Available online at www.sciencedirect.com
ScienceDirect

Chemometrics and Intelligent Laboratory Systems 89 (2007) 69–81

Chemometrics and intelligent laboratory systems

www.elsevier.com/locate/chemolab

Bootstrap-based tolerance intervals for application to method validation

Tabea Rebafka [a], Stéphan Clémençon [b], Max Feinberg [b,*]

[a] TSI — ENST Paris, France
[b] Unité Met@risk — INRA, France

Received 26 February 2007; received in revised form 5 June 2007; accepted 5 June 2007
Available online 23 June 2007

Abstract

Recently a new validation procedure was developed using a graphical statistical tool – the so-called accuracy profile – that makes interpretation of results easy and straightforward. Accuracy profiles are estimated by *tolerance intervals*. Most existing methods for constructing tolerance limits are confined to the restrictive case of normally distributed data. The present study is focused on a nonparametric approach based on bootstrap — in order to get out of this constraint. The Mathematical section recalls some definitions and presents the derivation of the new nonparametric bootstrap approach for setting two-sided mean coverage and guaranteed coverage tolerance limits for a balanced one-way random effects model. The section concludes with a simulation study assessing the performance of the bootstrap methods in comparison to classical methods. Finally, the applicability of the proposed intervals is demonstrated by application to the problem of quantitative analytical method validation based on the accuracy profile. This approach is illustrated by an example consisting in the HPLC determination of the vitamers of vitamin B3 (nicotinamide and nicotinic acid) in milk. The efficiency of the new tolerance intervals is demonstrated as well as the applicability of accuracy profiles in the delicate situation where a correction factor must be applied because there is not a full recovery of the analyte. The comparison of the various tolerance intervals also gives some indication on their interpretation.
© 2007 Published by Elsevier B.V.

Keywords: Method validation; One-way random effects model; Tolerance interval; Bootstrap; Accuracy profile

1. Introduction

In analytical chemistry, *method validation* is a very fruitful topic as indicated by the great number of publications tackling this subject. Nevertheless, there has still been some ambiguity concerning the use of this term and method validation is often interpreted as the characterization of the method by computing characteristics such as reproducibility or specificity without answering the question whether the method is fit for its purpose.

Recently a statistical approach was proposed with the idea to develop a trade-off between the definition of a valid method and the intuition that this must carry out results with a known level of liability (see [1]). The definition is based on user-defined acceptance limits and confidence levels. The new approach uses a graphical statistical tool – the so-called *accuracy profile* – that makes interpretation of results easy. Several publications present this approach with fully developed examples (see [2,3]).

Usually accuracy profiles are estimated by tolerance intervals. Statistical tolerance intervals are useful in validation of analytical procedures, in process reliability studies, pharmaceutical engineering and many fields where a control of the proportion of units falling within some limits is required. There are two basic types: the mean coverage tolerance interval and the guaranteed coverage tolerance interval. Most existing methods for constructing tolerance limits are restricted either to the case of independently and identically distributed data, or else, when a dependency structure for the observations must be furthermore assumed, to the case of normally distributed data. However, in practice situations are often encountered where normality and the i.i.d. assumption do not hold both at the same time. In the present study, we thus focus on a nonparametric approach for data drawn from a *one-way random effects model* accounting for dependency patterns.

* Corresponding author. Unite Met@risk — INAPG-16, rue Claude Bernard — 75231 Paris cedex 05, France. Tel.: +33 1 44 08 16 52; fax: +33 1 44 08 72 76.
E-mail address: feinberg@inapg.fr (M. Feinberg).

0169-7439/$ - see front matter © 2007 Published by Elsevier B.V.
doi:10.1016/j.chemolab.2007.06.001

For normally distributed data, tolerance intervals are first mentioned in Wilks [4] where a mean coverage tolerance interval for a simple i.i.d. sample is presented. By extending this result Mee [5] proposes an analytical approximate method of setting mean coverage tolerance limits for balanced one-way random effects models under normal distribution. The approach of Lin and Liao [6] for arbitrary linear mixed models with normally distributed balanced data is based on the concept of *general pivotal quantities*.

The problem of setting guaranteed coverage tolerance intervals from a nonparametric point of view is first addressed in Wilks [4], where the case of a simple i.i.d. random sample is considered. The idea is to use order statistics as endpoints of the interval. However, it is mainly the case of normally distributed that has given rise to most refinements. Refer to Wald and Wolfowitz [7] and Howe [8] for tolerance limits in the case of independent sample variables. Wallis [9] and Weissberg and Beatty [10] extend these procedures to more general normal linear models. Recently developed procedures that are based on these results are the constructions of Hoffman and Kringle [11] and Liao and Iyer [12] where the former is an analytical approach and the latter is based on a Monte-Carlo algorithm. Both procedures hold for arbitrary normal linear mixed model situations. Another procedure for constructing guaranteed coverage tolerance limits that is adapted to the normal balanced one-way layout is the analytical approach presented in Mee and Owen [13] and Mee [5].

The structure of most laboratory data corresponds to the one-way random effects model situation, but the often assumed normality distribution of data may be unrealistic in certain cases and should then be abandoned; as e.g. data from bacteria counting in food microbiology illustrate. Therefore, this paper presents an alternative to ‘classical’ tolerance intervals based on a nonparametric bootstrap-*t* procedure for setting mean coverage and guaranteed coverage tolerance limits. The method of generating bootstrap samples is adapted to the one-way random effects model in a way that preserves the dependency structure of the data. These new intervals turn out to generally maintain the nominal contents and confidence levels, and to be of quite short length.

Section 2 recalls some definitions and presents the derivation of the new nonparametric bootstrap approach for setting two-sided mean coverage and guaranteed coverage tolerance limits for a balanced one-way random effects model. The section concludes with a simulation study assessing the performance of the bootstrap methods in comparison to classical methods. Finally, the applicability of the proposed intervals is demonstrated in the Result section by application to the problem of quantitative analytical method validation based on the accuracy profile and empirical evidence of the interest of the method is provided by an illustrative example.

2. Mathematical setting

2.1. Definitions and concepts

We first recall a few key concepts related to the notion of *statistical tolerance intervals* that will be used in the sequel and provide a short description of the statistical model we shall consider.

2.1.1. Mean coverage tolerance interval

Let F denote the (continuous) cumulative distribution function of a random variable Z. Suppose that $\mathbf{X}=(X_1,\ldots, X_n)$ is a data sample drawn from the distribution F, independent from Z. Then an interval $[L(\mathbf{X}),U(\mathbf{X})]$ based on the data vector $\mathbf{X}$ is said to be a two-sided mean coverage tolerance interval for F at confidence level β if

$$\mathbb{P}(Z\in[L(\mathbf{X}),U(\mathbf{X})]) = \mathbb{E}[F(U(\mathbf{X})) - F(L(\mathbf{X}))] = \beta. \tag{1}$$

It is noteworthy that the probability on the left hand side of Eq. (1) refers to the randomness in both the X_i’s and Z. Hence, a proportion β of the population modelled by F is contained in the interval $[L(\mathbf{X}), \mathrm{U}(\mathbf{X})]$ *on average*. In other words, a future observation Z drawn from the distribution F is contained in $[L(\mathbf{X}), U(\mathbf{X})]$ with probability β, i.e. mean coverage tolerance intervals coincide with the notion of a prediction interval for a single future observation (see Paulson [14]). Mean coverage tolerance intervals are also called β-expectation tolerance intervals. The random quantity $F(U(\mathbf{X}))-F(L(\mathbf{X}))$ is called the *coverage* or *content* of the tolerance interval.

2.1.2. Guaranteed coverage tolerance interval

The notion recalled above may be refined as follows. A statistical interval that assures a content β not on average but with some guarantee γ is known as a *guaranteed coverage tolerance interval*. More precisely, an interval $[L(\mathbf{X}),\mathrm{U}(\mathbf{X})]$ is called a two-sided guaranteed coverage tolerance interval for F with content β at confidence level γ if

$$\mathbb{P}(\mathbb{P}(Z\in[L(\mathbf{X}),\, U(\mathbf{X})]|\mathbf{X})\geq\beta)=\mathbb{P}(F(U(\mathbf{X}))-F(L(\mathbf{X}))\geq\beta)=\gamma, \tag{2}$$

denoting $\mathbb{P}(.|\mathbf{X})$ the conditional probability given $\mathbf{X}$. Thus, it can be claimed with confidence coefficient γ that at least a proportion β of the population modelled by F lies within the interval $[L(\mathbf{X}),U(\mathbf{X})]$. In the statistical literature such a tolerance interval is also called a β-content, γ-confidence tolerance interval, or, for short, (β,γ)-tolerance interval.

2.1.3. One-way random effects model

In practice, data are rarely independent, since various sources of error create dependency among the observations. For an analytical procedure in chemistry e.g. those sources of error can be the time, the operator in charge of the experimental measurements or the used instruments. Random effects models have been precisely introduced for modelling the heterogeneity of the data due to those sources of errors. On grounds of parcimony, those sources are oftenly grouped together and models with a single source of error, so-called one-way random effects models, are generally considered in practice. In such a model, X_{ik} denotes the k-th observation on item i selected randomly from a population of items, i.e. the k-th measurement at the i-th day/effected by the i-th operator/using the i-th instrument. The observation can be written as

$$X_{ik} = m + B_i + \varepsilon_{ik}, \tag{3}$$

where m denotes the mean of X_{ik}, and B_i is the random effect of the i-th item and e_{ik} denotes the measurement error or residual of the observation. It is assumed that the B_i's are zero mean i.i.d. random variables as well as the ε_{ik}, and all random variables are jointly independent. Hence, all X_{ik} have a common distribution F, but repeated measurements on the same item are correlated.

