trends

Operational definitions of uncertainty

Edelgard Hund, D. Luc Massart, Johanna Smeyers-Verbeke*
ChemoAC, Farmaceutisch Instituut, Vrije Universiteit Brussel, Laarbeeklaan 103, B-1090 Brussels, Belgium

Very different approaches for the estimation of the uncertainty related to measurement results are found in the literature and in published guidelines. This article analyses and compares them. It is clear that 'uncertainty' is not, and should not be, the same in all situations. As a consequence, operational definitions of uncertainty are proposed that take into account the differences in the ways in which truth, uncertainty and error are conceived. **©2001 Elsevier Science B.V. All rights reserved.**

Keywords: Measurement uncertainty; Measurement error; Traceability; Bias

1. Uncertainties about uncertainty

Analytical chemists accept that a measurement cannot be interpreted properly without knowledge of its uncertainty. EURACHEM [1] defines uncertainty as "a parameter associated with the result of a measurement, that characterizes the dispersion of the values that could reasonably be attributed to the measurand". This definition follows the original definition of the Guide to the Expression of Uncertainty in Measurement (also known as the GUM) [2,3].

The uncertainty of the result of a set of measurements, e.g., a concentration, x, can be expressed in the form of a standard deviation, the so-called standard uncertainty, abbreviated as $u(x)$. The expanded uncertainty $U(x)$ defines an interval around the result of a measurement, $x \pm U(x)$, with $U(x) = ku(x)$. The constant k is called the coverage factor, and for $k = 2$, the expanded uncertainty is roughly equivalent to half the length of a 95% confidence interval. Thus, the probability that the mean value is included in the expanded uncertainty is about 95%.

The problem for analytical chemists is how to determine uncertainty in their specific situations and, unfortunately, there is much confusion and uncertainty (uncertainties about uncertainty!) about how to proceed. Very different approaches are found in the literature and in published guidelines. There are different reasons for this. The first lies in the way errors are treated. It is clear that uncertainty is related to measurement errors, but literature reports and guidelines differ in the sources of error that they take into account or to which they attribute uncertainty. In particular, the treatment of systematic error is a point of discordance. Errors refer to a divergence from the truth, and a second reason for the confusion in the literature lies in what is considered to be the truth: is one looking for the absolute truth or for a relative truth? In the former case, the reference is the true value, μ, while in the latter case the reference is a consensus value, e.g., one agreed upon for the purpose of comparative measurements. A third and very important reason is related to the way error is determined in practice. Analytical chemists are used to determining error by what they call 'method validation', and would like to use this for determining uncertainty. Metrologists have a different approach, which they apply to physical methods and would like to see applied also in analytical chemistry.

In this article, we try to analyse and compare the different approaches. We conclude that uncertainty is not, and should not be, the same to all practitioners and that definitions are therefore required that take into account the differences in the ways truth, uncertainty and error are conceived.

We only consider here the uncertainty of the measurement process. In several situations, the uncertainty related to, for example, sampling, homogeneity or stability of the samples also plays an important role, and should be taken into consideration.

*Corresponding author. Tel.: +32 (2) 477 4737; Fax: +32 (2) 477 4735.
E-mail: asmeyers@fabi.vub.ac.be

0165-9936/01/$ – see front matter
PII: S0165-9936(01)00089-9

© 2001 Elsevier Science B.V. All rights reserved.

2. Measurement errors

When an analytical chemist measures the concentration of a substance, the result should be considered an estimate of the true value. The error is then the difference between a stated result and the true value. There is a single error value for each result and part of it, the systematic error, can be corrected for. In contrast, the uncertainty derived from the errors is a range and no part of the uncertainty can be corrected for [4].

The measurement result obtained deviates from the true value because of the existence of a number of systematic and random errors. The following can be distinguished [5,6]:

- The method bias, a systematic error owing to the method used.
- The laboratory bias, which according to one's point of view is a systematic error (for an individual laboratory) or a random error (when the laboratory is viewed as a part of a population of laboratories, as is the case in an inter-laboratory study). In the latter case it is a component of the reproducibility of the method used.
- The run error, which is a random error owing to, among other factors, time effects, and is included in the intermediate precision estimate.
- The repeatability error, which is a random error occurring between replicate determinations performed within one run.

This list is sometimes called the 'ladder of errors' [6], because there is a hierarchy involved: the method as such; the method as it is applied by a certain laboratory; the method as it is applied by a certain laboratory on a certain day; and, finally, the error within that day for an individual determination.

A measurement result can therefore be decomposed as follows: Measurement result = true value + method bias + laboratory bias + run error + repeatability error.

Each of these steps on the ladder adds its own uncertainty. This is usually well understood for the random errors but less for the method bias. One guideline [7], for example, states that if systematic error occurs this should be corrected for, and from then on assumes that this has been done and the error does not need to be taken into account in establishing the uncertainty. However, the estimation of the systematic error is itself the result of measurements and is therefore subject to a degree of uncertainty. This uncertainty should then be taken into account in an uncertainty statement. When it is concluded that there is no bias, this in fact means in most practical cases that the bias is smaller than a certain limit, below which it may exist but cannot be detected. This, too, is a source of uncertainty and should be included in an exhaustive uncertainty statement.

