ELSEVIER

Journal of Pharmaceutical and Biomedical Analysis
20 (1999) 875–887

JOURNAL OF
PHARMACEUTICAL
AND BIOMEDICAL
ANALYSIS

www.elsevier.com/locate/jpba

Nested designs in ruggedness testing

Y. Vander Heyden [a], K. De Braekeleer [a], Y. Zhu [b], E. Roets [b], J. Hoogmartens [b], J. De Beer [c], D.L. Massart [a,*]

[a] *ChemoAC, Pharmaceutical Institute, Vrije Universiteit Brussel, Laarbeeklaan 103, 1090 Brussels, Belgium*
[b] *Laboratory for Pharmaceutical Chemistry and Drug Analysis, Katholieke Universiteit Leuven, Van Evenstraat 4, 3000 Leuven, Belgium*
[c] *Wetenschappelijk Instituut Volksgezondheid—Louis Pasteur, Juliette Wytmanstraat 14, 1050 Brussels, Belgium*

Received 8 August 1998; received in revised form 8 December 1998; accepted 1 January 1999

Abstract

Nested designs were performed in order to execute a ruggedness test according to the United States Pharmacopeia definition for ruggedness, in which mainly non-procedure related factors are examined. Several nested designs have been executed on a high performance liquid chromatography assay to determine tetracycline and related substances in bulk samples of tetracycline. Factors such as different laboratories, analysts, instruments, columns, days and batches were examined. The interpretation methods described in the literature were found to cause problems. In these methods the variances of the examined factors are estimated from the calculated mean square values and from the equation for the expected mean squares. Very frequently, negative variance estimates were obtained. Their absolute values were found to be dependent on the influence of the factor examined below it in the design, on the examined response. Therefore an alternative interpretation method for nested designs, based on pooled variances, was proposed and found to be appropriate to use for ruggedness testing purposes. Both approaches, the one from the literature and the one proposed here, were tested on simulated data coming from a nested design with four factors and on the experimentally measured data.

Keywords: Nested designs; Ruggedness test

1. Introduction

In the United States Pharmacopeia (USP) [1,2] ruggedness is defined as "The degree of reproducibility of test results obtained by the analysis of the same sample under a variety of normal test conditions, such as different laboratories, different analysts, different instruments, different lots of reagents, different elapsed assay times, different assay temperatures, different days, etc." This definition is different from that originally introduced by Youden and Steiner [3] under the term ruggedness, which in analytical chemistry is mainly called robustness and in which one evaluates the influence of small changes in the operating conditions. The use of two-level screening designs such

* Corresponding author. Tel.: +32-2-477-4737; fax: +32-2-477-4735.
E-mail address: fabi@vub.vub.ac.be (D.L. Massart)

0731-7085/99/$ - see front matter
PII: S0731-7085(99)00112-0

as Plackett–Burman [4] and fractional factorial designs [5], that are generally applied in robustness testing, is impossible when testing ruggedness as defined by the USP because impossible factor and level combinations could be required in these designs and the evaluation of a factor at more than two levels is recommended [6,7].

Nested designs [8,9] could then be an alternative and their use and analysis, called *nested analysis of variance* (nested ANOVA), are examined here. An example of a nested design is given in Fig. 1. The design is called nested because the subordinate classification is nested within the higher classification level. It is said that the levels of a factor (e.g. the analysts) are nested within the levels of another factor (laboratories in Fig. 1) if every level of the first factor appears within only one level of the second. This means, for instance, that each analyst appears only in one of the laboratories.

Nested designs can, according to the ISO (International Organisation for Standardisation) guidelines [9], be used in method validation for the analysis of intermediate precision estimates. The potential factors to be examined in the nested designs are then time, calibration, operator and equipment. These factors, together with other non-procedure related factors, as for instance batches and manufacturers of reagents, laboratories, and factors related to chromatographic columns (see USP definition) could also be considered in a ruggedness test set up following the USP definition. The factors examined in this kind of ruggedness test are mainly qualitative which explains why preferably they are examined at more than two levels.

In this study several nested designs were performed on the high performance liquid chromatography (HPLC) assay of the USP XXII for tetracycline HCl [1]. Only fully-nested (i.e. not staggered-nested designs) were examined [9]. The former ones are designs similar to the one shown in Fig. 1 while the latter ones are analogous designs but with some branches missing. The interpretations for nested designs, as described in the literature, were considered. We studied which interpretations were appropriate to make comparisons for the results of similar responses, both within (e.g. measured on different peaks) and between designs (e.g. on the same peak but in different designs). In analogy with the normalised effects in screening designs [10,11], the required interpretation criterion should for instance give comparable values when a factor is examined in different designs, regardless of the other factors examined in those designs.

2. Theory

An example of a nested design is given in Fig. 1. For the interpretation of such a design, an ANOVA table is created, as is shown in Table 1 for the design of Fig. 1. Details about the calculation of the different terms can be found in Refs. [8,9]. In Fig. 1 the symbols a,b,c and n represent

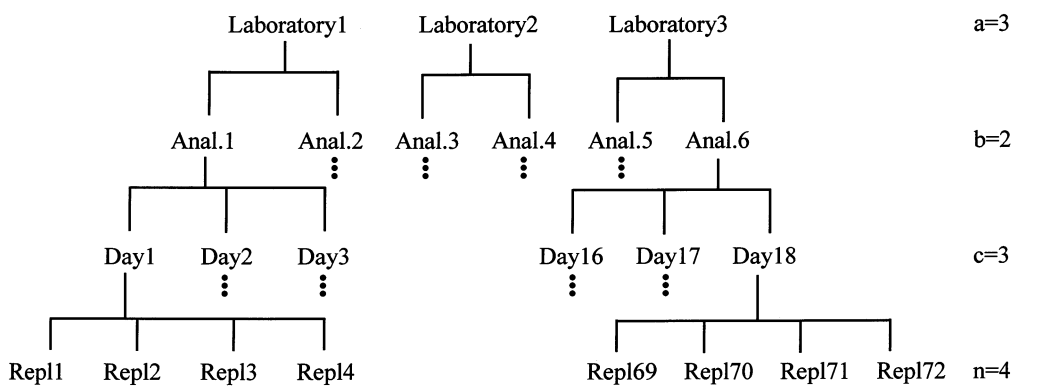

Fig. 1. Hypothetical design for a nested ANOVA in which the factors laboratories, analysts, days and replicates are examined.