In this work we always consider balanced data $\{X_{ik}, 1 \leq i \leq I, 1, \leq k \leq K\}$ where the number of repeated observations K is the same whatever the item considered, this means e.g. that every day one effects the same number of measurements. Then, an unbiased estimate of the variance of F is classically based on the between-mean squares MS_b and the within-mean squares MS_ε given by

$$\hat{\sigma}^2 = \frac{1}{K}\mathrm{MS}_\mathrm{b} + \left(1 - \frac{1}{K}\right)\mathrm{MS}_\varepsilon. \tag{4}$$

2.2. Tolerance intervals by a bootstrap-t procedure

The bootstrap method is a data resampling technique introduced by Efron in 1979 [15] providing an estimate of the distribution of a statistic. It forms an alternative to traditional statistical methods by considering data as if they were the population of interest and where statistical inferences are based on repeatedly resampling the original data. Moreover, they can even eliminate the need to impose a convenient statistical model that does not have a strong scientific basis. In model (3) for instance, it avoids to specify restrictive parametric forms for the distributions of the random effect and the measurement error. Bootstrap methods have emerged as powerful tools that offer the potential for highly accurate inferential methods. Because of the availability of inexpensive and fast computing, these computer-intensive methods have caught on very rapidly in recent years. Refer to Shao and Tu [16] for key concepts of the bootstrap theory and to Efron and Tibshirani [17] for a practice-oriented account. In particular, application of bootstrap methods to various problems arising in chemometrics is discussed in Wehrens et al. [18].

2.2.1. Bootstrapped mean coverage tolerance intervals

Initially, the bootstrap technique was used for samples of independent and identically distributed data. In this case Efron and Tibshirani [17] propose nonparametric mean coverage tolerance limits via the so-called bootstrap-*t* method (see also Fernholz and Gillespie [19]). In this section we further develop this approach in order to adapt it to the one-way random effects model on the one hand and to obtain intervals of shorter length by abandoning the symmetry constraint in the construction on the other hand.

In many applications tolerance intervals of short length are preferred since short intervals are more informative concerning the concentration of the underlying distribution. Clearly, under the normality assumption the shortest mean coverage tolerance interval is a symmetric interval around the mean. In many applications where tolerance intervals are seeked in a nonparametric setting, one may also assume that a high proportion of the population is in the neighbourhood of the mean. In such a case, in order to obtain short intervals, we may seek two-sided tolerance intervals around the sample mean $\bar{X}=(IK)^{-1}\Sigma_{1 \leq i \leq I}\Sigma_{1 \leq k \leq K}X_{ik}$ of the form $[\bar{X}+t_1\hat{\sigma}, \bar{X}+t_2\hat{\sigma}]$ where the structure of the observed data $\{X_{ik}\}$ is a balanced one-way random effects model with underlying distribution function F and $\hat{\sigma}^2$ denotes the unbiased variance estimate defined by Eq. (4). One may choose $t_1 \neq -t_2$ in order to allow for intervals that are non-symmetric around the empirical mean $\bar{X}$, which may sometimes be much shorter than symmetric ones when the underlying distribution F is skewed.

Now the interval $[\bar{X}+t_1\hat{\sigma}, \bar{X}+t_2\hat{\sigma}]$ becomes a mean coverage tolerance interval at confidence level β if t_1 and t_2 are determined such that the following equation holds:

$$\begin{aligned}\beta &= \mathbb{E}\left[F\left(\bar{X}+t_2\hat{\sigma}\right) - F\left(\bar{X}+t_1\hat{\sigma}\right)\right] \\ &= \mathbb{P}\left(Z \in \left(\bar{X}+t_1\hat{\sigma}, \bar{X}+t_2\hat{\sigma}\right)\right) = \mathbb{P}\left(t_1 < \frac{Z-\bar{X}}{\hat{\sigma}} < t_2\right),\end{aligned}$$

where Z denotes an independent future observation from the distribution F. Thus, the quantities t_1 and t_2 may be viewed as quantiles of the statistic $T=\frac{Z-\bar{X}}{\hat{\sigma}}$ of orders β_1 and $\beta_2=\beta+\beta_1$ respectively, the tuning parameter $\beta_1 \in [0, 1-\beta]$ being possibly picked so as to minimize t_2-t_1 (or equivalently the length $\hat{\sigma}(t_2-t_1)$ of such a mean coverage interval). For any $\beta_1 \in [0, 1-\beta]$, these quantities may be estimated in a nonparametric fashion by applying a bootstrap-*t* procedure as described below.

In order to generate new data, the so-called *bootstrap samples*, one has to resample the data so as to preserve the dependency structure of the observations. We recall that our sample consists of I subvectors $(X_{i1},\ldots, X_{ik})$ of length K where each subvector represents the data corresponding to one realization of the effect (one day or one lab for instance). In order to model the randomness of the effect one selects with replacement I subvectors from the original data vector, i.e. each subvector is drawn with probability $1/I$. Some vectors may be thus selected several times while others are left out. The second step consists of resampling each selected data vector by drawing K observations with replacement from each subvector, i.e. each value is drawn with probability $1/K$. Hence, in this way we obtain a bootstrap sample that has the same dependency structure as the original data. An alternative method for generating bootstrap datasets could classically consist in resampling 'residuals' in the linear model (3) (see a description of this resampling scheme in subsection 7.2 in [18] for instance). However, insofar as such an approach entirely relies on the additive structure of the one-way random effects model, it is much more "model-dependent" than the method proposed above. The required mean coverage tolerance interval may be estimated with the following algorithm.

Algorithm 1.

Step 1 Choose B, the number of bootstrap replicates, say 5000.

Step 2 Generate B independent bootstrap samples of size $I \times K$, $x^*(1),\ldots, x^*(B)$ by, sequentially, drawing I subvectors with replacement from the original data vector and then drawing K observations with replacement from each subvector thus obtained.

Step 3 Generate B independent sample variables $z^*(1),\ldots, z^*(B)$ each drawn from the data vector $\{X_{ik}\}$ with probability $1/N$ where $N=IK$.

Step 4 For b from 1 to B, evaluate the corresponding bootstrap replicate

$$T^*(b) = T(x^*(b), z^*(b)) = \frac{z^*(b) - \bar{x}^*(b)}{\hat{\sigma}(x^*(b))},$$

where $\bar{x}^*(b)$ and $\hat{\sigma}(x^*(b))$ respectively denote the bootstrapped versions of the sample mean and the standard error estimate computed from Eq. (4) based on $x^*(b)$.

Step 5 Derive the β_1- and β_2-sample quantiles $\hat{t}_{\beta_1}$ and $\hat{t}_{\beta_2}$ from $\{T^*(b)\}$ such that the distance between the sample quantiles $\hat{t}_{\beta_2} - \hat{t}_{\beta_1}$ is minimized under the constraint $\beta_2 - \beta_1 = \beta$.

Step 6 The estimated mean coverage tolerance interval is given by $[\bar{X} + \hat{t}_{\beta_1}\hat{\sigma}, \bar{X} + \hat{t}_{\beta_2}\hat{\sigma}]$.

Due to the minimization in Step 5 this algorithm provides the shortest tolerance interval of the form $[\bar{X} + t_1\hat{\sigma}, \bar{X} + t_2\hat{\sigma}]$.

2.2.2. Bootstrapped guaranteed coverage tolerance intervals

Similarly we construct guaranteed coverage tolerance intervals for balanced one-way random effects models. For the same reasons as above guaranteed coverage tolerance intervals of short length are preferred in most applications. We thus seek two-sided β-content, γ-confidence tolerance limits of the form $[\bar{X} + k_1\hat{\sigma}, \bar{X} + k_2\hat{\sigma}]$ where k_1 and k_2 satisfy

$$\begin{aligned}\gamma &= \mathbb{P}\left(F(\bar{X} + k_2\hat{\sigma}) - F(\bar{X} + k_1\hat{\sigma}) \geq \beta\right)\\ &= \mathbb{P}\left(\mathbb{P}\left(k_1 < \frac{Z - \bar{X}}{\hat{\sigma}} < k_2 | X_{11}, \ldots, X_{IK}\right) \geq \beta\right),\end{aligned}$$

where Z denotes a future observation drawn from the distribution F independent from the X_{ik}'s . The statistic of interest is again $T = \frac{Z - \bar{X}}{\hat{\sigma}}$ and the quantities k_1 and k_2 are determined by a *double bootstrap*. In a first step quantiles for the conditioned distribution $T|X_{ik}^*$ are derived by a bootstrap procedure where T is conditioned on a (fixed) bootstrap sample $\{X_{ik}^*\}$. This is repeated for a large number of different bootstrap samples. The second step consists of selecting k_1 and k_2 from the set of conditioned quantiles in a way that the guarantee γ is attained. To be more specific, for balanced data $\{X_{ik}\}$ from a one-way random effects model one applies the following algorithm.