For the uncertainties we can write:
Uncertainty = uncertainty associated with the method bias + uncertainty associated with the laboratory bias + uncertainty owing to the run effect + repeatability uncertainty.

Translated into variances this becomes for a single measurement:

$$u_x = \sqrt{u^2_{\text{meth.bias}} + \sigma^2_{\text{lab}} + \sigma^2_{\text{run}} + \sigma^2_r} \qquad (1)$$

Making use of the standard error of the mean [5,11], the standard uncertainty for the mean of $q \times p \times n$ measurements performed as n replicates (repeatability conditions) in each of p runs in each of q laboratories can be then calculated as:

$$u_{\bar{x}} = \sqrt{u^2_{\text{meth.bias}} + \sigma^2_{\text{lab}}/q + \sigma^2_{\text{run}}/qp + \sigma^2_r/qpn} \qquad (2)$$

2.1. Traceability and method bias

One of the most difficult problems in chemical measurement is to determine method bias. It is useful to consider two types of method bias, namely absolute or constant method bias, and relative or proportional method bias. There is a constant bias if a method leads to measurement results that deviate by a constant value from the corresponding true value. The observation of 110, 210 and 310 instead of 100, 200, 300 would be an example of a constant method bias. Often, problems with an inappropriate blank correction result in a constant bias. If the difference between the measurement result obtained with a certain method and the corresponding true value is proportional to the concentration of the analyte, a proportional method bias occurs. With a proportional bias, 110, 220 and 330 might, for example, be observed in the example mentioned above. A proportional bias is often caused by a matrix interference.

Method bias refers explicitly or implicitly to a reference or a standard that is considered to represent the truth. In the metrological vocabulary it is said that the result obtained can then be *traced* to a stated reference. 'Traceability' is defined as "the property of the result of a measurement or the value of a standard, whereby it can be related to stated references, usually national or international standards, through an unbroken chain of comparisons all having stated uncertainties" [2]. The second edition of the EURACHEM guideline [1] mentions four types of reference, which are important for chemical measurements. One of them requires the use of so-called primary methods. The CCQM [8] has identified isotope dilution with mass spectrometry (IDMS), coulometry, gravimetry and titrimetry as methods that have the potential to be primary methods. Such methods are in principle traceable to SI units, but it is not evident that this is also the case when these methods are applied under the conditions of routine laboratories. For most other methods the main possibilities are:

- using the method to make measurements on an appropriate certified reference material (CRM). The method is then traceable to that CRM;
- making measurements using defined procedures. In this case the method is traceable to the reference method;
- using the analytical procedure to make measurements on a known quantity of pure analyte. In practice, this usually means that recovery studies or standard addition methods are carried out.

Thus, there are always at least two components in the uncertainty associated with the method bias. The first is the uncertainty of the reference (which we will call here $u_{(\text{traceability})}$), the second is the uncertainty associated with the estimated bias:

$$u_{\text{meth.bias}} = \sqrt{u^2_{\text{traceability}} + u^2_{\text{est.bias}}} \tag{3}$$

The ultimate reference in the chemical analytical world is the value of the mole. A practical example would be the use of a pure standard substance as the reference. The uncertainty in that reference is due to the uncertainty of the purity of that standard substance. Very often, the use of such a reference is not practical and one uses as reference the concentration of a certain substance as determined by a given method in a given material. There is then an uncertainty owing to that determination.

3. Error budgets for determining the uncertainty

In the error budget method which, in analytical chemistry, seems to be used more in Europe than elsewhere, the variance is estimated of each variable that contributes to the total variance. This is also referred to as the 'bottom-up' approach. A simple example for an acid–base titration is given in the new version of the EURACHEM guideline [1]. The example treats the standardisation of a NaOH solution against the titrimetric standard potassium hydrogen phthalate (KHP). The concentration of NaOH is obtained from the following formula, with m_{KHP} the mass of KHP, P_{KHP} the purity of KHP, F_{KHP} the formula weight of KHP and V_{T} the volume of NaOH used for the titration of KHP:

$$c_{\text{NaOH}} = \frac{1000 \cdot m_{\text{KHP}} \cdot P_{\text{KHP}}}{F_{\text{KHP}} \cdot V_{\text{T}}} \tag{4}$$