Table 1
ANOVA table for the nested design of Fig. 1[a]

Source of variation	df	*SS*	*MS*	*F*-value	Critical *F*-value ($\alpha = 0.05$)
Among laboratories	$a-1$	SS_{lab}	MS_{lab}	$\frac{MS_{lab}}{MS_{anal}}$	$F_{[a-1,a(b-1)]}$
Analysts within laboratories	$a(b-1)$	SS_{anal}	MS_{anal}	$\frac{MS_{anal}}{MS_{day}}$	$F_{[a(b-1),ab(c-1)]}$
Days within analysts	$ab(c-1)$	SS_{day}	MS_{day}	$\frac{MS_{day}}{MS_{repl}}$	$F_{[ab(c-1),abc(n-1)]}$
Replicates within days	$abc(n-1)$	SS_{repl}	MS_{repl}		
Total	$abcn-1$	SS_{total}	MS_{total}		
	Expected MS				
Among laboratories	$s^2_{repl}+n.s^2_{day}+n.c.s^2_{anal}+n.c.b.s^2_{lab}$				
Analysts within laboratories	$s^2_{repl}+n.s^2_{day}+n.c.s^2_{anal}$				
Days within analysts	$s^2_{repl}+n.s^2_{day}$				
Replicates within days	s^2_{repl}				

[a] df, degrees of freedom; *SS*, sum of squares; *MS*, mean square; α, significance level; s^2_i, variance of a factor *i*.

the number of levels of a factor nested within the above ranked factor. In the following the interpretations of nested designs as described in the literature [8,9] are given. These interpretations are based on *F*-tests, estimation of the variance of each factor and expressing these variances relative to the sum of all variances.

A first interpretation criterion is the use of an *F*-test in which mean square values, which are in fact variances, are compared. The *F*-values for the factors are obtained by dividing the mean square (*MS*) of a factor with the *MS* exactly below it in the table. In these tests it is verified if a factor situated at a higher level in the hierarchical structure of the design causes a significant increase in the variance compared to the variance caused by all factors situated below.

In a second approach the variances of the different factors (s^2_i) are estimated. From the estimated mean square (*MS*) values and the formulas for the expected mean square (Table 1), one can estimate the different variance components. For the example of Fig. 1 this gives:

$$MS_{repl} = s^2_{repl} \quad (1)$$

$$MS_{days} = s^2_{repl} + n.s^2_{days} \rightarrow s^2_{days} = \frac{MS_{days} - MS_{repl}}{n} \quad (2)$$

$$MS_{anal} = s^2_{repl} + n.s^2_{days} + n.c.s^2_{anal} \rightarrow s^2_{anal} = \frac{MS_{anal} - MS_{days}}{n.c} \quad (3)$$

$$MS_{lab} = s^2_{repl} + n.s^2_{days} + n.c.s^2_{anal} + n.c.b.s^2_{lab} \rightarrow s^2_{lab} = \frac{MS_{lab} - MS_{anal}}{n.c.b} \quad (4)$$

A third approach is the use of the relative magnitude of the variance components. For instance, one determines

$$\frac{s^2_{lab} \times 100}{S^2}\%$$

with $S^2 = s^2_{repl} + s^2_{days} + s^2_{anal} + s^2_{lab}$.

In ruggedness testing, we consider the variance estimated for each factor as most interesting. In analogy with the effects [12] calculated from screening designs (which can also be expressed as variances [5]), these variances give information about the influence of a factor on a response. The knowledge of these influences allows one to control or standardize the most important factors by including a 'precautionary statement' [13] in the method description, in case a method is found not to be robust.

The ISO guidelines [9] require, in another context (intermediate precision estimation), that factors most affected by systematic effects should be arranged in the highest ranks of the hierarchy and those affected most by random effects should be

in the lowest ranks of the design. This may not seem important in the case of fully-nested experiments due to its symmetry [9]. The lowest factor is considered to have a residual variation.

3. Experimental

3.1. Substances

A tetracycline HCl (TC) standard with a purity of 98.2% and a tetracycline sample also containing 3% 4-epitetracycline HCl (ETC), 2% 4-epianhydrotetracycline HCl (EATC), 0.9% 2-acetyl-2-decarboxamidotetracycline HCl (ADTC) and 3% anhydrotetracycline HCl (ATC) were used. The pure substances were obtained from Acros Chimica (Beerse, Belgium). Potassium nitrate, ammonium oxalate, dibasic ammonium phosphate and *N*,*N*-dimethylformamide (DMF) of GR (pro analysis) quality, as well as dimethylformamide of HPLC quality were used.

3.2. Laboratories and instruments

The participating laboratories were (a) Laboratory for Pharmaceutical and Biomedical Analysis, Pharmaceutical Institute, Vrije Universiteit Brussel, Belgium; (b) Laboratory for Pharmaceutical Chemistry and Drug Analysis, Institute for Pharmaceutical Sciences, Katholieke Universiteit Leuven, Belgium, and (c) Wetenschappelijk Instituut Volksgezondheid—Louis Pasteur, Brussels, Belgium. The HPLC instruments used in lab (a) were (1) a Varian model 500 Liquid Chromatograph (Varian, Palo Alto, CA) with a Perkin-Elmer LC 90 UV spectrophotometric detector (Perkin-Elmer, Norwalk, CT) and a Merck Hitachi D-2000 Chromato-integrator (Darmstadt, Germany), and (2) a Merck Hitachi L-6000A pump with a Merck Hitachi L-4200 UV–VIS detector and a Merck Hitachi D2000 Chromato-integrator. The injectors were from Rheodyne (Cotati, CA). In lab (b) instrument (1) consisted of a L-6200 pump Merck-Hitachi with a Valco model CV-6-UHPa-N60 (Houston, TX) injector, a Model 441 (Waters, Milford, MA) detector and a HP 3396 Series II (Hewlett-Packard, PA) integrator, and instrument (2) of a Model 6000A (Waters, Milford, MA) pump, a Valco injector model CV-6-UHPa-N60 (Houston, TX) injector, a L-4000 UV (Merck-Hitachi) detector and a HP 3396 A (Hewlett-Packard, PA) integrator. In lab (c) instrument (1) consisted of a Waters 625 LC System with a Waters 990 PAD detector and a Marathon Autosampler (Spark Holland, Emmen, The Netherlands), while instrument (2) was a Waters 600-MS pump, a Waters 996 PAD detector, a Waters 717 Plus autosampler and the Waters Millenium Software. The injection volume was 20 μl and detection was performed at 280 nm on each of the instruments.