Algorithm 2.

Step 1 Choose B and C, the numbers of bootstrap replicates of the first and second bootstrap procedures respectively.

Table 1
Simulation study for the mean coverage tolerance interval that contrasts the bootstrap-*t* to the intervals of Mee and Lin and Liao for normal distributions

I	*K*	*N*	Mean coverage (std)			Expected length (std)		
			Mee	Lin	Bootstrap	Mee	Lin	Bootstrap
Normal distribution β=0.80 and R=0.10								
3	3	9	0.799 (0.124)	0.882 (0.113)	0.790 (0.115)	3.460 (1.293)	4.421 (1.681)	3.319 (1.101)
3	5	15	0.802 (0.108)	0.874 (0.107)	0.797 (0.107)	3.278 (0.792)	4.116 (1.200)	3.248 (0.700)
3	10	30	0.827 (0.102)	0.897 (0.088)	0.814 (0.097)	3.461 (0.951)	4.353 (1.332)	3.310 (0.795)
5	3	15	0.814 (0.082)	0.854 (0.079)	0.807 (0.089)	3.265 (0.647)	3.618 (0.733)	3.249 (0.643)
5	5	25	0.817 (0.073)	0.851 (0.072)	0.809 (0.084)	3.228 (0.603)	3.531 (0.694)	3.237 (0.609)
5	10	50	0.803 (0.057)	0.836 (0.059)	0.788 (0.059)	3.059 (0.403)	3.320 (0.480)	3.008 (0.398)
10	3	30	0.807 (0.064)	0.825 (0.061)	0.799 (0.068)	3.124 (0.439)	3.259 (0.451)	3.082 (0.453)
10	5	50	0.810 (0.042)	0.826 (0.041)	0.805 (0.043)	3.072 (0.293)	3.186 (0.310)	3.059 (0.306)
10	10	100	0.804 (0.036)	0.818 (0.038)	0.800 (0.037)	3.031 (0.259)	3.129 (0.283)	3.028 (0.264)
*Normal distribution β=*0.90 and *R*=1.0								
3	3	9	0.893 (0.121)	0.951 (0.078)	0.879 (0.140)	6.800 (2.862)	8.829 (3.544)	6.665 (2.689)
3	5	15	0.894 (0.095)	0.958 (0.060)	0.883 (0.098)	5.774 (1.940)	7.768 (2.520)	5.528 (1.953)
3	10	30	0.909 (0.111)	0.952 (0.097)	0.891 (0.113)	6.111 (2.296)	8.245 (3.109)	5.628 (1.963)
5	3	15	0.905 (0.097)	0.931 (0.080)	0.895 (0.093)	5.740 (1.702)	6.350 (1.827)	5.503 (1.559)
5	5	25	0.906 (0.069)	0.935 (0.055)	0.899 (0.071)	5.528 (1.597)	6.143 (1.686)	5.326 (1.449)
5	10	50	0.914 (0.063)	0.941 (0.049)	0.907 (0.059)	5.657 (1.736)	6.263 (1.797)	5.380 (1.372)
10	3	30	0.897 (0.061)	0.909 (0.058)	0.899 (0.052)	5.035 (0.807)	5.233 (0.815)	5.018 (0.725)
10	5	50	0.899 (0.060)	0.911 (0.057)	0.898 (0.062)	4.958 (0.828)	5.151 (0.860)	4.972 (0.838)
10	10	100	0.898 (0.049)	0.910 (0.046)	0.895 (0.050)	4.984 (0.875)	5.168 (0.895)	4.955 (0.861)
*Normal distribution β=*0.95 and *R*=5.0								
3	3	9	0.948 (0.124)	0.971 (0.104)	0.941 (0.136)	13.315 (6.467)	17.325 (7.334)	14.716 (8.820)
3	5	15	0.932 (0.090)	0.978 (0.043)	0.928 (0.090)	10.370 (6.232)	14.125 (6.983)	10.122 (6.117)
3	10	30	0.945 (0.096)	0.980 (0.052)	0.929 (0.096)	11.482 (5.207)	15.908 (6.380)	10.246 (4.756)
5	3	15	0.964 (0.056)	0.976 (0.042)	0.952 (0.061)	9.362 (2.473)	10.223 (2.505)	9.023 (2.693)
5	5	25	0.931 (0.068)	0.956 (0.047)	0.915 (0.075)	8.344 (2.760)	9.231 (2.815)	7.829 (2.512)
5	10	50	0.943 (0.069)	0.963 (0.050)	0.928 (0.076)	8.961 (3.040)	9.878 (3.095)	8.712 (3.244)
10	3	30	0.958 (0.043)	0.964 (0.038)	0.945 (0.047)	7.963 (1.434)	8.200 (1.428)	7.569 (1.258)
10	5	50	0.952 (0.045)	0.959 (0.040)	0.946 (0.050)	8.021 (1.784)	8.268 (1.779)	7.681 (1.560)
10	10	100	0.942 (0.034)	0.950 (0.029)	0.931 (0.043)	7.433 (1.277)	7.685 (1.270)	7.192 (1.338)

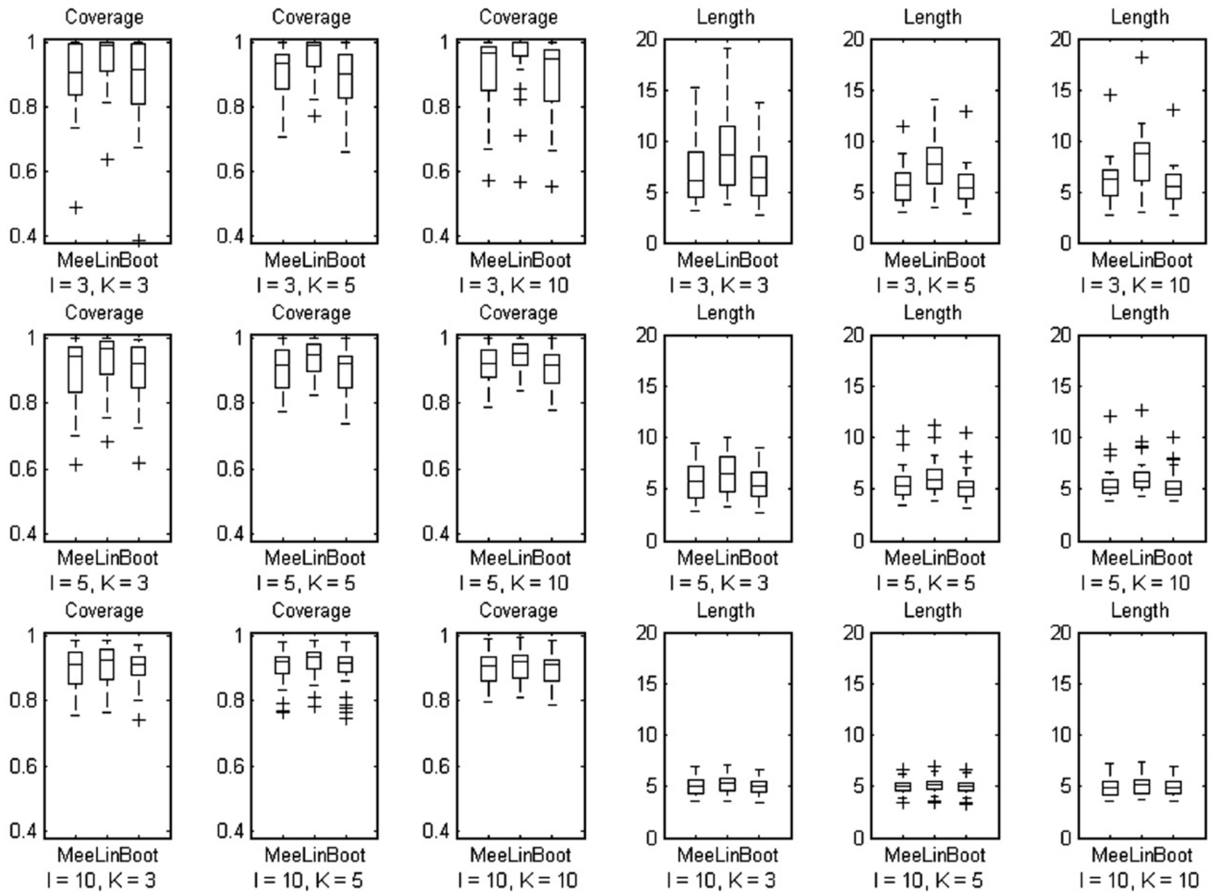

Fig. 1. Boxplots of the coverages and the lengths of the mean coverage tolerance interval for the various choices of $(I, K) \in \{3, 5, 10\}^2$ for the normal distribution with $R=1$ and $\beta=0.90$.

Step 2 Generate B independent bootstrap samples $x^*(1),\ldots, x^*(B)$ of size $I \times K$, by, sequentially, drawing I subvectors with replacement from the original data vector and then drawing K observations with replacement from each subvector thus obtained.