Eq. 4 describes a multiplicative relationship. While the variance of a sum or a difference is the sum of the variances of its components, the relative variance of a product or division is the sum of the relative variances of its components. Eq. 4 is of the type:

$$x = k\frac{a \cdot b}{c \cdot d} \tag{5}$$

where k is a numerical constant. The relative variance

$$\left(\frac{\sigma(x)}{x}\right)^2$$

is given by:

$$\left(\frac{\sigma(x)}{x}\right)^2 = \left[\left(\frac{\sigma(a)}{a}\right)^2 + \left(\frac{\sigma(b)}{b}\right)^2 + \left(\frac{\sigma(c)}{c}\right)^2 + \left(\frac{\sigma(d)}{d}\right)^2\right] \tag{6}$$

or

$$(\sigma(x))^2 = \left(k\frac{a \cdot b}{c \cdot d}\right)^2 \cdot \left[\left(\frac{\sigma(a)}{a}\right)^2 + \left(\frac{\sigma(b)}{b}\right)^2 + \left(\frac{\sigma(c)}{c}\right)^2 + \left(\frac{\sigma(d)}{d}\right)^2\right] \tag{7}$$

or

$$(\sigma(x))^2 = \left(k\frac{b}{c\cdot d}\sigma(a)\right)^2 + \left(k\frac{a}{c\cdot d}\sigma(b)\right)^2 + \left(k\frac{a\cdot b}{c^2\cdot d}\sigma(c)\right)^2 + \left(k\frac{a\cdot b}{c\cdot d^2}\sigma(d)\right)^2 \quad (8)$$

Notice that the coefficient for each variance $\sigma(y_i)$ is the partial derivative of x with respect to y_i, $\frac{\partial x}{\partial y_i}$,

so that:

$$(\sigma(c_{NaOH}))^2 = (1000)^2\cdot\left[\left(\frac{P_{KHP}}{F_{KHP}\cdot V_T}\sigma(m_{KHP})\right)^2 + \left(\frac{m_{KHP}}{F_{KHP}\cdot V_T}\sigma(b)\right)^2 + \left(\frac{m_{KHP}\cdot P_{KHP}}{F_{KHP}^2\cdot V_T}\sigma(F_{KHP})\right)^2 + \left(\frac{m_{KHP}\cdot P_{KHP}}{F_{KHP}\cdot V_T^2}\sigma(V_T)\right)^2\right] \quad (9)$$

It is more fashionable to write that the error budget approach models the uncertainty of a measurement result as a first-order Taylor series:

$$u(x) = \sqrt{\sum_{i=1}^{n}\left(\frac{\partial x}{\partial y_i}\cdot u(y_i)\right)^2 + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\frac{\partial x}{\partial y_i}\frac{\partial x}{\partial y_j}u(y_i, y_j)} \quad (10)$$

with y_i the value considered, and $u(y_i)$ the standard uncertainty related to this value. The second summand under the square root sign refers to the uncertainties related to the covariances. The first-order model is based on the assumption that the uncertainty of the uncertainty itself is negligible. In the example used by EURACHEM, the covariances are not taken into account. This yields the following error budget:

This equation is equivalent to Eq. 9. In the example, the uncertainty of the purity of the standard is obtained from the specifications of the provider; the other uncertainty sources were evaluated separately. For the uncertainty of the weight, both the full and the emptied container have to be considered. The uncertainty of each of the weights is derived from a balance repeatability and a contribution from the uncertainty of the balance's linearity. The other two sources of uncertainty require a comprehensive evaluation as well. The corresponding values are given in Table 1. The concentration of the NaOH solution is determined as 0.1021 mol l^{-1}. From Eq. 9 or Eq. 11 and using the data of Table 1, the corresponding standard uncertainty is computed as 0.0001 mol l^{-1}.

The error budget method as described in the GUM was developed by metrologists and physicists. It has indeed proved its value for physical measurements and is probably suitable for primary analytical methods. However, the example of Eq. 10 shows that, even for a primary method (titrimetry), the error budget approach becomes quite complex. To derive a complete uncertainty estimate requires that all relevant parameters be considered. For another primary method (IDMS), Dobney et al. [9] state that this leads to unwieldy expressions so that strict application of the uncertainty propagation law seems a daunting project. For methods with more uncertainty components, it is hardly feasible to construct an error budget. The error budget approach assumes that no important error is overlooked, and in many practical analytical situations this is not evident because, in practice, it is impossible to predict, for example, the effect of unanticipated matrix interferences. To estimate such sources of error requires systematic method validation.

For more complex analytical methods, it seems preferable to use the information obtained from method validation. Method validation leads to estimates of random and of systematic errors that encompass many different sources of uncertainty. From a practical point of view, it requires much less work, because the analytical chemist has to validate methods anyway. From method validation to

$$u(c_{NaOH}) = 1000\cdot\sqrt{\left[\left(\frac{P_{KHP}}{F_{KHP}\cdot V_T}\cdot u(m_{KHP})\right)^2 + \left(\frac{m_{KHP}}{F_{KHP}\cdot V_T}\cdot u(P_{KPH})\right)^2 + \left(\frac{m_{KHP}\cdot P_{KHP}}{(-F_{KHP})^2\cdot V_T}\cdot u(F_{KHP})\right)^2 + \left(\frac{m_{KHP}\cdot P_{KHP}}{-F_{KHP}\cdot(V_T)^2}\cdot u(V_T)\right)^2\right]} \quad (11)$$