3.3. Chromatographic conditions

The columns were all Alltima (Alltech, Deerfield, IL) C-8, 5 μm, 25 cm × 4.6 mm i.d. columns from the same batch. Within each laboratory four columns were used. The mobile phase consisted of 0.1 M ammonium oxalate, dimethylformamide and 0.2 M dibasic ammonium phosphate (680:270:50, V/V/V). The pH was adjusted to 7.65 with 3 N ammonium hydroxide or with 3 N phosphoric acid. The mobile phase was filtered through a 0.5-μm membrane filter or finer porosity and sonicated prior to use. The flow rate of the mobile phase was 1 ml/min.

3.4. Standard and sample solutions

A standard solution with concentration 0.5 mg/ml was prepared by dissolving the TC standard in a solvent consisting of 0.1 M ammonium oxalate and DMF (680:270, v/v). Potassium nitrate (1 mg/ml) was added to the solvent in order to determine the dead time. A sample solution, also with concentration 0.5 mg/ml, was prepared by dissolving the TC sample mixture in solvent. Daily one standard and two sample solutions were prepared and, between injections, stored in the dark at room temperature. These solutions are injected in both instruments under the conditions required for the different designs.

3.5. Designs

Several nested designs have been performed (Fig. 2). In Fig. 2a the influence of the factors analysts ($a = 3$), instruments ($b = 2$), columns ($c = 2$), days ($d = 3$) and replicates ($n = 2$) was examined. The experiments were performed within the same laboratory by three analysts. Four columns were used that were randomly distributed within analysts and instruments but in such a way that each analyst used each column once. In the design of Fig. 2b, the factors manufacturer of DMF ($a = 2$), instruments ($b = 2$), columns ($c = 2$), days ($d = 3$) and replicates ($n = 2$) were included. The design of Fig. 2c contains the factors laboratories ($a = 3$), instruments ($b = 2$), columns ($c = 2$), days ($d = 3$) and replicates ($n = 2$).

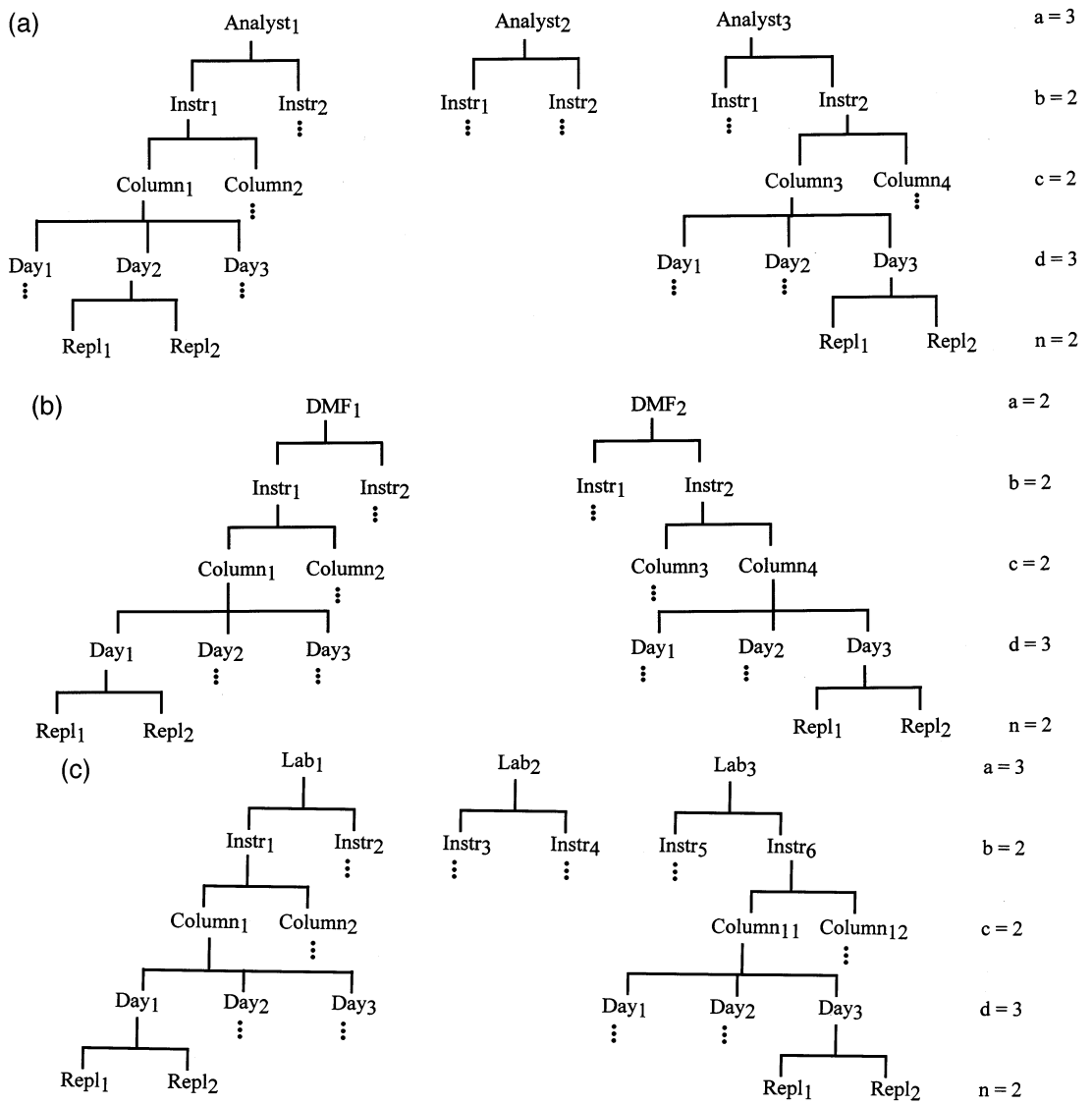

Fig. 2. The different executed nested designs.