Step 3 For b from 1 to B repeat:

1. Generate C independent sample variables $z^*(1),\ldots, z^*(C)$ drawn with replacement from the initial sample $\{X_{ik}\}$.
2. For $c=1,\ldots, C$, compute the corresponding bootstrap replicates

$$T_b^*(c) = T\left(x^*(b), z^*(c)\right) = \frac{z^*(c) - \overline{x}^*(b)}{\hat{\sigma}(x^*(b))},$$

   where $\overline{x}^*(b)$ and $\hat{\sigma}(x^*(b))$ denote the bootstrapped versions of the sample mean resp. the standard deviation estimate computed from Eq. (4) based on $x^*(b)$.
3. Derive the β_1- and β_2-sample quantiles $\hat{t}_{\beta_1}(b)$ and $\hat{t}_{\beta_2}(b)$ from $\{T_b^*(c)\}$ such that the distance between the sample quantiles $\hat{t}_{\beta_2}(b) - \hat{t}_{\beta_1}(b)$ is minimized under the constraint $\beta_2 - \beta_1 = \beta$.

Step 4 Choose k_1 and k_2 such that a proportion γ of the intervals $[\hat{t}_{\beta_1}(b), \hat{t}_{\beta_2}(b)]$ are completely included in the interval $[k_1, k_2]$.

Step 5 The estimated guaranteed coverage tolerance interval is given by $[\overline{X} + k_1\hat{\sigma}, \overline{X} + k_2\hat{\sigma}]$.

2.3. Simulation study

Several Monte-Carlo experiments have been carried out in order to investigate how the methods proposed above for building tolerance intervals perform, when compared to classical procedures.

2.3.1. The mean coverage tolerance interval

A 'good' tolerance interval is characterized by closeness of the attained mean coverage to the nominal value of β and a short interval's length both at the same time. Furthermore, the estimate should be good for small sample sizes, since collecting data through experimental measurements is always expensive and for technical reasons two or three repetitions only are available in most situations. Here the bootstrap mean coverage tolerance interval is compared to the intervals constructed using the method of Mee [5] and the one of Lin and Liao [6]. The classical approach of Mee is an analytical method using the Satterthwaite's approximation. Lin and Liao's intervals are based on the concept of generalized pivotal quantities and uses Monte-Carlo simulation. Here we use an improved version of the original method by choosing the quantiles so that they

Table 2
Simulation study for the mean coverage tolerance interval that contrasts the bootstrap-*t* to the intervals of Mee and Lin and Liao for Pareto distribution and a mixture of two normal distributions

I	*K*	*N*	Mean coverage (std)			Expected length (std)		
			Mee	Lin	Bootstrap	Mee	Lin	Bootstrap
Pareto distribution β=0.70 and R=0.10								
3	3	9	0.720 (0.195)	0.794 (0.157)	0.687 (0.186)	1.263 (0.816)	1.606 (1.068)	1.073 (0.603)
3	5	15	0.780 (0.158)	0.835 (0.130)	0.733 (0.133)	1.709 (2.095)	2.052 (2.564)	1.194 (1.129)
3	10	30	0.837 (0.090)	0.875 (0.071)	0.727 (0.104)	1.409 (0.630)	1.652 (0.743)	0.916 (0.324)
5	3	15	0.820 (0.118)	0.852 (0.100)	0.741 (0.126)	1.548 (1.112)	1.708 (1.238)	1.041 (0.530)
5	5	25	0.824 (0.119)	0.841 (0.110)	0.726 (0.103)	1.381 (0.638)	1.486 (0.694)	0.879 (0.306)
5	10	50	0.853 (0.077)	0.864 (0.067)	0.723 (0.064)	1.409 (0.511)	1.477 (0.547)	0.818 (0.276)
10	3	30	0.826 (0.118)	0.836 (0.110)	0.724 (0.073)	1.407 (0.757)	1.465 (0.787)	0.858 (0.278)
10	5	50	0.876 (0.066)	0.882 (0.061)	0.735 (0.071)	1.597 (0.667)	1.643 (0.683)	.874 (0.243)
10	10	100	0.907 (0.042)	0.910 (0.037)	0.756 (0.054)	1.716 (0.482)	1.755 (0.491)	0.869 (0.159)
Normal mixture distribution β=0.70 and R=0.20								
3	3	9	0.705 (0.172)	0.780 (0.145)	0.671 (0.153)	3.647 (2.432)	4.458 (2.815)	3.204 (1.801)
3	5	15	0.719 (0.102)	0.769 (0.102)	0.693 (0.100)	3.021 (0.726)	3.520 (0.950)	2.659 (0.587)
3	10	30	0.742 (0.046)	0.775 (0.049)	0.694 (0.064)	3.076 (0.488)	3.362 (0.564)	2.527 (0.397)
5	3	15	0.773 (0.059)	0.805 (0.055)	0.753 (0.073)	3.471 (0.599)	3.798 (0.650)	3.076 (0.571)
5	5	25	0.727 (0.139)	0.760 (0.130)	0.718 (0.133)	3.442 (1.629)	3.728 (1.716)	3.237 (1.397)
5	10	50	0.733 (0.106)	0.763 (0.105)	0.713 (0.095)	3.126 (1.013)	3.399 (1.135)	2.735 (0.680)
10	3	30	0.762 (0.074)	0.775 (0.072)	0.713 (0.092)	3.261 (0.682)	3.381 (0.705)	2.756 (0.598)
10	5	50	0.712 (0.096)	0.724 (0.095)	0.695 (0.075)	2.903 (0.825)	2.998 (0.857)	2.613 (0.550)
10	10	100	0.735 (0.108)	0.746 (0.106)	0.711 (0.104)	3.261 (1.191)	3.353 (1.211)	2.824 (0.871)

minimize the interval's length (in the original work one uses symmetric quantiles). Both methods assume normality of the data, i.e. both the effects B_i and the residuals e_{ik} have a normal distribution.

All routines have been implemented using the open source statistical software R (see http://www.r-project.org). Simulations are carried out by simulating data for different distributions, then computing the mean coverage and the expected length of the three estimated tolerance intervals. More precisely, given a certain distribution F, we simulated 30 independent data sets drawn from model (3) for I and K taking on all possible values in the set $\{3, 5, 10\}$, from which we computed β-tolerance limits via the three different methods. Then, by generating again a large number of values from the distribution (100,000), we derived the coverage of each interval, i.e. the proportion of variables included within the tolerance limits. Finally, the mean of those coverages appears in the tables as 'mean coverage', and the 'expected length' in the tables is computed by taking the mean of the lengths of the 30 intervals. Figures in brackets are the corresponding standard deviations.

Table 1 presents the results for normal distribution for different values of β (namely, 0.8, 0.9, 0.95) and different variance ratios $R=\sigma_b^2/\sigma_\varepsilon^2$ (namely 0.1, 1, 5) where σ_b^2 is the variance of the effect B_i and σ_ε^2 the variance of the residuals ε_{ik}. Each table contains the results for various numbers of items I (number of days/operators/instruments) and repetitions K per item. Obviously, all three methods provide mean coverages quite close to the nominal value β, and as the number of observations increases the mean coverages tend even more to the value of β. Further, the bootstrap almost always provides the shortest intervals. Varying the values of β has no influence on the performance of one of the methods. However, increasing the variance ratios R slightly lowers the convergence of the attained mean coverage.

In order to illustrate the typical distributions of the coverages and the lengths of the single tolerance intervals Fig. 1 presents the boxplots for each simulation level for the normal distribution when R=1. For all three methods one notes convergence of the coverages and the lengths when increasing the number of observations, i.e. all values become more concentrated.

Furthermore, we observe from Table 1 that Lin and Liao's method always yields the largest intervals. Hence, we cannot confirm Lin and Liao's statement that their method provides shorter intervals than Mee's method, and in addition, we have not found that their mean coverages are closer to β than Mee's.

Table 2 displays further empirical results obtained from simulated data sets drawn from Eq. (3) with $(I, K) \in \{3, 5, 10\}^2$

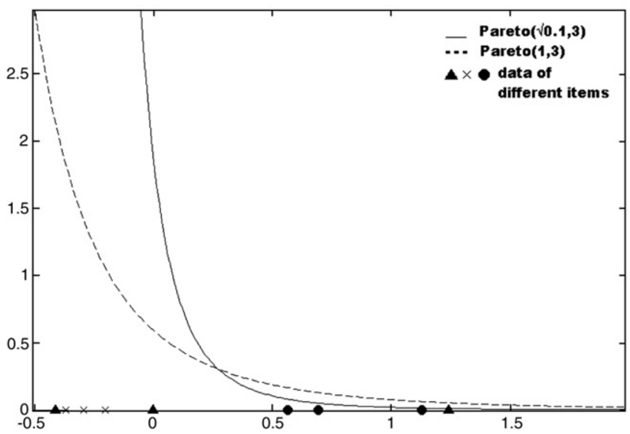

Fig. 2. Illustration of the Pareto based densities of the effects and the residuals. The effect follows a Par($\sqrt{0.1}$,3) minus its mean (solid line) and the residual a Par(1, 3) minus its mean (dashed line). A data set of 3 items and 3 repetitions per item from the corresponding one-way random effects model is indicated. The different symbols correspond to different items.