Table 1
Values and uncertainties in the standardization of an NaOH solution

	Description	Value x_i	Standard uncertainty $u(y_i)$	Relative standard uncertainty $u(y_i)/y_i$
m_{KHP}	Weight of KHP	0.3888 g	0.00013 g	0.00033
P_{KHP}	Purity of KHP	1.0	0.00029	0.00029
V_T	Volume of NaOH for KHP titration	18.64 ml	0.013 ml	0.00070
F_{KHP}	Formula weight of KHP	204.2212 g mol^{-1}	0.0038 g mol^{-1}	0.000019
c_{NaOH}	Concentration of NaOH solution	0.10214 mol l^{-1}	0.0001 mol l^{-1}	0.00097

uncertainty estimation is not an enormous step, and should overcome the mixture of scepticism and awe with which the introduction of uncertainty measurement and its metrological terminology has been greeted. Wood et al. [10] give a comparison of these two approaches, which was initiated by the UK Ministry of Agriculture, Fisheries and Food. For a variety of examples, such as the determination of total nitrogen in meat products using the Kjeldahl method, or the GFAA analysis of lead in wine, it was shown that both approaches lead to comparable results.

The discussion above does not mean that the error budget approach has no advantages at all. The fact that one has to consider all sources of uncertainty allows the largest contributions to be identified. These can then be considered particularly if the uncertainty has to be reduced. Method validation, on the other hand, focuses on the errors of the analytical procedure as such, but has a tendency to underestimate the importance of other sources of errors such as those related to sampling or some pre-treatment steps.

4. Should all sources of uncertainty always be included?

In many practical cases the analyst is not trying to determine the absolute truth. This is the case, for example, when standard methods are being used. The method has then been validated by a large group of laboratories known to be proficient in that method. In such a situation, the analyst can assume that the method bias is reduced to an acceptably small value. If a laboratory considers itself a member of the population of proficient laboratories, it can also regard the uncertainty related to the bias as negligible. From a metrological point of view, this means that one regards the problem in a relative way. Even though such an approach might seem a little too optimistic, it is acceptable in practice. The uncertainty to be considered is then the uncertainty associated with measurements performed with the standard method in the analyst's laboratory. This can be evaluated at an acceptable expense.

5. Operational definitions of uncertainty

Many years ago, analytical chemists described random errors using the general term 'precision'. They have now learned that there are different kinds of precision, such as reproducibility, repeatability, and intermediate forms, which should be used in certain well-defined situations. While the terms used indicate immediately what sources of variance are included in a precision statement, this is unfortunately not the case for uncertainty statements. The term 'uncertainty' can encompass different sources of uncertainty according to the situation considered. It is our opinion that different names should be given, or different qualifications added, to the term 'uncertainty' depending on the conditions under which an analyst is operating. In what follows, we will consider the following operational definitions of uncertainty (see also Table 2).

Within-laboratory uncertainty only considers the intermediate precision and therefore accounts for the repeatability and the run effect.

Reproducibility uncertainty additionally considers the reproducibility variance and therefore accounts for the within-laboratory uncertainty and the laboratory bias.

Bias-included uncertainty as well as **absolute uncertainty** additionally consider the uncertainty owing to the method bias. They also take into account the uncertainty owing to the estimated method bias and to the traceability. The bias-

Table 2
Some operational definitions of uncertainty

Type of uncertainty	Precision estimates considered	Uncertainty contributions accounted for	Uncertainty for the mean of n measurements performed under repeatability conditions $u_{\bar{x}}$
Within-laboratory uncertainty	Intermediate precision	Repeatability, run effect (e.g., analyst and time)	$\sqrt{s^2_{\text{analyst}} + s^2_{\text{time}} + s^2_r/n}$
Reproducibility uncertainty	Reproducibility	Repeatability, run effect (e.g., analyst and time), lab effect	$\sqrt{s^2_{\text{BL}} + s^2_r/n}$
Bias-included uncertainty	Intermediate precision	Repeatability, run effect (e.g., analyst and time), bias (lab and method)	$\sqrt{u^2_{\text{est.bias}} + s^2_{\text{run}} + s^2_r/n}$ [a]
			e.g., $\sqrt{s^2_d + s^2_{\text{run}} + s^2_r/n}$ [b]
			$\sqrt{s^2_{\text{obs}}/m_1 + s^2_{\text{run}} + s^2_r/n}$ [c]
			$\sqrt{s^2_{\text{obs}}/m_1 + s^2_{\text{native}}/m_2 + s^2_{\text{run}} + s^2_r/n}$ [d]
			$\sqrt{(\text{est.bias}/k)^2 + u^2_{\text{est.bias}} + s^2_{\text{run}} + s^2_r/n}$ [e]
Absolute uncertainty	Intermediate precision	Repeatability, run effect (e.g., analyst and time), bias (lab and method), traceability to CRM	$\sqrt{u^2_{\text{CRM}} + s^2_{\bar{x}} + s^2_{\text{run}} + s^2_r/n}$