3.6. Measured responses

For each design experiment, the standard and the two sample solutions were injected. Several responses were determined in each design experiment such as the retention times, peak areas and peak heights, capacity factors, relative retentions (with the ETC peak as reference peak), resolutions between ETC-EATC and EATC-TC, the peak widths at half height, and the contents of TC and degradation products in the sample solution. The content of TC is expressed as the percentage TC occurring in the sample, while the contents of the degradation products are expressed as fractions of the TC peak areas or heights.

The nested design interpretations and the influences of the examined factors on these responses were determined using Microsoft Excel 4.0 spreadsheets and the statistical software package Statgraphics® Plus 6 [14].

4. Results and discussion

4.1. Description of the problem

A nested design set up to perform a ruggedness test will not always fulfill all requirements specified in the literature because, for practical reasons, it is often not feasible. The requirement that in a nested design each level of a factor occurs only once in the level of another factor, for instance, is not always fulfilled since it would require an unreasonable number of columns and instruments. Therefore often a design is created which is situated between a pure nested design and a pure multi-way ANOVA design. This can also be observed in the designs we executed (Fig. 2).

Another requirement [9], namely that the factors affected most by systematic effects should be ranked highest in the nested hierarchy and those affected most by random effects in the lowest ranks, is not always fulfilled either since one does not always know in advance which are these factors. Otherwise a ruggedness test would not be necessary. A design, where one can presume that lower ranked factors have a higher influence on a number of responses than higher ones, is that of Fig. 2b, where it can be expected that the factor instruments will affect some responses to a larger extent than the factor manufacturer of DMF.

Moreover, since different responses are measured for the experiments, the same factor will not always be the most important for the different responses. This means that it can happen that for one response a higher ranked factor will be affected most by systematic effects, while for another it will be a lower ranked one.

The interpretation methods given in Section 2 (F-test, estimating individual variances and expressing relative magnitude of variances) were tried out for the different responses. The interpretation approach in which the variances of the different factors are estimated is discussed below. The two other approaches, i.e. the determination of the relative variance components and the F-tests, are not discussed since their results were found to be of minor interest in the context of ruggedness testing, as was already expected.

In earlier presented work on screening designs applied in robustness testing [10], it was found that the estimated effects of a factor, examined at the same levels in different screening designs, are similar. It was studied in this work if the variances estimated for a factor examined in different nested designs at a given number of levels also are similar. This means that it is verified if the variance estimates are similar for the same set of, for instance, chromatographic columns examined in different nested designs and occasionally situated at different ranks in those designs. However, from our experiments, very frequently negative variance estimates for some examined factors were obtained. The results for the variances calculated on a response measured in the design of Fig. 2c are given in Table 2. In this table, the MS values for the different factors are also shown. The calculated MS values are expected to increase from the lower to the upper level in the design (see formulas for expected MS in Table 1 and Eqs. (1)–(4)), because, when going to a higher level, a positive term is added to the MS of the lower level (Table 1), so that the variances calculated with equations similar to Eqs. (1)–(4), are expected to be positive.

Table 2
Variances and *MS*-values for the design of Fig. 2c, estimated for some responses[a]

Source of variation	Variance on		MS-values	
	k' (EATC)	k' (TC)	k' (EATC)	k' (TC)
Laboratories	0.0114	0.0376	0.2944	0.9433
Instruments	**−0.0129***	**−0.0284***	0.0205	0.0404
Columns	0.0261	0.0530	0.1751	0.3806
Days	0.0091	0.0310	0.0185	0.0624
Replicates	0.0004	0.0004	0.0004	0.0004

[a] k', capacity factor.
* Negative variance estimates.

It was observed that the above described requirements for *MS* and the variances often were not fulfilled. If a factor has a considerably smaller influence on a response than the one situated below it in the design, the *MS* calculated for it drops to a low value (Table 2). As a consequence the variance estimated for this factor is negative (Table 2). This can be observed in our example of Table 2, where the columns cause a much larger variation in the measured capacity factors than the instruments. Some statistical software packages (e.g. Statgraphics® Plus 6 [14]) systematically display each negative variance as zero. However, in our opinion, the variance of such a factor is not necessarily zero or negligible, as we will try to demonstrate further.

4.2. Alternative interpretation method

To avoid such problems an alternative interpretation method was proposed. It is based on pooled variances. For the design of Fig. 1, the pooled variance of the replicates was calculated by pooling all variances for replicates within days. Consecutively the average results for each day are calculated. The variance component of the days is then calculated as the pooled variance of those averages within analysts. The variances for analysts and laboratories are calculated in a similar way. By applying this method no negative variances can be obtained. When pooling the variances it is presumed that they are equal. However, given the low number of degrees of freedom with which a particular variance usually is estimated in a nested design, occasional differences are anyway difficult to demonstrate.

The idea to introduce this interpretation was based on the calculation of effects in screening designs. These effects are calculated in such a way that when the effect of a factor is calculated, the influence of all other factors is canceled. In the nested designs, we tried to obtain a similar phenomenon by evaluating the pooled variance at the lowest rank, then eliminate its influence by taking the average results within the different levels of the above rank and consecutively evaluate the influence (variance) of the second lowest rank, etc.

The same results as with the pooled variances are obtained from the ANOVA table by dividing the *MS* of a factor with the number of replicates within this factor. For the design of Fig. 1 this gives:

$$s^2_{\text{repl}} = MS_{\text{repl}} \tag{5}$$

$$s^2_{\text{days}} = \frac{MS_{\text{days}}}{\text{n}} = \frac{MS_{\text{days}}}{4} \tag{6}$$

$$s^2_{\text{anal}} = \frac{MS_{\text{anal}}}{\text{n.c}} = \frac{MS_{\text{anal}}}{12} \tag{7}$$

$$s^2_{\text{lab}} = \frac{MS_{\text{lab}}}{\text{n.c.b}} = \frac{MS_{\text{lab}}}{24} \tag{8}$$

4.3. Analysis of simulated data

For a given design with a given grand mean and certain given variances for the different factors (Table 3), experimental results were simulated. Four factors, A, B, C and N (Fig. 3) were

examined in the design and the number of levels were $a=3$, $b=2$, $c=3$ and $n=2$ respectively. The simulated data were then analysed applying both interpretation methods. The results are shown in Table 3. In Table 3(a) it can be seen that the method based on the pooled variances gives an estimate for the variances of the factors that is similar to the one originally entered, which is not the case for the literature method. When increasing the variances of factors B and N (Table 3(b)) the literature method yields a variance estimate for the factor above it in the hierarchy which is too low or even becomes negative, which is not the case for the pooled variance method.

