Available online at www.sciencedirect.com
SCIENCE DIRECT®
Talanta 65 (2005) 1022–1030
Talanta
www.elsevier.com/locate/talanta

Practical digest for evaluating the uncertainty of analytical assays from validation data according to the LGC/VAM protocol

A. Gustavo González*, M. Ángeles Herrador, Agustín G. Asuero

Department of Analytical Chemistry, University of Seville, E-41012 Seville, Spain

Received 1 June 2004; received in revised form 17 August 2004; accepted 26 August 2004
Available online 29 September 2004

Abstract

The estimation of the measurement uncertainty of analytical assays based on the LGC/VAM protocol from validation data is fully revisited and discussed in the light of the study of precision, trueness and robustness.

Keywords: Uncertainty measurement; LGC/VAM protocol; Precision; Trueness; Robustness

1. Introduction

Two of the best known approaches for estimating the uncertainty of analytical procedures are the ISO Guide to the expression of uncertainty measurement, also known as the GUM or "*bottom-up*" approach [1], and the Analytical Methods Committee inter-laboratory ("*top-down*") approach [2]. The GUM approach was originally intended for physical measurements, and although it has been adapted by EURACHEM to chemical measurements [3], it remains being tedious, time-consuming and analytically unrealistic [4]. GUM principles are significantly different from current procedures used in analytical chemistry, dealing with precision or recovery data [5]. Besides, the application of the "bottom-up" approach in routine analysis involves significant costs in time and effort, which often prevents its widespread application [6].

On the other hand, the "*top-down*" approach also exhibits several drawbacks. The first one is that it can only be applied when inter-laboratory exercises are available. In this approach, any individual laboratory is considered as a random variable and consequently, systematic and random errors within the laboratory become only random errors from this high level perspective. Maroto et al. [6] pointed out that within this approach, if we calculate the uncertainty of the results obtained in routine analysis by a given laboratory, the findings could be quite misleading. Thus, if the laboratory bias is very small, the uncertainty may be considerably underestimated. Conversely, if the laboratory has a large bias, the uncertainty could be largely overestimated.

Besides from these two main approaches, there are others where the measurement uncertainty is evaluated from the information extracted at the validation stage. This new paradigm is proposed by Ellison and Barwick [7], Ellison and Williams [8] and Maroto et al. [9]. The essential of this approach was elaborated by Barwick and Ellison within a LGC/VAM project [10] and constitutes an excellent alternative for evaluating the uncertainty of analytical procedures.

The aim of this paper is to provide to the routine and/or R&D analyst with a practical digest, clearly outlined, to evaluate the uncertainty of analytical assays from the validation data. Some considerations may deviate from the LGC/VAM protocol but are supported by theoretical dictates or researchers experience.

As Massart et al. [11] pointed out in their excellent handbook, there are three "golden rules" for method validation that can be summarized as follows:

* Corresponding author. Tel.: +34 954557173; fax: +34 954557168.
E-mail address: agonzale@us.es (A.G. González).

0039-9140/$ – see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.talanta.2004.08.038

(1) The analytical procedure has to be validated as a whole, with inclusion of sample treatments prior to the analysis.
(2) The analytical procedure has to be validated covering the full range of analyte concentrations specified in the method scope.
(3) The analytical procedure has to be validated for each kind of matrix where it will be applied.

These golden rules are the keys for elaborating the suitable scheme to be applied for evaluating uncertainty of the analytical procedure.

The ultimate objective of the validation of an analytical procedure is to guarantee that every single measurement that will be performed later in routine analysis will be close enough to the unknown true value of the sample [12]. Consequently, the objectives of the validation are not simply to obtain estimates of bias and precision but also evaluate these risks, which can be expressed by the uncertainty related to the result. As a part of the validation objective, the establishment of the *figures of merit* or *performance characteristics* and the identification and quantitation of the most contributing experimental factors is essential. From this perspective, the study of the main features derived from the modern meaning of accuracy is of utmost importance. According to Garcia-Campaña et al. [13] these features are: inertia (robustness), trueness and precision. This point of view is interesting because these terms just correspond to the three main studies considered by the LGC/VAM protocol [10] to estimate the measurement uncertainty of analytical assays: the precision study, the trueness study and the robustness study. As both precision and trueness are assessed within a single laboratory, the uncertainty due to laboratory transfer should be taken into account. This can be attained with the robustness study, which considers the changes in the variables of the analytical procedure (called factors) expected in a transfer among laboratories. According to the International Conference on Harmonization (ICH Q2B document) [14], the robustness of an analytical procedure is a measure of its capacity to remain unaffected by small, but deliberate variations in method parameters and provides an indication of its reliability during normal usage. Robustness tests can be considered to be intra-laboratory simulations of inter-laboratory studies, if the alterations introduced are suitably selected. This allows to a further approach developed by Hund et al. [15], which directly estimates the measurement uncertainty from the robustness study. They advise that usually, robustness tests that yields significant effects for the measurand leads to a further optimization of the method and, consequently, uncertainty evaluation should only be performed after the method has shown to be robust. This deviates from the LGC/VAM approach, which provides different algorithms for significant and no significant factors. In the following, the LGC/VAM approach is outlined and revised.