Table 3
Simulation study for the guaranteed coverage tolerance interval that contrasts the bootstrap-*t* to the intervals of Hoffman and Kringle and Liao and Iyer for normal distributions

I	*K*	*N*	Achieved guarantee (std coverage)			Expected Length (std)		
			Mee	Lin	Bootstrap	Mee	Lin	Bootstrap
Normal distribution β=0.70, γ=0.90 and R=1.0								
3	3	9	0.57 (0.168)	0.94 (0.124)	0.93 (0.127)	3.599 (1.306)	7.665 (3.466)	8.659 (5.431)
3	5	15	0.57 (0.157)	0.89 (0.124)	0.89 (0.122)	3.721 (1.364)	7.628 (3.780)	7.180 (4.934)
3	10	30	0.50 (0.147)	0.91 (0.126)	0.89 (0.129)	3.512 (1.167)	6.882 (3.365)	6.034 (3.305)
5	3	15	0.59 (0.127)	0.91 (0.111)	0.91 (0.108)	3.457 (0.916)	5.115 (1.505)	5.600 (1.763)
5	5	25	0.63 (0.115)	0.86 (0.113)	0.89 (0.109)	3.443 (0.850)	4.877 (1.478)	5.203 (1.625)
5	10	50	0.63 (0.107)	0.85 (0.110)	0.92 (0.105)	3.383 (0.783)	4.658 (1.389)	4.951 (1.464)
10	3	30	0.74 (0.077)	0.94 (0.074)	0.94 (0.078)	3.379 (0.548)	4.046 (0.691)	4.629 (0.874)
10	5	50	0.70 (0.089)	0.86 (0.092)	0.94 (0.087)	3.332 (0.604)	3.916 (0.785)	4.601 (1.024)
10	10	100	0.67 (0.071)	0.88 (0.078)	0.92 (0.073)	3.256 (0.471)	3.758 (0.638)	4.308 (0.736)
Normal distribution β=0.80, γ=0.90 and R=5.0								
3	3	9	0.51 (0.152)	0.97 (0.061)	0.89 (0.092)	5.763 (2.303)	12.935 (6.170)	16.218 (13.264)
3	5	15	0.51 (0.180)	0.88 (0.118)	0.88 (0.125)	5.800 (2.402)	12.795 (6.470)	12.278 (8.530)
3	10	30	0.56 (0.131)	0.91 (0.092)	0.86 (0.106)	5.531 (2.109)	11.838 (5.890)	9.766 (6.363)
5	3	15	0.60 (0.134)	0.88 (0.095)	0.92 (0.091)	5.658 (1.727)	8.573 (2.864)	9.699 (4.168)
5	5	25	0.66 (0.126)	0.87 (0.093)	0.90 (0.097)	5.710 (1.739)	8.546 (2.921)	9.375 (3.871)
5	10	50	0.59 (0.104)	0.93 (0.076)	0.90 (0.079)	5.392 (1.398)	7.921 (2.400)	8.265 (2.943)
10	3	30	0.71 (0.088)	0.87 (0.073)	0.91 (0.070)	5.432 (1.119)	6.611 (1.426)	7.307 (1.722)
10	5	50	0.74 (0.083)	0.89 (0.072)	0.89 (0.071)	5.386 (1.055)	6.485 (1.356)	7.120 (1.653)
10	10	100	0.64 (0.085)	0.85 (0.077)	0.89 (0.072)	5.247 (1.044)	6.281 (1.359)	6.988 (1.651)

and illustrates the advantage of the bootstrap. A one-way random effects model based on the Pareto distribution is considered. Let Y denote a random variable following a Pareto distribution Par(x_m, k) with location parameter x_m and scale parameter k, then both effects and residuals follow the distribution of $Y-\mathbb{E}[Y]$, i.e. a Pareto distribution minus its mean. Fig. 2 illustrates the densities of the effect, a $\mathrm{Par}\left(\sqrt{0.1},3\right)$ minus its mean, and the residual, a Par(1, 3) minus its mean. Further, a data set of size (I, K)=(3, 3) is indicated in Fig. 2 where each symbol corresponds to a different item $i\in\{1,2,3\}$. Here Pareto distributions are considered because they are typical skewed and heavy-tailed distributions in opposition to the normal distribution. The methods of Mee and Lin and Liao fail completely, whereas the bootstrap intervals are still quite

Table 4
Simulation study for the guaranteed coverage tolerance interval that contrasts the bootstrap-*t* to the intervals of Hoffman and Kringle and Liao and Iyer for Pareto distribution and a mixture of two normal distributions

I	*K*	*N*	Achieved guarantee (std coverage)			Expected Length (std)		
			Mee	Lin	Bootstrap	Mee	Lin	Bootstrap
Pareto distribution β=0.90, γ=0.60 and R=1.0								
3	3	9	0.47 (0.154)	0.55 (0.124)	0.64 (0.144)	3.196 (3.222)	4.202 (4.295)	10.375 (25.910)
3	5	15	0.49 (0.088)	0.52 (0.074)	0.62 (0.092)	3.063 (2.179)	3.795 (2.820)	4.865 (8.515)
3	10	30	0.56 (0.090)	0.64 (0.081)	0.60 (0.102)	3.197 (2.209)	3.757 (2.760)	4.000 (6.650)
5	3	15	0.51 (0.079)	0.59 (0.068)	0.64 (0.073)	3.184 (3.104)	3.742 (3.703)	6.398 (21.488)
5	5	25	0.52 (0.074)	0.50 (0.069)	0.62 (0.072)	2.872 (1.495)	3.239 (1.742)	3.370 (3.246)
5	10	50	0.52 (0.059)	0.66 (0.056)	0.60 (0.069)	3.818 (4.860)	4.243 (5.765)	7.977 (44.840)
10	3	30	0.71 (0.068)	0.74 (0.064)	0.72 (0.063)	3.263 (1.519)	3.512 (1.638)	3.470 (1.859)
10	5	50	0.70 (0.056)	0.72 (0.054)	0.68 (0.059)	3.771 (2.175)	3.988 (2.307)	3.863 (3.083)
10	10	100	0.65 (0.041)	0.58 (0.040)	0.62 (0.048)	3.248 (1.390)	3.397 (1.485)	3.103 (2.055)
Normal mixture distribution β=0.80, γ=0.60 and R=5.0								
3	3	9	0.52 (0.193)	0.89 (0.110)	0.83 (0.137)	7.640 (3.056)	17.939 (7.927)	24.545 (17.586)
3	5	15	0.57 (0.085)	0.96 (0.056)	0.95 (0.062)	7.496 (1.412)	11.717 (3.572)	11.115 (2.559)
3	10	30	0.73 (0.063)	0.95 (0.066)	0.93 (0.058)	7.428 (0.915)	9.737 (1.968)	9.573 (1.561)
5	3	15	0.69 (0.087)	0.95 (0.063)	0.95 (0.068)	7.684 (1.513)	10.389 (2.327)	11.252 (2.619)
5	5	25	0.64 (0.138)	0.88 (0.098)	0.87 (0.098)	8.320 (2.657)	12.845 (4.371)	15.653 (7.295)
5	10	50	0.56 (0.090)	0.93 (0.068)	0.90 (0.071)	7.389 (1.527)	9.933 (2.517)	10.588 (2.973)
10	3	30	0.72 (0.059)	0.93 (0.052)	0.95 (0.054)	7.346 (0.894)	8.437 (1.062)	9.462 (1.357)
10	5	50	0.44 (0.099)	0.69 (0.094)	0.80 (0.086)	6.963 (1.574)	7.972 (1.931)	8.948 (2.254)
10	10	100	0.61 (0.088)	0.90 (0.073)	0.92 (0.065)	7.486 (1.406)	9.141 (1.791)	10.209 (2.165)

short and the attained mean coverage is close to the required value β. For the mixture of two normal distributions, similar results are obtained, see Table 2.

Finally, we conclude within the limits of these experiments that for normal data the bootstrap-*t* and Mee's method appear as the most competitive. However, in the nonparametric case only the bootstrap-*t* method is recommended, since it still performs well for nongaussian data and provides almost exact mean coverage tolerance intervals of short length.

2.3.2. The guaranteed coverage tolerance interval

Concerning the guaranteed coverage tolerance interval the bootstrap intervals are contrasted to the intervals obtained using the methods of Hoffman and Kringle [11] and Liao and Iyer [12] in a simulation study. The former is an analytical approach using approximate confidence intervals for the variance, and the latter is based on generalized confidence intervals using a Monte-Carlo algorithm.

In the simulation study we compare the achieved guarantee, i.e. the probability that a coverage of β is achieved, and the expected interval's length for different distributions. In particular, we simulated 100 independent data sets from a given distribution. For each data set we derived the interval estimates via the three different procedures. Then a large number of values (100,000) from the same distribution are simulated and the proportion of variables falling within the intervals is evaluated, i.e. is the coverage of each interval. The 'achieved guarantee' denotes the proportion of intervals with a coverage exceeding the nominal value β. This number is supposed to be close to the given value of γ. Moreover, the expected length and the standard deviations are computed as in the case of mean coverage tolerance intervals.

We first studied the performance of these tolerance intervals for normally distributed data (see Table 3). Even though Hoffman's method has been specifically designed for the gaussian case, it does not perform well in practice: the intervals are systematically too short and, hence, do not attain the required guarantee γ. For Liao's and the bootstrap-*t* methods, the attained guarantee is acceptable. Based on empirical evidence, one may claim that intervals obtained via Liao's method are shorter than the bootstrap intervals. In the normal situation Liao and Iyer's method thus provides the most appropriate intervals, nevertheless the bootstrap-*t* also gives acceptable results.