[a] General expression if the bias is not significant or a significant bias is corrected for.
[b] Bias estimated from a comparison with the reference method; the uncertainty in the reference method is considered negligible. If a *t*-test is used in the evaluation of the bias, the pooled variance of both methods is used in s^2_d.
[c] Bias estimated from recovery experiments with a blank.
[d] Bias estimated from recovery experiments; the uncertainty of the spiked concentration is considered negligible. If a *t*-test is used in the evaluation of the bias, s^2_{obs} and s^2_{native} are pooled.
[e] General expression if a significant bias is not corrected for.
Notations used:
s^2_r, repeatability variance.
n, number of experiments performed under repeatability conditions.
s^2_{time}, variance component between days.
s^2_{analyst}, variance component between analysts.
s^2_{BL}, variance component between laboratories.
s^2_{run}, variance component between runs (e.g., different days and/or analysts).
s^2_d, variance of the difference of the mean results of the two methods.
$s^2_{\text{est.bias}}$, uncertainty related to an estimated bias.
s^2_{obs}, variance observed for the analyses of a spiked sample.
m_1, number of determinations performed on a spiked sample.
s^2_{native}, variance observed for the analyses of a native sample.
m_2, number of determinations performed on the native sample.
$s^2_{\bar{x}}$, variance of the mean concentration observed for the CRM (e.g., repeatability or intermediate precision conditions).
u^2_{CRM}, uncertainty related to a certified reference material, specified on the certificate.
k, coverage factor.

included uncertainty and the absolute uncertainty differ in the level of traceability of the measurement results.

We do not propose that these terms, as such, should be generally used, but we do propose that an international body of analytical chemists should consider the situation and define generally recognised operational definitions of uncertainty. We shall now consider the references to which the measurement is traced in each of those cases, what sources of error are considered, and how uncertainty could be determined.

5.1. Within-laboratory uncertainty

A guideline prepared by the NMKL [7] states that it wants to de-dramatise the subject of measurement uncertainty (which describes very well how

Table 3
Results of the inter-laboratory study

Laboratory	x_{i1}	x_{i2}	$\overline{x_i}$
1	7.9	12.0	9.95
2	9.1	10.4	9.75
3	9.7	10.2	9.95
4	12.0	9.9	10.95
5	10.6	12.1	11.35
6	10.4	8.0	9.2
7	9.4	7.7	8.55
8	10.1	10.5	10.30
9	12.0	13.0	12.50
10	8.1	7.3	7.70
11	10.3	11.6	10.95
12	10.9	12.5	11.70
13	15.0	14.0	14.50
14	10.7	7.1	8.90

many analytical chemists react when confronted with the subject). It tries to do so by using a simple procedure based on the determination of a confidence interval with a precision measure. The precision statement is preferably what is called in the NMKL guideline the within-laboratory reproducibility. In ISO terms, this is the intermediate precision. If this precision estimate is not available, then it is suggested that repeatability can be used as the precision statement, but the guideline warns that, in that way, the uncertainty is underestimated. The reference in this guideline is the mean of the laboratory for that determination. Only the lower rungs of the ladder of errors are included, namely those owing to the repeatability and the run errors. Laboratory and method bias are not considered, and there is no traceability to SI units or at least to stated references. In our opinion, this is perfectly acceptable as long as the limitations in the level of reference and sources of uncertainty are recognised.

A minor remark which could be made about this guideline is that it would have been preferable to separate the intermediate precision into its components, i.e., the repeatability and the between-run (such as between time+between analysts) components, by using a designed experiment in the same way as for the repeatability and between-laboratory components in Section 5.2. If this is done, it is possible to compute the uncertainty when the analyst afterwards performs n replicate measurements under repeatability conditions, and uses the mean of the replicates as the stated result. The within-laboratory uncertainty on the mean becomes:

$$u_{\overline{x}} = \sqrt{s_{run}^2 + s_r^2/n} = \sqrt{s_{analyst}^2 + s_{time}^2 + s_r^2/n} \quad (12)$$

For the mean of n replicate measurements performed by a single analyst on each of p days the within-laboratory uncertainty would then be:

$$u_{\overline{x}} = \sqrt{s_{analyst}^2 + s_{time}^2/p + s_r^2/pn} \quad (13)$$

The NMKL guideline states that if systematic error occurs, this should be corrected for, and from then on, it assumes that this has been done. The bias estimate is not considered to constitute a source of uncertainty. However, as explained earlier, even if the bias is determined to be zero (since it is lower than a certain limit) an uncertainty is related to this factor. Therefore, the NMKL guideline is somewhat misleading because a less expert user might consider that the reference level is higher than it is for this guideline. When no uncertainty related to the bias is included in the uncertainty estimate, this should be clear, e.g., from the term used in identifying the uncertainty estimate.

5.2. Reproducibility uncertainty

In reference [11] the authors describe an intercomparison experiment for polyunsaturated fatty acids in oil. Each participating laboratory carried out duplicate analyses on the sample. The results of these determinations after removal of outliers are given in Table 3.