Consider Z_{found} the analyte concentration obtained by applying the analytical procedure to a sample. If in the trueness study, there are both proportional bias (recovery R significantly different from 1) and constant bias (offset δ_c significantly different from 0), then the corrected concentration value for the analyte will be:

$$Z = \frac{Z_{found} - \delta_c}{R} \tag{1}$$

If we are interested to evaluate the uncertainty of Z, the application of the variance propagation law leads to:

$$\frac{u(Z)}{Z} = \sqrt{\frac{u^2(Z_{found}) + u^2(\delta_c)}{(Z_{found} - \delta_c)^2} + \frac{u^2(R)}{R^2}} \tag{2}$$

Now, the term $u(Z_{found})/(Z_{found} - \delta_c)$ can be ascribed to the relative dispersion mainly due to random errors. Both the precision study and the robustness study account for this spread. If the relative uncertainties for intermediate precision and for the robustness study are, respectively, $R.S.D._I$ and $R.S.D._{rob}$, and can be considered fixed for a future determination, then the following relationship can be easily derived:

$$\frac{u(Z)}{Z} = \sqrt{R.S.D._I^2 + R.S.D._{rob}^2 + \frac{u^2(\delta_c)}{(Z_{found} - \delta_c)^2} + \frac{u^2(R)}{R^2}} \tag{3}$$

If the trueness study concludes that there is neither constant nor proportional method bias, then $R = 1$, $\delta_c = 0$, $Z = Z_{found}$ and Eq. (3) becomes

$$\frac{u(Z)}{Z} = \sqrt{R.S.D._I^2 + R.S.D._{rob}^2 + \frac{u^2(\delta_c)}{Z^2} + u^2(R)} \tag{4}$$

Eq. (3) or (4) indicates how the uncertainty in the analyte concentration can be suitably evaluated from the precision study ($R.S.D._I$), the trueness study ($u(\delta_c)$ and $u(R)$) and the robustness study ($R.S.D._{rob}$).

2. Intermediate precision study

Following the validation golden rules, the analytical procedure should be validated separately for each class of matrix considered, covering the full range of analyte concentrations. Accordingly, the suitable way to perform the intermediate precision study is the consideration of a single sample matrix and a range of analyte concentrations. It is recommended at least three concentration levels (low, medium and high) covering the dynamic working range, with a number of replicates at each concentration. The ICH Q2B document recommends three replicates [16] and the FDA document on bioanalytical validation considers five replications [17]. Thus, a series of replications within 3–5 is advised. Calculations of intermediate precision must be carried out on results instead of responses.

If samples containing the suitable analyte concentrations are not available, spiked samples can be prepared. In case of

pharmaceutical formulations (or other manufactured products), where the "placebo" is available, the precision study should be carried out from spiked placebos. We refer to the samples used in the intermediate precision study as "control samples". Control samples must be stable, homogeneous and as similar as possible to the future samples to be analysed. Control samples can be considered as validation standards and they represent in the validation phase, the future samples that the analytical procedure will have to quantify. Each validation standard must be prepared and treated independently as a future sample. The independence is essential for a good estimation of the between-series variance. Indeed, the analytical procedure is not developed to quantify in routine with the same operator and on the same equipment a single sample unknown during a day but a very large number of samples through time and so, implying often several operators and several equipments.

Consider m concentration levels and n replicates for each level. As indicated above, typical recommended values are $m = 3$ (low, medium and high) and $3 \leq n \leq 5$ in all cases. The replication measurement should be suitably designed in order to obtain time-different intermediate precision estimates. Thus, control samples should be analysed n times over different days to obtain the intermediate precision values according to the USP requirements [18]. If the analytical procedure is not direct (gravimetry, titrimetry, coulometry or isotope dilution mass spectrometry), the calibration curve should be established before the control sample analysis from freshly prepared reagent solutions. From the replication, the standard deviation, s, and the relative standard deviation, R.S.D., are calculated for each concentration level.

In some instances, the intermediate precision seems to be independent of the analyte concentration. Hence, the variances obtained at each concentration level should be uniform. There are several tests described for testing the homogeneity of variances. The Cochran's assay [19] is very suitable when the number of observations is equal for all concentration levels (as commonly carried out by following the design with m and n fixed) but otherwise, the Levene's test is an interesting alternative [12]. According to the Cochran assay, the value:

$$C = \frac{s^2_{\max}}{\sum_{i=1}^{m} s_i^2} \tag{5}$$

is computed (where s_i^2 is the variance at the concentration level i and $s^2_{\max}$ is the maximum value) and is compared with the critical tabulated value $C_{\text{crit}}(m, n, P)$, P being the selected confidence level, commonly 95%. If $C \leq C_{\text{crit}}$, then there is no significant difference between the variance and therefore, the intermediate precision is a constant amount, independent on analyte concentration. In this case, an estimate of the pooled standard deviation is obtained:

$$s^2_{\text{pool}} = \sqrt{\frac{\sum_{i=1}^{m}(n_i - 1)s_i^2}{\sum_{i=1}^{m}(n_i - 1)}} \tag{6}$$

In many cases, the intermediate precision seems to be proportional to the analyte concentration. Thus, the relative variances, R.S.D.^{2}, should be constant. The Cochran's test can be applied in a similar way for R.S.D.:

$$C = \frac{\text{R.S.D.}^2_{\max}}{\sum_{i=1}^{m}\text{R.S.D.}_i^2} \tag{7}$$

If $C \leq C_{\text{crit}}$, then there is no significant difference between the relative variance and therefore, the intermediate precision is proportional to the concentration of the analyte. A pooled value can be estimated:

$$\text{R.S.D.}_{\text{pool}} = \sqrt{\frac{\sum_{i=1}^{m}(n_i - 1)\text{R.S.D.}_i^2}{\sum_{i=1}^{m}(n_i - 1)}} \tag{8}$$

Sometimes, neither of the two situations considered above applies. Then it could be necessary to evaluate separate intermediate precision estimation at different concentration ranges.