As already mentioned earlier, not always all requirements for nested designs are fulfilled. Suppose, for instance, that the factors from Fig. 3 are instruments (A), analysts (B), columns (C) and replicates (N) and that three instruments, two analysts and six columns were available (Fig. 4). In this example one has an intermediate situation between a nested and a multiway ANOVA. Both designs in Fig. 4 contain the same factor combinations, i.e. they require the same experiments.

Table 3
Variance estimates for some simulated design results[a]

	Source of variance	Entered variances	Estimation of the variance components	
			Literature method	Pooled variances method
(a)	Factor A	21.13	12.22	21.16
	Factor B	17.83	15.54	17.88
	Factor C	7.00	1.78	7.02
	Factor N	10.46	10.49	10.49
(b)	Factor A	21.13	0.73	21.16
	Factor B	40.80	38.52	40.86
	Factor C	7.00	−3.22	7.02
	Factor N	20.46	20.48	20.48
(c)	Factor B	17.83	10.83	17.88
	Factor A	21.13	18.82	21.16
	Factor C	7.00	1.78	7.02
	Factor N	10.46	10.49	10.49
(d)	Factor C	7.00	−1.92	7.02
	Factor B	17.83	10.83	17.88
	Factor A	21.13	15.92	21.16
	Factor N	10.46	10.49	10.49

[a] The grand mean of the designs was always defined to be 100.

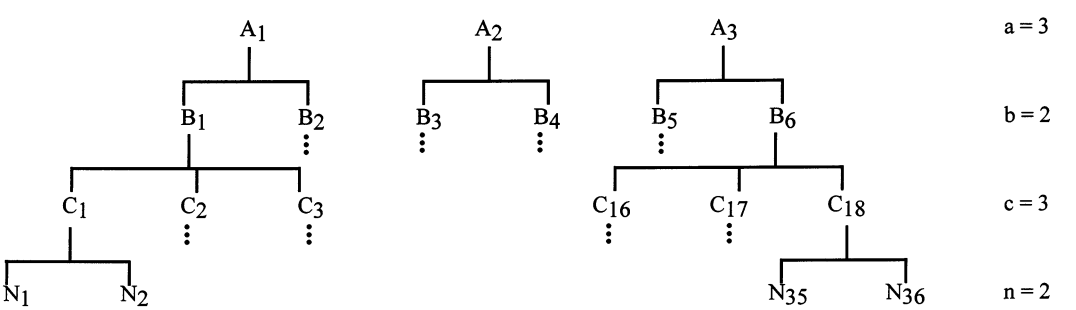

Fig. 3. Design used in the simulations.

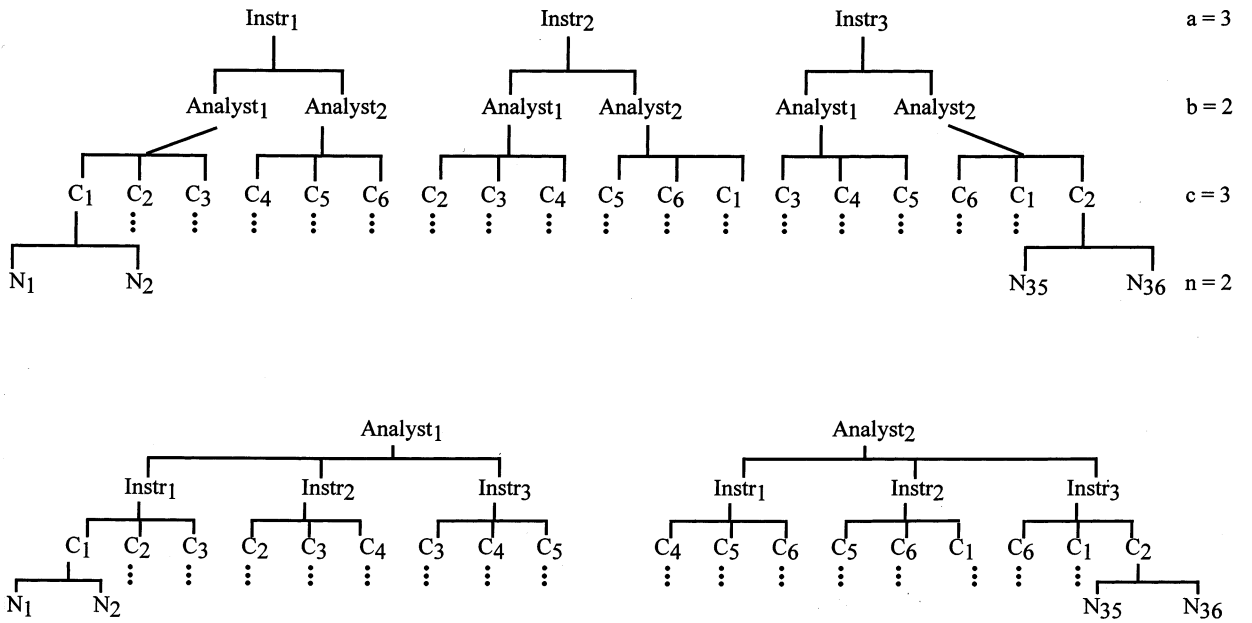

Fig. 4. Two designs requiring the same experiments but between which analysts and instruments are switched. C, column; N, replicate.