The analysis of variance yields two mean squares. By dividing them by the appropriate number of degrees of freedom one obtains s_r^2 and $(s_r^2 + 2s_{BL}^2)$ as estimates of σ_r^2 and $(\sigma_r^2 + 2\sigma_{BL}^2)$, from which σ_r and σ_{BL} can be computed. They are respectively the repeatability variance and the between-laboratory variance. The reproducibility

Table 4
Results of the analysis of variance (ANOVA) for the inter-laboratory study

Variance estimate	Variance	Corresponding standard deviation
Within laboratory σ_{WL}^2	1.9	1.4
Between laboratory σ_{BL}^2	2.1	1.4
Reproducibility σ_R^2	4.0	2.0

variance can then be obtained as the sum of the two variance estimates: $s_R^2 = s_r^2 + s_{BL}^2$. Notice that $s_{BL}^2 = s_{run}^2 + s_{lab}^2$. The experimental set-up does not, however, allow one to separately estimate s_{run}^2 and s_{lab}^2. The variance estimates obtained for the example are summarised in Table 4.

The uncertainty of an individual analysis is then:

$$u_{\bar{x}} = \sqrt{s_r^2 + s_{BL}^2} \quad (14)$$

If n replicate determinations under repeatability conditions have been carried out, then this uncertainty becomes:

$$u_{\bar{x}} = \sqrt{s_r^2/n + s_{BL}^2} \quad (15)$$

The within- and between-laboratory components are included in the uncertainty estimate. It should be noted that the repeatability is the one computed from the inter-laboratory experiment and not the repeatability obtained in an individual laboratory, which means that individual laboratories should use precision estimates of an inter-laboratory study for computing uncertainties only when they are sufficiently proficient in the particular procedure. This implies, of course, that they have access to the results of the inter-laboratory study. The sources of uncertainty included in the uncertainty statement are the repeatability variance, the run effect, and the laboratory bias, but not the method bias. If the result is defined by a standard procedure applied, such as would be the case, for example, for the determination of crude fibre, then the highest level of reference is reached, since in this empirical method there is, by definition, no method bias. If the standard procedure determines the concentration of a well-defined substance, as was the case for the fatty acids, then it is possible in principle to try tracing the result back to a higher reference level. It should therefore be understood that the reference in the method of determining uncertainty described in the example is the mean value of the result of a large group of qualified laboratories, using the standard procedure concerned, for a given matrix and level of concentration.

A frequently asked question is how to perform an inter-laboratory precision study when there are fewer than eight laboratories, the number required by ISO for an inter-laboratory precision experiment. It has been suggested [6] that in that situation one should estimate the reproducibility and the repeatability from the Horwitz function [12]. This function was deduced from a consideration of more than 7500 method performance studies. It relates the relative reproducibility standard deviation to the concentration. The corresponding repeatability standard deviation is considered to be one-half to two-thirds of the reproducibility, and s_{BL}^2 is then about 0.5–0.75 of s_R^2. Another possibility, when an inter-laboratory experiment cannot be set up, consists in the consideration of data obtained from robustness tests. Since robustness tests simulate the changes that can be expected when transferring an analytical method between laboratories (or instruments, or operators), the variances observed for adequately chosen modifications of the method parameters should give an indication of the reproducibility variance.

5.3. Bias-included uncertainty

In the previous sections, only random errors (intermediate precision and reproducibility) have been considered in the uncertainty statements. Systematic error or bias additionally adds uncertainty to a measurement result.

We will consider here the uncertainty estimation using information from the in-house validation of the analytical method. This implies (i) that usually no estimate of the reproducibility, but only an estimate of the intermediate precision is available and (ii) that the bias estimated is the overall bias, which is a combination of the laboratory bias and the method bias. As mentioned earlier inter-laboratory studies are required to separate the method bias from the laboratory bias.

Eq. 2 can then be written as follows:

$$u_{\bar{x}} = \sqrt{u_{bias}^2 + s_{run}^2 + s_r^2/n} \quad (16)$$

in which u_{bias} combines the uncertainty in the method bias with the uncertainty in the laboratory bias.

5.3.1. Method comparison

If the bias for a new routine method is evaluated by a comparison with a reference method, the uncertainty associated with the bias (see Eq. 3) is:

$$u_{bias} = \sqrt{u_{ref.meth}^2 + u_{est.bias}^2} \quad (17)$$

Indeed the traceability uncertainty, $u_{traceability}^2$, here is the uncertainty in the reference method. If

the latter is a reference method as defined by ISO [13] it has been evaluated in an inter-laboratory study and therefore the reproducibility variance is known. Most often, the reference method will be considered to be not biased and the uncertainty in the reference method to be negligible. Traceability to the reference method is achieved by comparing the results obtained with the routine and reference procedure.