The final result of the intermediate precision study is a relative standard deviation value for future real samples (R.S.D.$_{\text{I}}$). Thus, if the intermediate precision is non-depending on the analyte concentration on the full concentration range, the expected result is:

$$\text{R.S.D.}_{\text{I}} = \frac{s_{\text{pool}}}{Z} \tag{9}$$

where Z is the analyte concentration found in the analysed sample. If, conversely, the intermediate precision is proportional to the analyte concentration in the whole working range, the expected value is:

$$\text{R.S.D.}_{\text{I}} = \text{R.S.D.}_{\text{pool}} \tag{10}$$

When the intermediate precision must be estimated separately, the R.S.D.$_{\text{I}}$ value is computed from the standard deviation belonging to the concentration range that corresponds to the analyte concentration found in a real sample. For example, the whole concentration range $[Z_{\text{low}}, Z_{\text{high}}]$ is split into three intervals, where the standard deviation has been separately evaluated: s_1 belongs to the interval $[Z_{\text{low}}, Z_1]$, s_2 to $[Z_1, Z_2]$ and s_3 to $[Z_2, Z_{\text{high}}]$. If the concentration of analyte found in the real sample is $Z \in [Z_1, Z_2]$ then we get:

$$\text{R.S.D.}_{\text{I}} = \frac{s_2}{Z} \tag{11}$$

The intermediate precision seems effectively to be independent of the analyte concentration when the dosing range is narrow. When the range widens, the standard deviation increases and the R.S.D. decreases with the concentration, respectively.

3. Trueness study

The trueness study covers the uncertainties related to method bias. The LGC/VAM protocol shows several ways to estimate the uncertainty related to the trueness:

(1) the analysis of certified reference materials (CRM);
(2) the comparison with a reference method; and
(3) the recovery assays with spiked samples.

Unfortunately, within the field of environmental, toxicological and pharmaceutical analysis, neither CRMs nor alternate methods are available for new contaminants, toxic and drug-related analysis. Accordingly, the procedure of recovery assays from spiked samples is the considered here because it can be commonly applied anywhere. We consider the case of a non-direct analytical procedure that needs calibration. A standard calibration curve is obtained by plotting the signal or analytical response of different standard solutions of the analyte. Let us assume that the standard calibration relationship is linear within a given concentration range of analyte, and consequently, the analytical response (Y) can be expressed as a linear straight line against the analyte concentration (Z):

$$Y = a_{SC} + b_{SC}Z \tag{12}$$

The subscript SC refers to the standard calibration, established by preparing different (generally aqueous) standard solutions of the analyte from a primary standard if possible. In the assessment of trueness, both proportional and constant method bias have to be evaluated. If the placebo is not available, proportional bias is assessed from recovery assays and constant bias is estimated according to Youden's method [20].

3.1. Recovery assays on spiked samples

Consider the application of the analytical procedure to a dissolved test portion of a given sample within the linear working range. Assuming that (i) the matrix cannot contribute to the signal as an interfering agent [21] and (ii) there is no interaction between the analyte and the matrix, the analytical response (Y) can be now expressed as:

$$Y = A + BZ \tag{13}$$

where A and B are sample constants. A is a constant that does not change when the concentration of the analyte and/or the sample change [22]. It is called the "true sample blank" [23] and can be evaluated from the Youden's sample plot [20,24–25]. B is the fundamental term that justifies the analytical procedure and is directly related to the analytical sensitivity [26]. If both constant and proportional bias method are absent, then $A = a_{SC}$ and $B = b_{SC}$. In order to assess the absence of proportional bias, a homogeneous bulk spiked sample from a matrix, which contains the analyte, is used. The analyte have to be spiked at several concentration levels in order to cover the concentration range of the method scope. As in the precision study, a given homogeneous control sample is fortified at "m" levels (for instance, low, medium and high) and the analyses are performed in "n" ($3 \leq n \leq 5$) replicates by following a nested scheme as indicated in Fig. 1. The control sample is treated according to the analytical procedure, fortified by spiking a given amount i (from 1 to m) of the analyte with j (from 1 to n) replicates for each addition. The corresponding elemental recovery is calculated as:

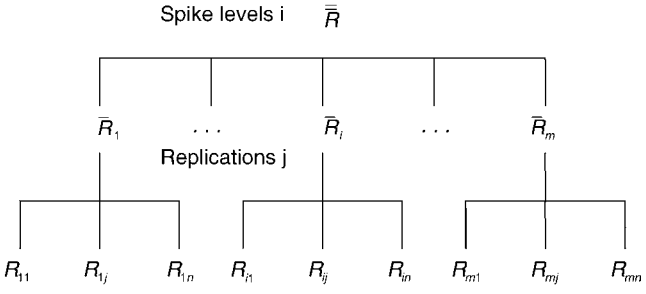

Fig. 1. Nested scheme for the recovery study.