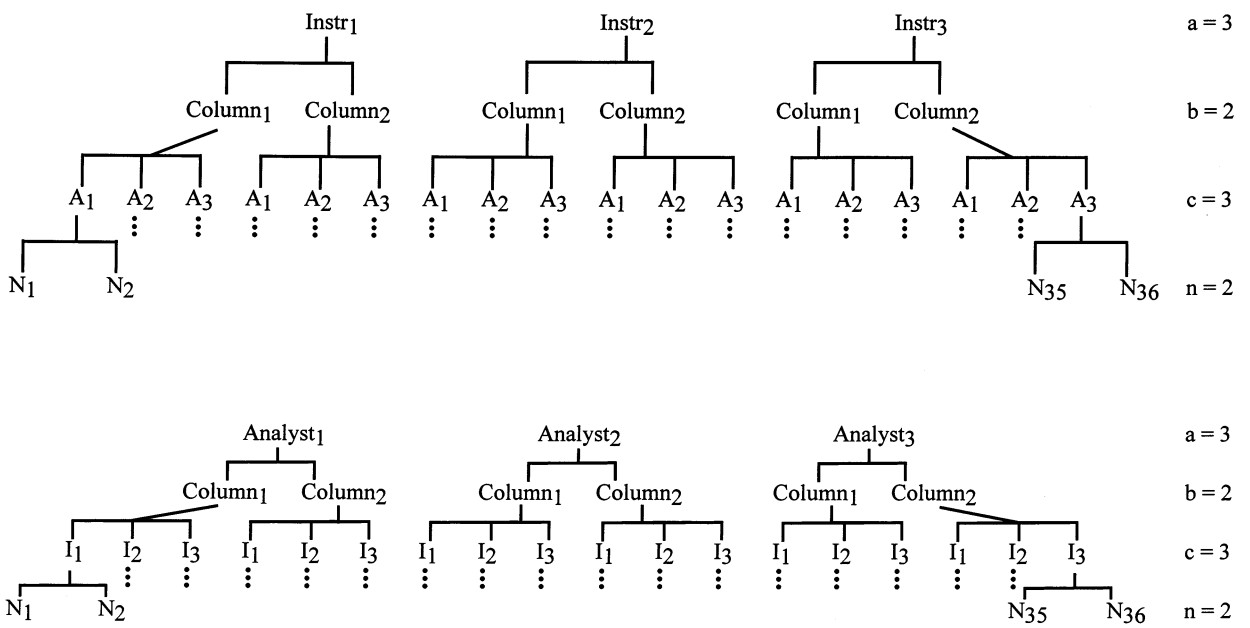

Fig. 5. Two designs requiring the same experiments but between which analysts and instruments are switched. Analysts and instruments situated in different levels compared to Fig. 4. A, analyst; I, instrument; N, replicate.

However when treating both designs according to the literature method, different variance estimates are obtained while the pooled variance method leads to exactly the same estimates (Table 3(a) and (c)).

The design of Fig. 3 can be considered a pure multiway ANOVA design for three instruments, two columns and three analysts (Fig. 5a). In that case Fig. 5b represents the same experimental set-up. However treating them according to the nested design interpretation again leads to different variance estimates while the pooled variances approach showed similar results between both set-ups (Table 3, (a) and (d)).

4.4. Analysis of experimental data

The results of four different designs performed with the TC method are shown in Tables 4 and 5. The designs are shown in Fig. 2. Since in one of the laboratories several analysts performed the same experiments (Fig. 2a) it is possible to create three designs equivalent to the one given in Fig. 2c. The variances and percent relative standard deviations (R.S.D.) estimated with both interpretations for two of these three are shown in Table 4 (c1 and c2). In the R.S.D. values, the standard deviations were expressed relative to the grand mean of the design. From Table 4 it can be observed that the pooled variance method does not yield negative values while the literature method does. For these latter, no R.S.D. values were calculated (*** in Table 4).

Designs c1 and c2 include the factor laboratories while designs a and b are performed within one lab. This explains for instance why the s^2_{instr}

Table 4
Results for four nested designs for the responses peak areas of TC and EATC[a]

Design	Source of variance	EATC peak				TC peak			
		Literature method		Pooled variances method		Literature method		Pooled variances method	
		s^2	R.S.D.	s^2	R.S.D.	s^2	R.S.D.	s^2	R.S.D.
a	Analysts	−2.8E+10	***	7.82E+08	6.35	−3.0E+13	***	1.40E+12	7.73
	Instruments	5.76E+10	54.53	5.77E+10	54.56	6.50E+13	53.47	6.50E+13	53.50
	Columns	2.05E+07	1.03	1.38E+08	2.67	1.40E+10	0.77	1.50E+11	2.57
	Days	2.64E+08	3.69	3.51E+08	4.26	3.50E+11	3.94	4.10E+11	4.26
	Replicates	1.74E+08	3.00	1.74E+08	3.00	1.20E+11	2.28	1.20E+11	2.28
b	DMFs	−2.5E+10	***	1.64E+07	0.99	−2.6E+13	***	1.41E+10	0.86
	Instruments	5.05E+10	54.60	5.06E+10	54.65	5.29E+13	52.46	5.30E+13	52.53
	Columns	−2.4E+07	***	1.83E+08	3.28	3.22E+10	1.29	2.82E+11	3.83
	Days	5.61E+08	5.75	6.20E+08	6.05	6.88E+11	5.98	7.48E+11	6.24
	Replicates	1.18E+08	2.64	1.18E+08	2.64	1.21E+11	2.51	1.21E+11	2.51
c1	Laboratories	2.46E+12	116.40	2.93E+12	126.88	2.97E+15	116.43	3.58E+15	127.69
	Instruments	9.28E+11	71.43	9.28E+11	71.44	1.21E+15	74.14	1.21E+15	74.14
	Columns	−1.6E+08	***	1.55E+08	0.92	−2.8E+11	***	1.91E+10	0.29
	Days	2.08E+08	1.07	9.49E+08	2.28	4.81E+11	1.48	8.88E+11	2.01
	Replicates	1.48E+09	2.85	1.48E+09	2.85	8.13E+11	1.92	8.13E+11	1.92
c2	Laboratories	2.50E+12	118.56	2.96E+12	129.01	3.04E+15	119.14	3.63E+15	130.36
	Instruments	9.21E+11	71.91	9.21E+11	71.92	1.20E+15	74.84	1.20E+15	74.84
	Columns	−1.8E+08	***	2.21E+08	1.11	−2.6E+11	***	1.53E+11	0.85
	Days	4.81E+08	1.64	1.21E+09	2.61	8.29E+11	1.97	1.24E+12	2.41
	Replicates	1.46E+09	2.86	1.46E+09	2.86	8.20E+11	1.96	8.20E+11	1.96

[a] The letters to indicate a design refer to those given in Fig. 2.