If the bias of the routine method as compared to the reference method is not significant, which in fact means that it is smaller than a certain limit, this does not mean that there is no uncertainty associated with the estimated bias. The uncertainty in the estimated bias, $u_{\text{est.bias}}$, corresponds to the uncertainty associated with the measured difference between the mean results obtained by using both methods. As specified by EURACHEM [1] and IUPAC [14] this corresponds to the standard deviation term, $s_d = s_{(x_A - x_B)}$, that appears in the *t*-test applied to test whether the difference is statistically significant:

$$t = \frac{|x_A - x_B|}{s_d} = \frac{|x_A - x_B|}{s_p\sqrt{\frac{1}{n_A} + \frac{1}{n_B}}} \tag{18}$$

with

$$s_p = \sqrt{\frac{(n_A - 1)s_A^2 + (n_B - 1)s_B^2}{n_A + n_B - 2}}$$

the pooled standard deviation of both methods, n_A and n_B, being the number of measurements performed with the reference and routine method, respectively.

$\bar{x}_A$ and $\bar{x}_B$ are the mean results obtained in the laboratory for the reference and routine method, respectively, and s_A^2 and s_B^2 are the variances observed in the laboratory for the reference and the routine method, respectively. Most often, measurements are performed under repeatability or intermediate precision conditions [15].

This of course implies that s_A^2 and s_B^2 are estimates of the same σ so that they can be pooled. If this is not the case,

$$s_d = \sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}$$

and an appropriate test such as Cochran's test has to be used [5].

If the test reveals that the bias of the method under study is not significant, Eq. 16 therefore becomes:

$$u_{\bar{x}} = \sqrt{u_{\text{ref.meth}}^2 + s_d^2 + s_{\text{run}}^2 + s_r^2/n} \tag{19}$$

which reduces to:

$$u_{\bar{x}} = \sqrt{s_d^2 + s_{\text{run}}^2 + s_r^2/n} \tag{20}$$

if, as mentioned before, the uncertainty in the reference method is considered negligible.

EURACHEM [1] recommends the same approach for the uncertainty statement if the bias is found to be significant and if, as required by ISO, the measurement result is corrected for the bias. This approach is also preferred by IUPAC [14]. If no correction is applied, the observed bias and its associated uncertainty are reported in addition to the result [1]. When a significant bias is not considered relevant, and therefore is not corrected for, the approach proposed by IUPAC [14] for recovery experiments (explained in Section 5.3.2) could also be applied.

One should note that in the uncertainty expressed as in Eq. 20 some terms appear more than once. Indeed, if in the evaluation of the bias the measurements performed with the routine method are obtained, e.g., under repeatability conditions, s_A^2 is equal to s_r^2. Some documents, such as that by the BCR concerning the application of metrology in chemistry and biology [16], do not then additionally include the repeatability uncertainty into the uncertainty statement, since it has already been considered in the process of assessing the bias. This does not, however, seem to be correct, since the bias is to be considered a source of uncertainty additional to the random errors.

5.3.2. Recovery

Another approach to evaluating the bias is by a recovery experiment. The IUPAC report discusses the problems related to recovery estimation and the possibilities of corrections [14]. Barwick and Ellison [17] focus on the evaluation of the uncertainties associated with recovery. Even though recovery is usually regarded in a relative way, absolute expressions are used in the following, since this corresponds with the other expressions used. In the recovery experiment, a known amount of analyte is added to the sample matrix. The bias can then be estimated as the difference between the concentra-

tion recovered, $\bar{x}_{rec}$, and the known spiked concentration, c_{spike}:

$$\text{est.bias} = x_{rec} - c_{spike} \quad (21)$$

If blank sample material is available, $\bar{x}_{rec}$ is the concentration observed for the spiked sample. The uncertainty associated with the bias can then be calculated as:

$$u_{bias} = \sqrt{s_{obs}^2/m + u_{spike}^2} \quad (22)$$

where s_{obs}^2 represents the variance of the replicate analyses performed on the spiked sample, and m is the number of replicate analyses performed on the spiked sample. This means that the uncertainty of the traceability of Eq. 3 is here the uncertainty of the spiked concentration. However, this uncertainty will usually be negligible.

If no blank material is available the concentration recovered is obtained from the measurement of the sample before and after the addition of the analyte, $\bar{x}_{rec} = \bar{x}_{obs} - \bar{x}_{native}$. The bias is estimated according to Eq. 21. In this situation, the uncertainty associated with the bias contains a contribution from the repeated measurements of the sample after the addition of the analyte as well as before the addition:

$$u_{bias} = \sqrt{s_{obs}^2/m_1 + s_{native}^2/m_2} \quad (23)$$

where s_{obs}^2 again represents the variance of the replicate analyses performed on the sample after the addition, and m_1 is the number of replicate analyses performed; s_{native}^2 is the variance of the replicate analyses performed before the addition and m_2 is the number of replicates. As already mentioned, the uncertainty of the spiking is considered negligible.

If s_{obs}^2 and s_{native}^2 are estimates of the same variance, they can be pooled. The uncertainty associated with the bias

$$u_{bias} = \sqrt{\frac{(m_1-1)s_{obs}^2 + (m_2-1)s_{native}^2}{m_1 + m_2 - 2}\left(\frac{1}{m_1} + \frac{1}{m_2}\right)} \quad (24)$$

is then again the standard deviation that appears in the t-test applied to test whether the spiked concentration has been recovered.