$$R_{ij} = \frac{(C_{obs} - C_{native})_{ij}}{C_{spike,i}} = \frac{(\hat{C}_{spike,i})_j}{C_{spike,i}} \tag{14}$$

C_{native} is the concentration of the analyte in the unspiked control sample, C_{obs} the concentration of analyte in the spiked sample and C_{spike}, the concentration of the spiked analyte. The found difference $C_{obs} - C_{native} = \hat{C}_{spike}$ should be very close to the added amount C_{spike}, if the recovery is significant. Subscript i refers to the spiked amount and subscript j indicates replication of this amount.

The Standard Addition Method (SAM) can be suitably used to estimate the recovery of spiked samples [22,27]. Thus, for a spiked sample, Eq. (13) may be rewritten as:

$$Y = A + B(C_{native} + C_{spike}) = A + BC_{native} + BC_{spike}$$
$$= a_{SAM} + b_{SAM}C_{spike} \tag{15}$$

where, a_{SAM} and b_{SAM} are the intercept and slope of the SAM calibration straight line. An important requirement for this technique is that all solutions, unspiked and spiked test portions of the sample be diluted to the same final volume. By using this calibration function, from the analytical signal at each addition i and at each replicate j, Y_{ij}, the amount of analyte found is estimated as:

$$(\hat{C}_{spike,i})_j = \frac{Y_{ij} - a_{SAM}}{b_{SAM}} \tag{16}$$

and consequently from Eq. (14), the recovery for the j-replication of the i-spike can be obtained as:

$$R_{ij} = \frac{Y_{ij} - a_{SAM}}{b_{SAM}C_{spike,i}} \tag{17}$$

The uncertainty of R_{ij} can be obtained by applying the law of variance propagation andconsidering the covariance between

a_{SAM} and b_{SAM} [28]:

$$u^2(R_{ij}) = \frac{u^2(Y_{ij}) + u^2(a_{SAM})}{(b_{SAM}C_{spike,i})^2} + \left(\frac{Y_{ij} - a_{SAM}}{b_{SAM}^2 C_{spike,i}}\right)^2 \times u^2(b_{SAM}) + \frac{2(Y_{ij} - a_{SAM})}{b_{SAM}^3 C_{spike,i}^2}\mathrm{cov}(a_{SAM}, b_{SAM}) + \left(\frac{Y_{ij} - a_{SAM}}{b_{SAM}C_{spike,i}^2}\right)^2 u^2(C_{spike,i}) \quad (18)$$

$u^2(b_{SAM})$, $u^2(a_{SAM})$ are taken as the corresponding variances $s^2(a_{SAM})$ and $s^2(b_{SAM})$ of the SAM calibration straight line. The covariance is evaluated according to $\mathrm{cov}(a_{SAM}, b_{SAM}) = -\bar{C}_{spike}s^2(b_{SAM})$, and $\bar{C}_{spike}$ is taken as $\sum_{i=1}^{m} C_{spike,i}/m$ because the n replications cancel with the division by n.

The uncertainty for the i-level spiked amount, $u(C_{spike,i})$ can be easily calculated. A working standard solution of the analyte is prepared by weighing the required mass (w) from a primary standard of purity P (fraction scale) and dissolving it to a volume V_0. The i-level spiked concentration is carried out by dissolving the sample test portion in a volumetric flask of volume V and before filling to the mark, to add a volume v_i of the analyte working standard solution. Accordingly, the spiked concentration is:

$$C_{spike,i} = \frac{wPv_i}{V_0 V} \quad (19)$$

Thus, the law of variance propagation gives the corresponding uncertainty:

$$\frac{u(C_{spike,i})}{C_{spike,i}} = \sqrt{\frac{u^2(w)}{w^2} + \frac{u^2(P)}{P^2} + \frac{u^2(v_i)}{v_i^2} + \frac{u^2(V)}{V^2} + \frac{u^2(V_0)}{V_0^2}} \quad (20)$$

The remaining uncertainty values $u(w)$, $u(P)$, $u(v_i)$, $u(V)$ and $u(V_0)$ are estimated in the usual way by following the EURACHEM/CITAC guide [3]. These uncertainties are mainly type B ones with some exceptions such as in-house balance calibration or volume repeatability, and can be evaluated from confidence/tolerance intervals by assuming a given probability distribution function (normal, rectangular, triangular,...). The uncertainty associated to mass measurement, $u(w)$, is estimated from the calibration certificate supplied by a calibration laboratory. Generally, this certificate provides us with the relative uncertainty $u_{rel}(w)$ to be assigned to the value indicated by the weighing instrument [29]. Thus, the value of the uncertainty of our weight may be easily obtained in the conventional scale (without applying buoyancy corrections) as

$$u(w) = wu_{rel}(w) \quad (21)$$

The uncertainty of purity, $u(P)$ is estimated from the supplier's certificate tolerance of purity ($\pm a_P$) assuming a rectangular distribution:

$$u(P) = \frac{a_P}{\sqrt{3}} \quad (22)$$

When the certificate only indicates a purity interval [a_{min}, a_{max}],$a_P = (a_{max} - a_{min})/2$

If the purity is expressed as "...higher than..." a given value, this is taken as a_{min} and $a_{max} = 1$. For instance, if a reagent certificate indicates that the purity is higher than 99%, then in the fractional scale, $a_{min} = 0.99$, $a_{max} = 1.00$, $a_P = (1.00 - 0.99)/2 = 0.005$.