Table 5
Results (R.S.D. values) from the pooled variance method for a number of responses[a]

Design	Source of variance	Responses							
		[EATC]	[TC]	k'(EATC)	k'(TC)	t_R(TC)	R_s(EATC-TC)[b]	R_s(EATC-TC)[b]	W(TC)
a	Analysts	1.39	1.91	7.04	7.50	4.72	15.12	13.52	3.21
	Instruments	1.82	0.93	4.21	4.26	3.02	8.52	9.14	4.21
	Columns	1.51	1.40	8.08	8.92	6.44	8.62	7.88	4.85
	Days	1.57	2.95	4.27	5.58	3.88	13.15	11.50	4.08
	Replicates	2.08	2.09	0.69	0.55	0.36	1.83	1.70	1.64
b	DMFs	0.02	0.26	2.47	3.07	2.05	4.07	3.48	1.73
	Instruments	2.48	0.64	5.39	5.27	3.69	8.32	8.71	3.44
	Columns	0.64	1.02	2.62	3.83	2.06	10.88	10.12	2.36
	Days	1.14	1.89	2.80	1.63	1.19	10.20	9.38	3.34
	Replicates	1.53	2.25	0.64	0.39	0.28	1.80	1.54	1.69
c1	Laboratories	0.61	1.61	4.77	6.12	6.48	16.60	14.85	6.81
	Instruments	3.12	0.80	1.78	1.79	1.24	3.06	2.83	2.94
	Columns	1.77	0.62	7.36	7.77	5.80	5.74	5.36	5.10
	Days	3.22	2.45	4.15	5.45	3.98	8.01	6.98	4.56
	Replicates	2.08	1.68	0.81	0.64	0.54	1.31	1.45	1.33
c2	Laboratories	1.17	0.36	7.20	12.54	9.71	28.35	25.34	7.57
	Instruments	3.10	0.57	3.20	3.09	2.24	4.08	3.89	1.55
	Columns	1.37	0.79	1.99	2.15	1.48	3.42	3.10	3.00
	Days	3.10	1.57	3.68	3.45	2.61	5.50	5.10	4.10
	Replicates	1.63	1.72	0.77	0.57	0.50	1.17	1.16	1.21

[a] [EATC], [TC], content of EATC and of TC; k', capacity factor; t_R, retention time; R_s, resolution; W, peak width at half height.
[b] Resolutions of the two columns R_s(EATC-TC) were calculated differently (see text).

($R.S.D_{instr}$) is larger in designs c1 and c2. Direct comparisons of R.S.D. values between designs c1 and c2 on the one hand and designs a and b on the other are difficult because the grand means of these designs are quite different since some factors (qualitative ones!) were examined at different numbers of levels. However when comparisons are made within designs c1 and c2 on the one hand, and within designs a and b on the other, it is observed that similar R.S.D. values are obtained. Similar R.S.D. values for a factor are also found within one design but for different peaks. This is not the case when applying the interpretation from the literature.

From Table 4 it can be observed that for the important factors both methods give similar results. For less important factors considerable differences can be seen between both methods, especially when such a factor is situated in the hierarchy of the design above an important one.

In Table 5 the R.S.D.-values obtained from the pooled variance method for a number of responses are given. The responses were determined from chromatograms as shown in Fig. 6, obtained for the standard and sample solutions. The responses for which results are given are the contents of EATC and TC, the capacity factors of EATC and TC, the retention time of TC, the resolution between EATC and TC determined in two different ways, and the peak width at half height of TC. The results for the peak areas of EATC and TC were already shown in Table 4. The resolution (R_s) was determined in the first column (b) according to the equation

$$R_S = \frac{1}{4}\left(\frac{k'_{TC}}{k'_{EATC}} - 1\right)\sqrt{N_{TC}}\left(\frac{k'_{EATC}}{k'_{EATC}+1}\right) \quad (9)$$

where k' is a capacity factor and N_{TC} the number of theoretical plates derived from the TC peak, and in the second column (b) according to

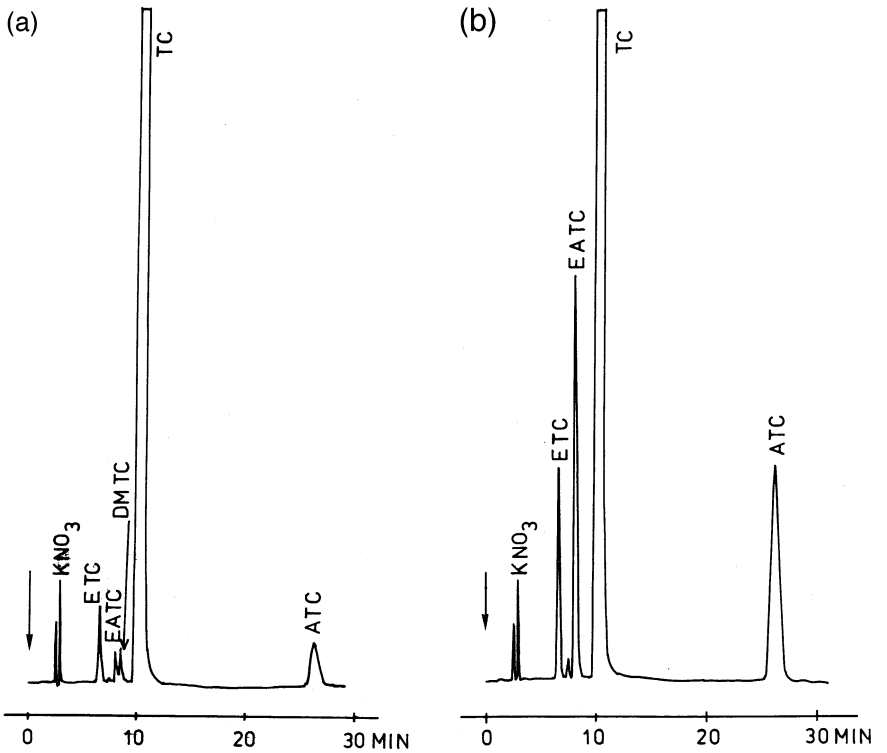

Fig. 6. Chromatogram obtained under the conditions described in the text, (a) for the standard solution, and (b) for the sample solution. AUFS, 0.032. ETC, EATC, TC, ATC, see text; DMTC, demethyltetracycline.

$$R_S = \frac{1.177(t_{TC} - t_{EATC})}{W_{EATC} + W_{TC}} \quad (10)$$

where t is a retention time and W a peak width at half height.

