6), the minimum requirement for baseline separation [5], it is obvious that the requirements set are not sufficient. It is also probable that a minor impurity, seen as a shoulder in the peak of TC in Fig. 1 and Fig. 3, is partly responsible for the problems. To account for these problems, a possible alternative here is to define a SST limit independently from the robustness result. The Food and Drug Administration (FDA) Center for Drug Evaluation and Research (CDER) also generally recommends a peak resolution larger than 2.0 'between the peak of interest and the closest potential interfering peak' [21]. For the separation between TD and TAD, this recommendation could be followed, too, since there similar problems as for $TC-X_2$ were observed.

However, the above suggests that slight modifications in the mobile phase composition might be required prior to the start of an analysis.

5. Conclusions

The case study showed that the strategy proposed in Ref. [8] can also be applied for complex samples. If the SST limits derived are less strict than those prescribed by the pharmacopoeias, the use of the latter is recommended. When the limits according to Ref. [8] are quite strict, e.g. when no significant effects are observed on a response, the use of the most extreme design experiments can be a somewhat less strict alternative for peaks with a robust content determination.

For the drug substance considered, tylosin, three samples with different composition were analysed. Comparable SST limits were derived from all three samples. Consequently, it would be sufficient to consider only one representative sample in the robustness test to derive SST limits. If the content determination of one of the major components is affected, it might be necessary to define SST limits independently from the robustness test, which require a given minimum separation between well-defined peaks. These limits are then stricter than those derived from the robustness test.

Acknowledgements

The authors thank Prof. J. Hoogmartens, Katholieke Universiteit Leuven, Belgium, for providing tylosin standard and samples, and Mrs V. Reynders for technical assistance. This work was supported by the Research contract No. NM/03/24 of the Belgian government (The Prime Ministers Service—Federal Office for Scientific, Technical and Cultural Affairs, Standardisation Program). Y. Vander Heyden is a postdoctoral fellow of the Fund for Scientific Research—Flanders (Belgium) (F.W.O.)

References

[1] R.H. King, L.T. Grady, J.T. Reamer, J. Pharm. Sci. 63 (1974) 1591–1596.
[2] M. Dong, R. Paul, L. Gershanov, Today's Chemist at Work, 2001, pp. 38–42.
[3] ICH Topic Q2B Validation of analytical procedures: Methodology. Step 4, consensus guideline, the International conference on Harmonisation of Technical Requirements for registration of Pharmaceuticals for human use, November 6, 1996.
[4] T.D. Wilson, J. Pharm. Biomed. Anal. 8 (1990) 389–400.
[5] European Pharmacopoeia, third ed., Supplement 2001, European Pharmacopoeia, Strasbourg, 2001, pp. 30–33.
[6] The United States Pharmacopoeia, 23rd ed., National Formulary 18, United States Pharmacopoeial convention, Rockville, MD, 1995, pp. 1566.
[7] Y. Vander Heyden, M. Jimidar, E. Hund, N. Niemeijer, R. Peeters, J. Smeyers-Verbeke, D.L. Massart, J. Hoogmartens, J. Chromatogr. A 845 (1999) 145–154.
[8] Y. Vander Heyden, A. Nijhuis, J. Smeyers-Verbeke, B.G.M. Vandeginste, D.L. Massart, J. Pharm. Biomed. Anal. 24 (2001) 723–753.
[9] A. Chalgeri, H.S.I. Tan, J. Pharm. Biomed. Anal. 14 (1996) 835–844.
[10] Y. Vander Heyden, F. Questier, D.L. Massart, J. Pharm. Biomed. Anal. 17 (1998) 153–168.
[11] European Pharmacopoeia, third ed., Supplement 1998, European Pharmacopoeia, Strasbourg, 1998, pp. 504–505.
[12] J. Paesen, W. Cypers, K. Pauwels, E. Roets, J. Hoogmartens, J. Pharm. Biomed. Anal. 13 (1995) 1153–1159.
[13] R.L. Plackett, J.P. Burman, Biometrika 33 (1946) 305–325.
[14] Y. Vander Heyden, D.I. Massart, in: M.W.B. Hendriks, J.H. de Boer, A.K. Smilde (Eds.), Robustness of Analytical Chemical Methods and Pharmaceutical Technological Products, Elsevier, Amsterdam, 1996, pp. 79–147.
[15] C. Daniel, Technometrics 1 (1959) 311–341.
[16] A. Nijhuis, H.C.M. van der Knaap, S. de Jong, B.G.M. Vandeginste, Anal. Chim. Acta 391 (1999) 187–202.
[17] E. Hund, Y. Vander Heyden, M. Haustein, D.L. Massart, J. Smeyers-Verbeke, Anal. Chim. Acta 404 (2000) 257–271.
[18] European Pharmacopoeia Commission: System Suitability Criterion for Repeatability, non-published document PA/PH/Exp. MG/T (01) 1, 2001.
[19] S. Ebel, Latest developments of general separation methods, lecture given at the Conference 'The Future Face of the European Pharmacopoeia—Current Concerns in Pharmaceutical Analysis', 8–9, February 2001, Cannes, France.
[20] 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, pp. 109–114.
[21] Center for Drug Evaluation and Research (CDER): Reviewer Guidance, Validation of Chromatographic Methods, FDA, Rockville, MD, November 1994, available from http://www.fda.gov/cder/guidance/cmc3.pdf, as given in December 2001.