The uncertainty of volume measurements has three major influences and can be split into three contribution terms: calibration tolerance, volume repeatability and temperature effect. Thus, for a given volume measurement, V for instance, we have the following uncertainty contributions: $u(V_{cal})$, $u(V_{rep})$ and $u(V_{temp})$.

The calibration term comes from the manufacturer's tolerance in volume ($\pm a_{cal}$). Assuming a triangular distribution, this uncertainty is computed as:

$$u(V_{cal}) = \frac{a_{cal}}{\sqrt{6}} \quad (23)$$

The volume repeatability is a type A uncertainty obtained from a series of filling and weighing the volume V of the container (flask, pipette,...) and is expressed as the corresponding standard deviation:

$$u(V_{rep}) = s_{rep} \quad (24)$$

The temperature term arises from the difference between the calibration temperature (T_{cal}) of the glassware and the temperature of the laboratory (T_{lab}). The uncertainty of this effect is calculated from the difference of temperature ($\Delta T = |T_{lab} - T_{cal}|$), the coefficient of volume expansion for the glass ($\kappa = 2.1 \times 10^{-4}\,^{\circ}C^{-1}$) and the volume, V, assuming a rectangular distribution:

$$u(V_{temp}) = \frac{\kappa \Delta T V}{\sqrt{3}} \quad (25)$$

The final uncertainty for the volume is computed by combining the three separated contributions:

$$u^2(V) = u^2(V_{cal}) + u^2(V_{rep}) + u^2(V_{temp}) \quad (26)$$

This procedure is repeated for evaluating the other volume uncertainties $u(v_i)$ and $u(V_0)$. Once $u(C_{spike,i})$ is calculated according to equation (20), the uncertainty of the elemental recovery, $u(R_{ij})$ is computed from Eq. (18). Aside from this combined uncertainty, a pure repetition one, $u_{rep}(R_{ij})$ can be considered:

$$u_{rep}^2(R_{ij}) = \frac{\sum_{j=1}^{n}(R_{ij} - \bar{R}_i)^2}{n-1} \quad (27)$$

with

$$\bar{R}_i = \frac{1}{n}\sum_{j=1}^{n} R_{ij} \quad (28)$$

The uncertainty corresponding to these intermediate $\bar{R}_i$ values is given by

$$u^2(\bar{R}_i) = \frac{1}{n^2}\left[\sum_{j=1}^{n} u^2(R_{ij}) + u^2_{\text{rep}}(R_{ij})\right] \quad (29)$$

As above, another pure repetition term can be considered:

$$u^2_{\text{rep}}(\bar{R}_i) = \frac{\sum_{i=1}^{m}(\bar{R}_i - \bar{\bar{R}})^2}{m-1} \quad (30)$$

Commonly, the different intermediate recovery values are in good agreement and are combined to give a consensus $\bar{\bar{R}}$ value:

$$\bar{\bar{R}} = \frac{1}{m}\sum_{i=1}^{m} \bar{R}_i \quad (31)$$

Accordingly, the final uncertainty for the consensus recovery is:

$$u^2(\bar{\bar{R}}) = \frac{1}{m^2}\left[\sum_{i=1}^{m} u^2(\bar{R}_i) + u^2_{\text{rep}}(\bar{R}_i)\right] \quad (32)$$

Consensus recovery may be tested for significance [30] using the Student's *t*-test:

$$t = \frac{|\bar{\bar{R}} - 1|}{u(\bar{\bar{R}})} \quad (33)$$

According to the LGC/VAM protocol [10], if the degrees of freedom associated with the uncertainty of consensus recovery are known, t is compared with the two-tailed critical value, t_{crit}, for the appropriate number of degrees of freedom at 95% confidence. If $t \leq t_{\text{crit}}$, the consensus recovery is not significantly different from 1. Alternatively, instead of t_{crit}, the coverage factor k may be used for the comparison. Typical values are $k = 2$ or $k = 3$ for 95% or 99% confidence, respectively [31]. Thus:

- If $|\bar{\bar{R}} - 1|/u(\bar{\bar{R}}) \leq k$, the recovery is not significantly different from 1.
- If $|\bar{\bar{R}} - 1|/u(\bar{\bar{R}}) > k$, the recovery is significantly different from 1 and the analytical result must be corrected by $\bar{\bar{R}}$. In any case, the relative uncertainty for proportional bias from the trueness study is taken as $u(\bar{\bar{R}})/\bar{\bar{R}}$.