For the one-way random effects model based on Pareto distributions with location parameter $x_m = 1$ and $k = 3$ for both effect and residual (see Table 4), we observe that neither Hoffman's nor Liao's method is well performing; a result that is

Table 5
Experimental data used to illustrate the calculation of the accuracy profile

Day	Level	Nicotinamide				Nicotinic acid						
		Calibration		Validation (milk A)		Calibration		Validation (milk A)		Validation (milk B)		
		Concentration (mg/L)	Peak area (AU)	Added conc. (mg/L)	Peak area (AU)	Concentration (mg/L)	Peak area (AU)	Added conc. (mg/L)	Peak area (AU)	Added conc. (mg/L)	Peak area (AU)	Peak area (corrected) (AU)
1	1	0.4	22.68	0.4	14.37	0.2	11.56	0.2	5.78	0.2	6.00	12.00
1	1	0.4	23.15	0.4	13.84	0.2	13.80	0.2	6.90	0.2	7.10	14.19
1	1	0.4	22.58	0.4	11.15	0.2	13.25	0.2	6.63	0.2	6.93	13.86
1	2	2.0	137.95	2.0	117.61	2.0	131.37	2.0	77.08	2.0	68.09	136.18
1	2	2.0	136.56	2.0	117.70	2.0	137.49	2.0	71.96	2.0	70.48	140.96
1	2	2.0	100.73	2.0	112.80	2.0	134.81	2.0	68.85	2.0	70.53	141.06
1	3	4.0	280.44	4.0	251.13	4.0	272.19	4.0	145.61	4.0	141.97	283.93
1	3	4.0	270.46	4.0	252.75	4.0	259.97	4.0	128.07	4.0	129.64	259.28
1	3	4.0	263.13	4.0	237.95	4.0	271.56	4.0	138.36	4.0	140.59	281.17
2	1	0.4	22.51	0.4	14.57	0.2	10.92	0.2	5.46	0.2	5.69	11.37
2	1	0.4	23.19	0.4	12.07	0.2	12.38	0.2	6.19	0.2	6.61	13.22
2	1	0.4	21.79	0.4	13.56	0.2	13.62	0.2	6.81	0.2	6.91	13.81
2	2	2.0	135.94	2.0	120.27	2.0	112.73	2.0	61.42	2.0	57.15	114.30
2	2	2.0	131.35	2.0	117.96	2.0	119.05	2.0	65.42	2.0	61.00	121.99
2	2	2.0	118.58	2.0	121.44	2.0	120.70	2.0	60.81	2.0	62.50	124.99
2	3	4.0	275.32	4.0	244.56	4.0	234.72	4.0	124.20	4.0	121.65	243.29
2	3	4.0	273.52	4.0	246.10	4.0	243.12	4.0	115.94	4.0	127.01	254.02
2	3	4.0	220.90	4.0	252.90	4.0	248.44	4.0	123.03	4.0	126.98	253.95
3	1	0.4	22.60	0.4	14.47	0.2	12.89	0.2	6.45	0.2	6.51	13.02
3	1	0.4	23.17	0.4	12.96	0.2	13.63	0.2	6.81	0.2	7.00	14.00
3	1	0.4	22.18	0.4	12.35	0.2	13.59	0.2	6.79	0.2	6.99	13.99
3	2	2.0	136.94	2.0	118.94	2.0	133.61	2.0	77.24	2.0	69.99	139.98
3	2	2.0	133.95	2.0	117.83	2.0	135.84	2.0	67.91	2.0	70.08	140.16
3	2	2.0	109.65	2.0	117.12	2.0	136.48	2.0	79.97	2.0	70.34	140.67
3	3	4.0	277.88	4.0	247.84	4.0	268.89	4.0	147.28	4.0	139.25	278.49
3	3	4.0	271.99	4.0	249.42	4.0	272.53	4.0	147.76	4.0	142.57	285.13
3	3	4.0	242.02	4.0	245.43	4.0	270.72	4.0	142.58	4.0	140.06	280.13

All results are expressed as mg/L for the concentration and dimensionless peak area ratio (Arbitrary Unit AU) for the analytical response. Intermediate precision condition is here different days of experiments within a single laboratory.

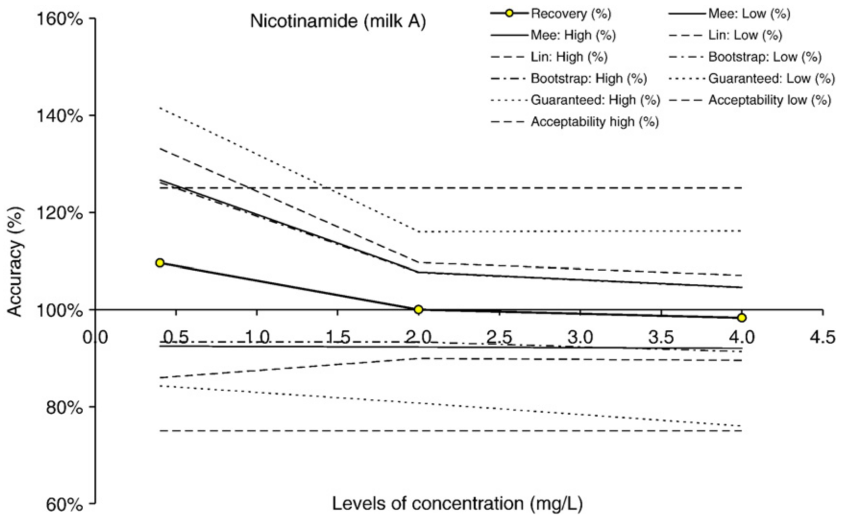

Fig. 3. Illustration of the different tolerance intervals computed from nicotinamide data. Mee and bootstrap algorithm give very comparable intervals and the method can be concluded as valid between 0.4 and 4.0 mg/L because these tolerance intervals are included within the acceptability interval.

not astonishing since those methods are developed for the normal situation. The results of Liao's method for the normal mixture model (Table 4) are relatively good which is due to the fact that the normal mixture distribution resembles the normal distribution. However, the nonparametric bootstrap method always provides good results, i.e. intervals of short length and a coverage that assures the required guarantee γ. However, the results are not as good as in the case of the mean coverage tolerance intervals. The reason might be that for this method a larger number of observations is necessary.

3. Result and discussion

3.1. Accuracy profiles of vitamin B3 in dairy products

3.1.1. Accuracy profile

A recently developed new definition of a valid analytical method is based on the intuition that a well performing method quantifies accurately and with reliability the unknown quantities that the laboratory will have to determine. Depending upon the application the analyst defines the domain of acceptable measurements by acceptance limits. Typically one considers a symmetric interval, at the original data scale, around the true concentration level μ, i.e. $(\mu-\lambda,\ \mu+\lambda)$. Second, the analyst determines the maximal risk of the occurrence of an unacceptable measurement that he is willing to take by fixing a confidence level, say β. So according to this definition a method is declared valid if the risk of having an unacceptable measurement is at most $1-\beta$.

One way to verify if the definition holds is to determine the accuracy profile of the analytical procedure. The accuracy profile consists of intervals containing a proportion β of the probability distribution of the measurements for each concentration level covering the assumed domain of application of the method. As a result the method is valid for the concentration levels where the accuracy profile is entirely included within the acceptance limits.

Unfortunately, the accuracy profile depends upon the unknown probability distribution of the measurements. So statistical estimates are required to derive an approximation of the accuracy profile. Indeed

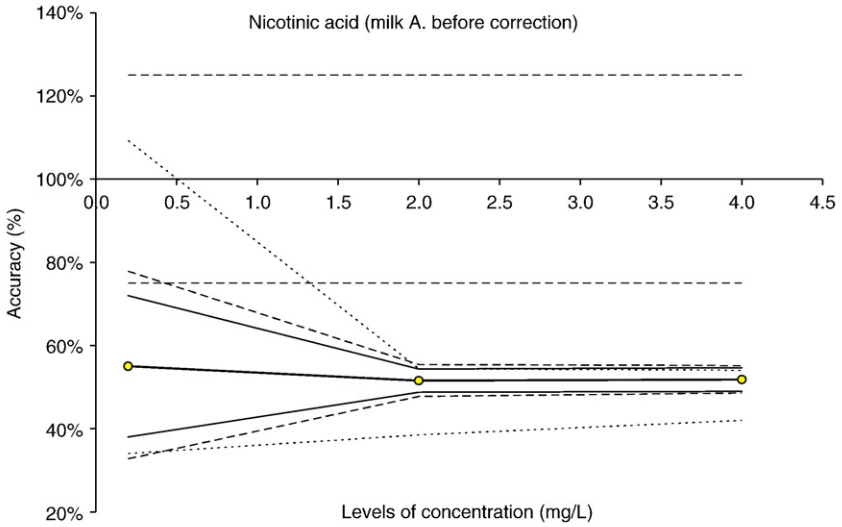

Fig. 4. Tolerance intervals in the case of a strong matrix effect (recovery yield around 52%) as observed for nicotinic acid. The method is not valid.