From Table 5 it can be observed, from comparison of the variances of the factors with that of the replicates, that in the four designs, none of the factors had an important effect on the contents determined for EATC and TC. From comparison of the results for the capacity factors it is again observed that within one design, similar results are obtained for the two peaks. The differences that for certain factors occur between designs can be explained by the influence that the ageing of the columns had on the considered response since no correction for their drift [15,16] was made.

5. Conclusions

The interpretation of nested designs, described in the literature, causes problems when these designs are performed for ruggedness testing purposes. Therefore an alternative interpretation was proposed. This pooled variance method worked properly in our case study.

The R.S.D. values obtained with the pooled variances method remain constant when a factor is examined in different designs, when examined at the same levels, and also for different peaks within one design. This in analogy with the use of normalised effects in screening designs [10,11].

Setting up a ruggedness test by means of a fully nested design is often not feasible for practical reasons. It requires many experiments while, in general, one tries to minimize this number [17]. It

may be more interesting to try staggered-nested designs which are also called nested designs with unequal sample size [8]. This kind of design could then be used to interpret some aspects of the ruggedness of a method without having to execute a fully nested design. One could use the measurements which have to be carried out for other method validation purposes to create a staggered-nested design. For instance there are repeatability measurements that could be combined with measurements at different days, performed by different analysts and/or on different instruments, with different columns, etc. to form a staggered-nested design. The interpretation described in the literature for the staggered-nested designs is even more complicated than for the fully nested ones. In particular, the determination of the numbers of degrees of freedom is not evident. Therefore it would be worthwhile to evaluate a similar interpretation as proposed here, based on the pooled variances.

Another possibility to examine a limited number of factors at, for instance, three levels while the majority are at only two, and that could be an alternative for the nested designs, are the so-called asymmetrical fractional factorial or asymmetrical screening designs [18,19]. In those designs one has the possibility to evaluate the effect of the different factors when these factors are studied at different numbers of levels.

References

[1] The United States Pharmacopeia XXII, The National Formulary XVII, United States Pharmacopeial Convention, Rockville, MD, 1990, p. 1337 and 1712.

[2] The United States Pharmacopeia 23, The National Formulary XVIII, United States Pharmacopeial Convention, Rockville, MD, 1995, p. 1510.

[3] W.J. Youden, E.H. Steiner, Statistical Manual of the Association of Official Analytical Chemists, The Association of Official Analytical Chemists, Arlington, VA, 1975, pp. 33–36, 70–71, 82–83.

[4] R.L. Plackett, J.P. Burman, The design of optimum multifactorial experiments, Biometrika 33 (1946) 305–325.

[5] E. Morgan, Chemometrics: Experimental Design, Analytical Chemistry by Open Learning, Wiley, Chichester, 1991, pp. 118–188.

[6] Y. Vander Heyden, D.L. Massart, Review of the use of robustness and ruggedness in analytical chemistry, in: A. Smilde, J. de Boer, M. Hendriks (Eds.), Robustness of Analytical Methods and Pharmaceutical Technological Products, Elsevier, Amsterdam, 1996, pp. 79–147.

[7] Y. Vander Heyden, F. Questier, D.L. Massart, Ruggedness testing of chromatographic methods: selection of factors and levels, J. Pharm. Biomed. Anal. 18 (1998) 43–56.

[8] R.R. Sokal, F.J. Rohlf, Biometry, 2nd edn, W.H. Freeman, San Francisco, CA, 1981, pp. 271–320.

[9] ISO, International Organization for Standardization, Accuracy (trueness and precision) of measurements methods and results. Part 3: Intermediate measures of the precision of a standard measurement method, International Standard ISO/DIS 5725-3, 1994.

[10] Y. Vander Heyden, K. Luypaert, C. Hartmann, D.L. Massart, J. Hoogmartens, J. De Beer, Ruggedness tests on the HPLC assay of the United States Pharmacopeia XXII for tetracycline hydrochloride. A comparison of experimental designs and statistical interpretations, Anal. Chim. Acta 312 (1995) 245–262.

[11] Y. Vander Heyden, C. Hartmann, D.L. Massart, L. Michel, P. Kiechle, F. Erni, Ruggedness tests on an HPLC assay: comparison of tests at two and three levels by using two-level Plackett–Burman designs, Anal. Chim. Acta 316 (1995) 15–26.

[12] 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. 78–86.

[13] ICH Harmonised Tripartite Guideline, Validation of Analytical procedures: methodology, The International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use, 6 November 1996 (http:/www.ifpma.org/ich1.html).

[14] Statgraphics® Plus, Statistical Graphics System by Statistical Graphics Corporation, version 6, Manugistics Inc., Rockville, USA.

[15] J.L. Goupy, Methods for Experimental Design, Principles and Applications for Physicists and Chemists, Elsevier, Amsterdam, 1993, pp. 159–177, 421–427.

[16] Y. Vander Heyden, A. Bourgeois, D.L. Massart, Influence of the sequence of experiments in a ruggedness test when drift occurs, Anal. Chim. Acta 347 (1997) 369–384.

[17] Y. Vander Heyden, F. Questier, D.L. Massart, A ruggedness test strategy for procedure related factors: experimental set-up and interpretation, J. Pharm. Biomed. Anal. 17 (1998) 153–168.

[18] S. Addelman, Symmetrical and asymmetrical fractional factorial plans, Technometrics 4 (1962) 21–46.

[19] J.M. Pernot, H. Brun, B. Pouyet, M. Sergent, R. Phan-Tan-Luu, Influence of experimental parameters on the microencapsulation of a photopolymerizable phase, J. Microencapsulation 10 (3) (1993) 323–328.