The slope of the SAM calibration (b_{SAM}) is an estimate of the product of the slope of the standard calibration (b_{SC}), which represents the sensitivity of the analytical method and the method recovery. Thus the following equation holds true:

$$b_{\text{SAM}} = \bar{\bar{R}} b_{\text{SC}} \quad (34)$$

3.2. Constant bias assessment

As mentioned above, in presence of sample matrix, the relationship between the analytical response and the analyte concentration is given by Eq. (13). The independent term "A" is the true sample blank because it is determined when both the native analyte and the matrix are present [27]. The SAM calibration indicated by Eq. (15) includes this term within the intercept ($a_{\text{SAM}} = A + b_{\text{SAM}} C_{\text{native}}$). The Youden's plot [20,24,25] consists of plotting the instrumental response (Y) against the amount of sample (the weight or volume of sample test portion to be dissolved up to the assay volume):

$$Y = A + b_{\text{Youden}} w_{\text{sample}} \quad (35)$$

The intercept of the plot is an estimation of the Total Youden Blank (TYB), which is the sum of the System Blank (SB) corresponding to the intercept of the standard calibration (a_{SC}) and the Youden Blank (YB) associated to the method constant bias [27,32]. Thus, we can equate TYB = A, SB = a_{SC} and and YB = $A - a_{\text{SC}}$. The method constant bias is defined as [27]:

$$\delta_{\text{c}} = \frac{\text{YB}}{b_{\text{SC}}} = \frac{A - a_{\text{SC}}}{b_{\text{SC}}} \quad (36)$$

The uncertainty of the constant bias can be obtained by the law of the variance propagation:

$$u(\delta_{\text{c}}) = \sqrt{\frac{u^2(A)}{b^2_{\text{SC}}} + \frac{u^2(a_{\text{SC}})}{b^2_{\text{SC}}} + \frac{(A - a_{\text{SC}})^2 u^2(b_{\text{SC}})}{b^4_{\text{SC}}} + \frac{2(A - a_{\text{SC}})}{b^3_{\text{SC}}}\text{cov}(a_{\text{SC}}, b_{\text{SC}})} \quad (37)$$

The uncertainties $u^2(A)$, $u^2(a_{\text{SC}})$ and $u^2(b_{\text{SC}})$ are obtained from the statistical parameters of the straight line fits: $s^2(A)$ from the Youden's plot; $s^2(a_{\text{SC}})$ and $s^2(b_{\text{SC}})$ from the standard calibration plot. Also cov(a_{SC}, b_{SC}) is computed from the standard calibration.

Once the uncertainty $u(\delta_{\text{c}})$ is evaluated, the method constant bias is tested for significance in a way very similar to recovery:

- if $|\delta_{\text{c}}|/u(\delta_{\text{c}}) \leq k$, the constant bias is not significantly different from 0; and
- if $|\delta_{\text{c}}|/u(\delta_{\text{c}}) > k$, the constant bias is significantly different from 0

and the analytical result should be corrected by δ_{c}. As Maroto et al. [9] pointed out, even if the analytical procedure is free from constant bias, its uncertainty must be included in the overall uncertainty budget for future determinations. The same applies for the absence of proportional bias. Thus, as it was already mentioned, the correction indicated by Eq. (1) becomes $Z = Z_{\text{found}}$ in case of absence of both constant and proportional method bias, but their uncertainties remain in the final uncertainty expression as shown in Eq. (4).

4. Robustness study

Robustness [33,34], considered in the sense of internal validation, deals with the effect of experimental variables

called factors, inherent to the analytical procedure (temperature, mobile phase composition, detection wavelength, pH, ...) on the analytical result. Robustness study examines the alteration of these factors as expected in a transfer between laboratories and therefore, is of utmost importance in the uncertainty budget. Within the field of experimental design, to study the main effects of factors, screening designs are applied. Screening designs are two-level saturated fractional factorial designs centred on the nominal analytical conditions [35]. Plackett and Burmann [36] developed such designs for studying f factors in $N = f + 1$ experiments when N is any multiple of four less than 100 (except for 92) [37]. Plackett–Burmann designs are very useful tools for the robustness study of analytical procedures. However, these designs cannot deal with factor interactions and hence, they are suitable when the interactions are negligible or for the sake of consider the key set of dominant factors [38].

The strategy to carry out the robustness study is based on a landmark procedure suggested by Youden [39,40]:

(i) Identify the influential factors.
(ii) For each factor, define the nominal (x^0) and extreme levels (x^{max}, x^{min}) expected in routine work. These values are encoded as follows: $x^0 = 0$, $x^{max} = +1$ and $x^{min} = -1$.
(iii) Arrange the experimental design by using a two-level 2^{7-4} fractional Plackett–Burmann matrix.
(iv) Perform the experiments in random order on a control sample with analyte concentration halfway the concentration range of the method scope.

According to the definition of robustness [14], the interval under investigation is very short (−1, +1; e.g. pH from 3.8 to 4.2). Under these conditions, it must be stressed that no quadratic effect is generally observed and thus, a linear model can be used. It is one of fundamental differences between the robustness study and that one of optimization, in which the investigated interval is wider.