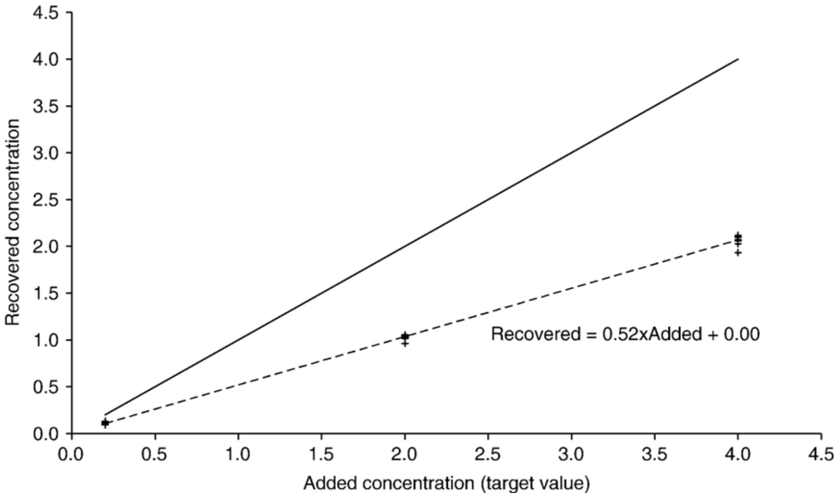

Fig. 5. Verification of the linearity (of accuracy). The fitted linear relationship can be used to compute a correction factor of 1/0.52 2.0.

tolerance intervals can be used. As we will see in the following example it makes a great difference whether one uses the mean coverage or the guaranteed coverage tolerance interval. The parameter γ indicates that the accuracy profile is included in the computed guaranteed coverage tolerance interval with probability γ, so for a value of γ close to one the interval is naturally rather large. However, the mean coverage tolerance interval has the property that if one repeats the computation for different data sets one obtains in average an interval that coincides with the accuracy profile. These two different interpretations of tolerance intervals must be kept in mind when interpreting the different estimates of the accuracy profile. The data of the example hereafter presented were recently published in [20] with more specific details about accuracy profile methodology.

3.1.2. Analytical method

The determination of vitamin B3 (or vitamin PP or niacin) in foods is based on two chemical forms: nicotinamide (natural vitamer) and nicotinic acid (synthetic form used for food fortification). The total concentration is expressed by the sum of the concentrations of the two analytes. The herein-presented study concerns milk whose concentration of vitamin B3 varies between 0.6 and 1.0 mg/L in natural products. Due to thermal treatments and authorized supplementation, the concentration of market products rather ranges between 0.4 and 2 mg/L. So in order to reasonably cover the application field, an analytical method should be valid for the range from 0.2 to 4.0 mg/L.

The official method for the determination of vitamin B3 is based upon a delicate microbiological technique with high uncertainty (40%) referenced as AOAC 94-413 (1990). This validation study aims at checking whether HPLC is acceptable as an alternative method. For this reason the acceptability limits have been set at ±25%, which is clearly more demanding than the usual performances of the reference method which stand rather around ±40%.

Nicotinic acid and nicotinamide are simultaneously extracted from milk by an acid hydrolysis in HCl 0.1 M at 100 °C during 1 h. Analyte separation is performed on a LiChrospher 60 RP Select B column (5 m, 4 mm diameter, 250 mm length) using an eluent system comprising 0.07 mol/L phosphate buffer containing 0.075 mol/L hydrogen peroxide and 5.10^{-6} mol/L copper sulphate. Detection is preceded by a post-column photochemical derivatization carried out with a UV lamp, eliminating the absorption line at 254 nm. Detectable derivatives are measured by a spectrofluorimetric detector at 380 nm after excitation at 322 nm. Standard calibration is obtained by passing

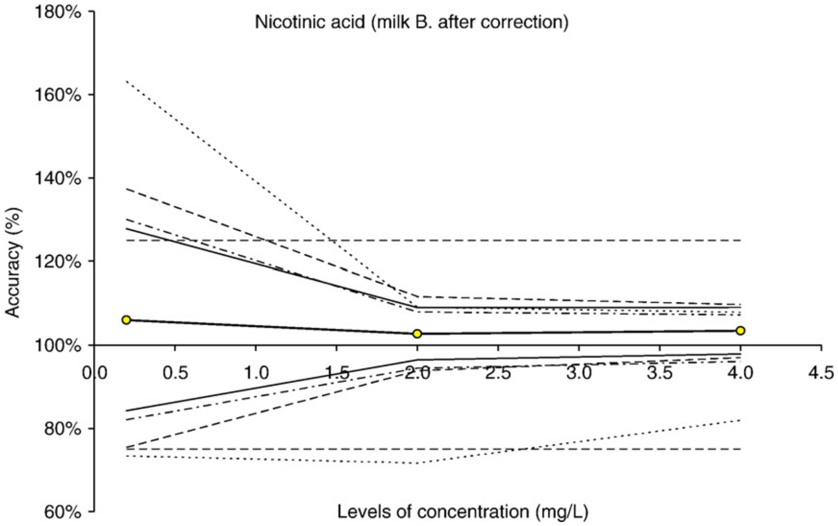

Fig. 6. Tolerance intervals computed from data of milk B after correction. The method can be decided as valid between 0.3 and 4.0 mg/L.

Table 6
Limits of the Mee's tolerance interval for nicotinamide data

	Levels		
Number of replicates	3	3	3
Theoretical concentration (mg/L)	0.4	2.0	4.0
Recovered concentration (mg/L)	0.4383	2.0004	3.9328
Intermediate precision standard deviation	0.0292	0.0572	0.1101
Lower limit of the tolerance interval	0.3700	1.8460	3.6802
Upper limit of the tolerance interval	0.5066	2.1549	4.1855

known amounts of amounts of nicotinic acid and nicotinamide in the chromatographic system. All computations are carried out by the software R interfaced with Excel using the R-COM extension package.

3.1.3. Experimental design

For both calibration and validation data, the experimental design consists in 27 trials. In particular, measurements for three concentration levels covering the validation range have been carried out on 3 days with three repetitions each (3 × 3 × 3). This choice is motivated on the one hand by an economically acceptable total number of trials, and on the other hand by the operating procedure that recommends a minimum of three concentration levels for calibration.

Given that adapted reference material does not exist, spiked standards on two milk samples, say milk A and milk B, are carried out. Here two different samples are necessary since a significant matrix effect for nicotinic acid has been observed during the method development. No matrix effect was observed for nicotinamide. Altogether, three series of data were collected:

1. Calibration etalons obtained by dissolving known amounts of pure nicotinamide and nicotinic acid in distilled water;
2. Spiked (with nicotinamide and nicotinic acid) calibration standards in milk A;
3. Spiked (only with nicotinic acid) validation standards in milk B.

Table 5 presents the complete experimental design and all collected data.

3.1.4. Nicotinamide

For nicotinamide validation data are only measured on milk A. According to [1], firstly calibration data are used to estimate the recovered concentrations from the validation data by inverse-prediction. In the present application, applicable calibration model was the straight line and was calculated by simple ordinary least-squares regression. Then mean coverage tolerance intervals are computed using the three different methods presented in the previous section; the parameter β is chosen at 0.90. Also the guaranteed coverage tolerance limits are calculated by using the bootstrap method where $\beta = \gamma = 0.90$. All intervals are illustrated in Fig. 3. In the case of β-tolerance intervals we observe that both bootstrap and Mee's intervals are fairly close one to another, but Lin and Liao's interval is larger. This is in accord with our simulation results where Lin and Liao's method often provides larger intervals than the other methods since their intervals tend to contain more than the required β% of the mass of the distribution. As well the guaranteed coverage tolerance interval is much larger than all β-tolerance intervals.

Now the analytical method is declared valid where tolerance intervals fall within the acceptance limits. Hence, method validation depends strongly on the used method for computing tolerance limits. And as the analyst might be interested in validating the analytical method on a large range of concentration levels, he will tend to use either the bootstrap or Mee's method since they provide the narrowest intervals for comparable coverage levels.

Furthermore, the lowest point where tolerance interval cuts the acceptability limits can be interpreted as the limit of quantification (LOQ). From Fig. 3 it can be deduced that LOQ for nicotinamide is close to 0.4 mg/L and the validated domain ranging from 0.4 up to 4.0 mg/L. So, in order to prove a small LOQ, it is also necessary that tolerance intervals are narrow. Therefore, when (β, γ)-tolerance limits are included within the acceptance limits one can have a good guarantee that for this domain if the validation study was replicated after some time, the conclusions would remain the same, and the validation domain be kept unchanged. This is an important complementary feature, whereas the reliability of a validation study always remains questionable for the analyst and the β-tolerance limits alone do not provide confirmatory elements.