Youden selected a 2^{7-4} Plackett–Burmann design because enables the study up to seven factors in eight experiments. The corresponding matrix design is illustrated in Table 1. The eight runs are split into two groups of four runs on the basis of the levels +1 or −1. Every factor x_k is estimated as the difference of the mean result obtained at the level +1 from the obtained at the level −1.

$$D(x_k) = \tfrac{1}{4}\left(\sum_i Z_i\right)_{(x_k=+1)} - \left(\sum_i Z_i\right)_{(x_k=-1)} \quad (38)$$

For example, for the factor x_5:

$$D(x_5) = \left(\frac{Z_1 + Z_3 + Z_6 + Z_8}{4}\right) - \left(\frac{Z_2 + Z_4 + Z_5 + Z_7}{4}\right) \quad (39)$$

Youden chosen this design because he thought seven factors are enough for any robustness study. But in case of dealing with a less number of influential factors ($f < 7$), what is the way to perform the study? Youden explained the procedure in a funny manner:

"Suppose only six factors are explored. In that event, associate with factor $x_7 = -1$ [factor g in the original] some meaningless operation such as solemnly picking up the beaker, looking at it intently, and setting it down again. Omit this meaningless operation for determinations that involve factor $x_7 = +1$ [factor G in the original]"

Once the effects $D(x_k)$ have been estimated, to determine whether variations in a factor have a significant effect on the analytical result, a significance test is used [41]:

$$t(x_k) = \frac{\sqrt{n}|D(x_k)|}{\sqrt{2}s} \quad (40)$$

Here, n is the number of experiments carried out at each level for each factor ($n = 4$ in the 2^{7-4} Plackett–Burmann design) and s is the estimation of the method precision. It can be used the result obtained from the precision study (R.S.D.$_I$). In that case: $s = \text{R.S.D.}_I \bar{Z}$, by taking $\bar{Z} = \sum_i Z_i/8$. The t value is compared with the two-tailed critical one, t_{crit}, for $N - 1$ degrees of freedom at 95% confidence, where N is the number of determinations used in the evaluation of R.S.D.$_I$.

If $t(x_k) \leq t_{crit}$, the factor effect is not significantly different from 0 and therefore, slight changes in the factor do not have a significant effect on method performance (the method is robust against this factor). If $t(x_k) > t_{crit}$, the effect is significantly different from 0 and the method is not

Table 1
Arrangement of factor levels for a 2^{7-4} Plackett–Burmann design

Runs, N	Factors x_k (k = 1 to f)							Response, Z_i (i = 1 to N)
	X_1	X_2	X_3	X_4	X_5	X_6	X_7	
1	+1	+1	+1	+1	+1	+1	+1	Z_1
2	+1	+1	−1	+1	−1	−1	−1	Z_2
3	+1	−1	+1	−1	+1	−1	−1	Z_3
4	+1	−1	−1	−1	−1	+1	+1	Z_4
5	−1	+1	+1	−1	−1	+1	−1	Z_5
6	−1	+1	−1	−1	+1	−1	+1	Z_6
7	−1	−1	+1	+1	−1	−1	+1	Z_7
8	−1	−1	−1	+1	+1	+1	−1	Z_8

robust against this factor. In both cases, there is an uncertainty in the final result Z associated with the factor x_k, $u(Z(x_k))$.

For the factors whose variations do not affect the method performance, the uncertainty is evaluated according to [10]:

$$u(Z(x_k)) = \frac{\sqrt{2}t_{\text{crit}}s}{\sqrt{n}1.96}\frac{\delta_{\text{real}}(x_k)}{\delta_{\text{test}}(x_k)} \quad (41)$$

In this expression, n, t_{crit} and s play the same role as above. δ_{real} is the change in the factor level which would be expected when the method is operating under control routine analysis. δ_{test} is the change in the factor level specified in the robustness study, that is, $\delta_{\text{test}} = x_k^{\max} - x_k^{\min}$. For example, consider the extreme values for the detection wavelength in a HPLC procedure to be $x^{\max}$ = 219 nm and $x^{\min}$ = 209. The specifications of the detector indicate that the expected tolerance for the wavelength is ±2 nm. If the effect of wavelength does not affect the method performance, Eq. (41) will be used to estimate the contribution uncertainty. The values of δ_{real} and δ_{test} will be: δ_{real} = (+2 nm) − (−2 nm) = 4 nm, and δ_{test} = (219 nm) − (209 nm) = 10 nm.

For those factors affecting significantly the method performance, the uncertainty is computed as:

$$u(Z(x_k)) = u(x_k)c_k \quad (42)$$

Here, $u(x_k)$ is the type B uncertainty estimated from the specifications of the corresponding item assuming a given probability density function, and c_k is an estimation of the sensitivity coefficient, taken as the quotient of the observed effect on the result ($D(x_k)$) and the change in the factor level ($x_k^{\max} - x_k^{\min}$):

$$c_k = \frac{D(x_k)}{x_k^{\max} - x_k^{\min}} = \frac{D(x_k)}{\delta_{\text{test}}} \quad (43)$$

When all the uncertainty contributions of the factors have been estimated, the relative uncertainty of the robustness study is calculated as:

$$\text{R.S.D.}_{\text{rob}} = \sqrt{\frac{\sum_{k=1}^{f} u^2(Z(x_k))}{\overline{Z}^2}} \quad (44)$$

The total estimation for the uncertainty on the analyte concentration Z for a future real sample is then computed by applying Eq. (3).

5. Conclusion

The estimation of measurement uncertainty for chemical assays according to the LGC/VAM protocol fromvalidation data is a suitable approach and can be considered as an alternative to other well established schemes such as “bottom-up” and “top-down”.