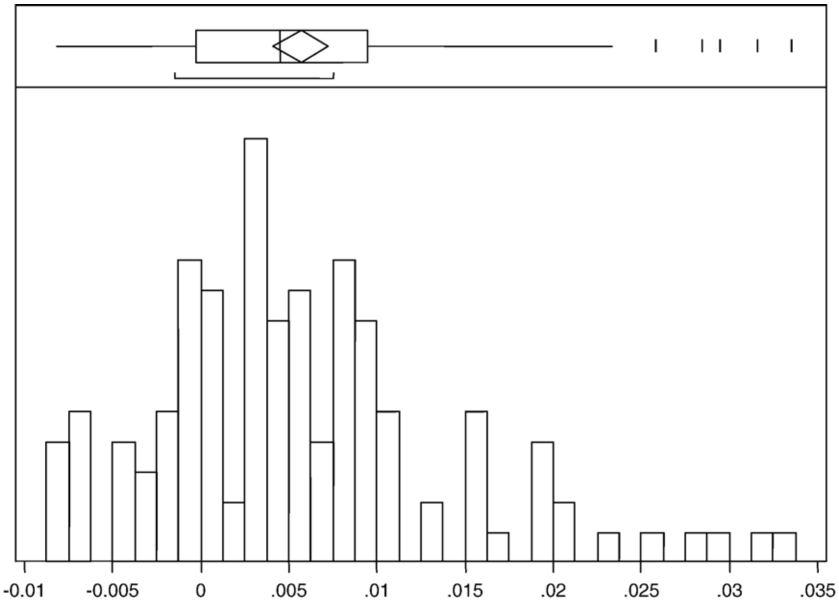

Fig. 7. Distribution of relative differences between the reference value and the bootstrap values for the nicotinamide study.

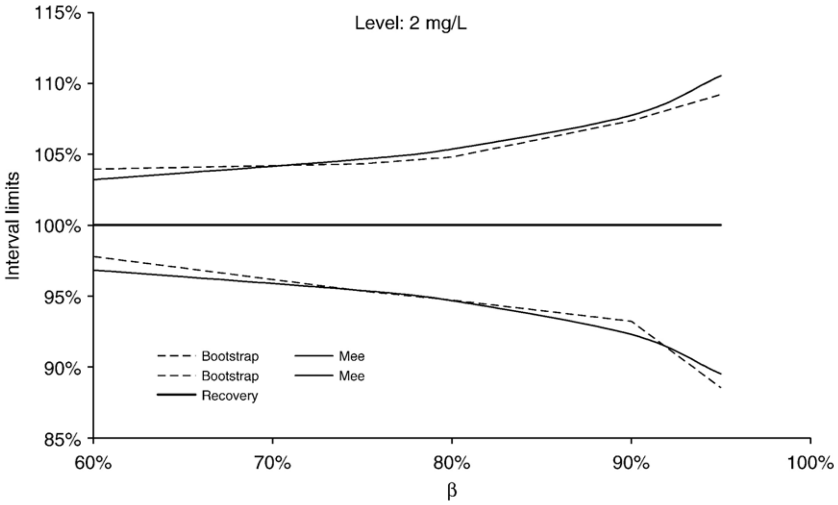

Fig. 8. Comparative evolution of Mee and bootstrap tolerance intervals as a function of at level of nicotinamide study.

3.1.5. Nicotinic acid

3.1.5.1. Estimation of the recovery for nicotinic acid. For nicotinic acid, accuracy profiles are computed on the milk A data. We stated a matrix effect that is confirmed in Fig. 4 by the high number of unacceptable results. None of the intervals are included within the acceptability limits and one may conclude that the method is not valid at all. However, the average recovery is about 52%. In order to correct the matrix effect we proceed as follows: i) use the milk A data to compute a correction factor; ii) correct the milk B data by applying this correction factor and compute the tolerance intervals based on the corrected data.

When considering the average recovery yield, it is around 52%. By simply relating the added concentration to the recovered, it is possible to: i) verify the linearity of the method in terms of bias; ii) define a correction factor as the inverse of the slope of the least-squares regression line. Fig. 5 illustrates this technique and gives as a regression line:

$$[\text{Recovered}] = 0.52 \times [\text{Added}] + 0.00.$$

Here the correction factor yields 1/0.52 = 1.92, however, other than LS regression techniques can be applied and in order to simplify, it was decided to use a rounded correction factor of 2.0.

3.1.5.2. Corrected data profile. After correction of the milk B data by this factor, as reported in Table 5, the various accuracy profiles are derived with the same β and γ parameter values as previously, see Fig. 6. We observe that the results are comparable to those of the nicotinamide data, even though there is a slight shift.

3.2. Discussion and conclusions

The major difference between the bootstrapped tolerance interval and the traditional approach proposed by Mee consists in the fact that, based on normal distribution assumption Mee's interval leads to analytical formulae to compute limits while bootstrap uses an algorithmic procedure. Therefore, when considering the possible differences between repeated calculations of bootstrapped tolerance interval it is interesting to verify whether it presents an important variability. Using the nicotinamide data, we have repeated 60 times the calculations of the bootstrap upper and tolerance interval limits for β=90% and compared the values obtained to the values of Mee used as references. Table 6 gives these Mee values.

When computing the relative differences between the reference values of Table 6 and the bootstrap values we get 120 data which are illustrated on the histogram of Fig. 7. The average relative difference is positive and equal to 0.57% with a maximum of 3.36% and a minimum of −0.82%. This explains why both TI are always very close. This also means that bootstrap TI are a little bit wider and more conservative than Mee's.

On the other hand, the choice of a reasonable value for β appears as a critical issue for the analysts. Therefore, the influence of the value used for β was also studied. Both Mee and bootstrap TI were calculated for values of β ranging from 60% to 95% using the nicotinamide data at level 2 mg/L. Fig. 8 illustrates the evolution of the β–TI and the role of this parameter. This figure demonstrates the strong influence of β and the need for analytical chemists to define guidelines. The value of β must be related to the risk for end-users to get unacceptable results and the risk for the laboratory to have to replicate a measurement. It does not mean that an unacceptable result is always wrong, whereas β does not represent a risk of error but the proportion of future results obtained following the operating procedure. It seems evident that some international harmonization will necessary define either consensus values for β and recognized values for the acceptability limit λ in respect of the analytical technique and its application domain.

The question of validating methods is changing and, in the field of food chemistry, the recent discussion initiated by the Commission of Codex Alimentarius (http://www.fao.org/docrep/meeting/005/X0830E/x0830e0a.htm) and echoed by the US FDA or the European Commission about the so-called "criteria approach" is an indication of these changes. The concept of accuracy profile and the statistical notion of tolerance interval are attempts to participate to this evolution.

References

[1] P. Hubert, J.J. Nguyen-Huu, B. Boulanger, E. Chapuzet, P. Chiap, N. Cohen, P. Compagnon, W. Dewe, M. Feinberg, M. Lallier, M. Laurentie, N. Mercier, G. Muzard, C. Nivet, L. Valat, Harmonization of strategies for the validation of quantitative analytical procedures: a SFSTP proposal — part I, Journal of Pharmaceutical and Biomedical Analysis 36 (3) (2004) 579–586.

[2] M. Feinberg, B. Boulanger, W. Dewé, P. Hubert, New advances in method validation and measurement uncertainty aimed at improving the quality of chemical data, Analytical and Bioanalytical Chemistry 380 (3) (2004) 502–514.

[3] M. Feinberg, M. Laurentie, A global approach to method validation and measurement uncertainty, Accreditation and Quality Assurance 11 (2006) 3–9.

[4] S.S. Wilks, Determination of sample sizes for setting tolerance limits, Annals of Mathematical Statistics 12 (1) (1941) 91–96.

[5] R.W. Mee, β-expectation and β-content tolerance limits for balanced one-way ANOVA random model, Technometrics 26 (3) (1984) 251–254.

[6] T.Y. Lin, C.T. Liao, A β-expectation tolerance interval for general balanced mixed linear models, Computational Statistics & Data Analysis 50 (2006) 911–925.

[7] A. Wald, J. Wolfowitz, Tolerance limits for a normal distribution, Annals of Mathematical Statistics 17 (2) (1946) 208–215.

[8] W.G. Howe, Two-sided tolerance limits for normal populations — some improvements, Journal of American Statistical Association 64 (1969) 610–620.

[9] W.A. Wallis, Tolerance intervals for linear regression, Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, 1951, pp. 43–52.

[10] A. Weissberg, G. Beatty, Tables of tolerance–limit factors for normal distributions, Technometrics 2 (4) (1960) 483–500.

[11] D. Hoffman, R. Kringle, Two-sided tolerance intervals for balanced and unbalanced random effects models, Journal of Biopharmaceutical Statistics 15 (2005) 283–293.

[12] C.T. Liao, H.K. Iyer, A tolerance interval for the normal distribution with several variance components, Statistica Sinica 14 (2004) 217–229.

[13] R.W. Mee, D.B. Owen, Improved factors for one-sided tolerance limits for balanced one-way ANOVA random model, Journal of American Statistical Association 78 (384) (1983) 901–905.

[14] E. Paulson, A note on control limits, Annals of Mathematical Statistics 14 (1943) 90–93.

[15] B. Efron, Bootstrap methods: another look at the jackknife, Annals of Statistics 7 (3) (1979) 1–26.

[16] J. Shao, D. Tu, The Jacknife and Bootstrap, Springer Series in Statistics, Springer, 1993.

[17] B. Efron, R.J. Tibshirani, An Introduction to the Bootstrap, Monographs on Statistics and Applied Probability, Chapman & Hall, 1993.

[18] R. Wehrens, H. Putter, L. Buydens, The bootstrap: a tutorial, Chemometrics and Intelligent Laboratory Systems 54 (2000) 35–52.

[19] L.T. Fernholz, J.A. Gillespie, Content-corrected tolerance limits based on the bootstrap, Technometrics 43 (2) (2001) 147–155 (9).

[20] M. Feinberg; Validation of analytical methods based on accuracy profiles. J. Chromatogr. A (in press), doi:10.1016/j.chroma.2007.02.021.