References

[1] Guide to the Expression of Uncertainty in Measurement, ISO, Geneva, Switzerland, 1993, corrected and reprinted in 1995.
[2] Analytical Methods Committee, Analyst, 120 (1995) 2303–2308.
[3] EURACHEM/CITAC Guide, Quantifying Uncertainty in Analytical measurement, second ed., 2000, http://www.vtt.fi/kct/eurachem/quam2000-pl.pdf.
[4] M. Valcárcel, A. Ríos, Anal. Chim. Acta 400 (1999) 425–432.
[5] V.J. Barwick, L.R. Ellison, Analyst 124 (1999) 981–990.
[6] A. Maroto, R. Boqué, J. Riu, F.X. Rius, TrAC 18 (1999) 577–584.
[7] L.R. Ellison, V.J. Barwick, Analyst 123 (1998) 1387–1392.
[8] L.R. Ellison, A. Williams, Accred. Qual. Assur. 3 (1998) 6–10.
[9] A. Maroto, J. Riu, R. Boqué, F.X. Rius, Anal. Chim. Acta 391 (1999) 173–185.
[10] V.J. Barwick, L.R. Ellison, VAM Project 3.2.1, Development and Harmonisation of Measurement Uncertainty Principles. Part D: Protocol for Uncertainty Evaluation from Validation Data, January 2000, Report No.: LGC/VAM/1998/088.
[11] D.L. Massart, B.G.M. Vandeginste, L.M.C. Buydens, S. de Jong, P.J. Lewi, J. Smeyers-Verbeke, Handbook of Chemometrics and Qualimetrics, Part A, Elsevier, Amsterdam, 1997.
[12] Ph. Hubert, J.J. Nguyen-Huu, B. Boulanger, E. Chapuzet, P. Chiap, N. Cohen, P.A. Compagnon, W. Dewé, M. Feinberg, M. Lallier, M. Laurentie, N. Mercier, G. Muzard, C. Nivet, L. Valat, STP Pharma Pratiques 13 (2003) 101–138.
[13] A.M. García-Campaña, J.M. Bosque-Sendra, L.C. Rodríguez, E.A. López, Biomed. Chromatogr. 14 (2000) 27–29.
[14] http://www.fda.gov/cder/guidance/ichq2a.pdf.
[15] E. Hund, D.L. Massart, J. Smeyers-Verbeke, Anal. Chim. Acta 480 (2003) 39–52.
[16] http://www.fda.gov/cder/guidance/1320fnl.pdf.
[17] http://www.fda.gov/cder/guidance/4252fnl.htm.
[18] USP, United States Pharmacopeia 24th ed., United States Pharmacopeial Convention, Rockville, MD, 2000.
[19] G.W. Snedecor, W.G. Cochran, Statistical Methods, eighth ed., Iowa State University Press, 1989.
[20] W.J. Youden, Anal. Chem. 19 (1947) 946–950.
[21] L. Huber, LC-GC Int. 11 (1988) 96–105.
[22] A.G. González, M.A. Herrador, Talanta 48 (1999) 729–736.
[23] M.J. Cardone, Anal. Chem. 58 (1986) 438–445.
[24] W.J. Youden, Biometrics 3 (1947) 61.
[25] W.J. Youden, Mater. Res. Stand. 1 (1961) 268–271.
[26] K.S. Booksh, B.R. Kowalski, Anal. Chem. 66 (1994) 782A–791A.
[27] A. Maroto, R. Boqué, J. Riu, F.X. Rius, Anal. Chim. Acta 446 (2001) 133–145.
[28] Reference [3], Appendix E (Statistical Procedures), p. 106.
[29] Guide for the determination of the mass within the scope of reference, measurement procedures in medical reference measurement laboratories, PTB Mitteilungen 109 (1999) 379–383. http://www.dkd.info/de/med_messgr/guidemass.pdf.
[30] T.J. Farrant, Statistics for the Analytical Scientist. A Bench Guide, Royal Society of Chemistry, Cambridge, UK, 1997.
[31] A. Maroto, R. Boqué, J. Riu, F.X. Rius, Analyst 128 (2003) 373–378.
[32] R.C. Castells, M.A. Castillo, Anal. Chim. Acta 423 (2000) 179–185.
[33] M.B. Sanz, L.A. Sarabia, A. Herrero, M.C. Ortiz, Talanta 56 (2002) 1039–1048.

[34] L.C. Rodríguez, R. Blanc, A.M. García, J.M. Bosque, Chemom. Intell. Lab. Sys. 41 (1998) 57–68.
[35] J.M. Bosque, M. Nechar, L. Cuadros, Fresenius J. Anal. Chem. 365 (1999) 480–488.
[36] R.L. Plackett, J.P. Burmann, Biometrica 33 (1946) 305–325.
[37] D.C. Montgomery, Design and Analysis of Experiments, third ed., Wiley, NY, 1991.
[38] A.G. Gonzalez, Anal. Chim. Acta 360 (1998) 227–241.
[39] W.Y. Youden, Statistical Techniques for Collaborative Tests, Association of Official Analytical Chemists (AOAC), Washington DC, 1967.
[40] I. García, M.C. Ortiz, L. Sarabia, C. Vilches, E. Gredilla, J. Chromatogr. A 992 (2003) 11–27.
[41] Y. Vander Heyden, K. Luypaert, C. Hartmann, D.L. Massart, J. Hoogmartens, J. De Beer, Anal. Chim. Acta 312 (1995) 245–262.