As for the method comparison, the uncertainty associated with a non-significant bias should be considered in the uncertainty statement. If the bias is not significant, or if a significant bias is corrected for, the uncertainty of the mean result of n replicated measurements is expressed by Eq. 16:

$$u_{\bar{x}} = \sqrt{u_{bias}^2 + s_{run}^2 + s_r^2/n}$$

IUPAC [14] mentions a pragmatic approach for the situation when a significant bias is not considered relevant, and therefore is not corrected for. It consists in adding the absolute value of the uncorrected bias to the expanded uncertainty, the latter being calculated assuming that the bias is zero. This approach is also followed by Barwick and Ellison [17], but they include the recovery (expressed in a relative way) in the standard uncertainty by taking into account the coverage factor, k, that will be used in the calculation of the expanded uncertainty. With the same approach applied here, Eq. 23, for example, would become:

$$u_{bias} = \sqrt{\left(\frac{\text{est.bias}}{k}\right)^2 + \frac{s_{obs}^2}{m_1} + \frac{s_{native}^2}{m_2}} \quad (25)$$

However, according to the GUM [2], corrections should always be applied. Consequently, the later IUPAC document [6] as well as EURACHEM [1] do not recommend this approach.

Alternatively, when a significant bias is not considered relevant and is not corrected for, IUPAC [14] proposes that one should increase u_{bias} and proceed as if the bias was not significant. The increased u_{bias} is calculated as $|\text{est.bias}| / t_{crit}$ with t_{crit} the tabulated t value used in the significance test. This increased uncertainty, u_{bias}, thus corresponds to the uncertainty that in the significance test would just lead to the conclusion that the bias is not significant. However, according to IUPAC, all these approaches lead to an overstatement of the uncertainty. Therefore, as already mentioned earlier, IUPAC [6] as well as EURACHEM [1] recommend that one should always correct for the bias, or report the observed bias and its uncertainty in addition to the result.

5.4. Absolute uncertainty

The way in which bias was assessed in Section 5.3.2 does not always provide full traceability to the SI or to the highest metrological level possible. Primary methods allow for a direct traceability to the SI

[16]. A direct traceability to the SI is also possible with primary standards [18]. However, these primary standards only comprise pure chemicals, and therefore they cannot be used as direct reference for analyses with complicated matrices. The reference materials with the highest level in the hierarchy that is suitable for matrix analysis are CRMs [18]. In a method-independent context, Valcárcel and Rios [19] also consider the value of a CRM as the highest achievable real reference in the metrological hierarchy. The high reliability of these materials stems from the fact that their content is determined by different laboratories using different methods, so that laboratory and method biases are eliminated to the highest degree possible. If the bias of a certain method is assessed by the analysis of a CRM with a matrix similar to the one under study, the most reliable estimate of the bias is obtained. If a precision estimate of the routine method is available, the uncertainty can be estimated using Eq. 16. The uncertainty associated with the bias again splits up into the uncertainty of the estimated bias and the traceability, which here is the uncertainty specified for the CRM by the certification organisation:

$$u_{\text{bias}} = \sqrt{u_{\text{CRM}}^2 + u_{\text{est.bias}}^2} \qquad (26)$$

The uncertainty in the estimated bias, $u_{\text{est.bias}}$, corresponds to the uncertainty associated with the difference between the concentration measured for the CRM and the certified value. This again corresponds to the standard deviation term $s_d = s_{(\bar{x}-\mu)} = s_{\bar{x}}$ that appears in the t-test often applied to test whether the difference is statistically significant:

$$t = \frac{|\bar{x}-\mu|}{s_{\bar{x}}} \qquad (27)$$

with μ being the certified value, $\bar{x}$ the mean result obtained in the laboratory for the CRM and $s_{\bar{x}}^2$ the variance of the mean concentration observed for the CRM. Most often, measurements are performed under repeatability or intermediate precision conditions.

Notice that in this approach the uncertainty in the certified value is considered negligible compared with the method precision and, therefore, it is not taken into account in the significance test.

Alternative approaches that take the uncertainty in the certified value into account have been proposed [16,20,21]. The following formula is then used as the criterion for acceptance:

$$-2\sqrt{u_{\text{CRM}}^2 + \sigma^2} \le \bar{x}-\mu \le 2\sqrt{u_{\text{CRM}}^2 + \sigma^2} \qquad (28)$$

with σ^2 being the precision of the measurement results which correspond to $s_{\bar{x}}^2$ defined beforehand. The uncertainty in the CRM, u_{CRM}, has to be obtained from the certificate. As pointed out by Jorhem [22], the uncertainty intervals described in the certificates may have different meanings and are not always easy to understand. Jorhem [22] also shows that the often reported 95% confidence interval calculated on the basis of the mean value for each participating laboratory is suitable for the characterisation of the CRM but not for the evaluation of individual laboratory measurements. More detailed certification reports including information on how to use the CRM are required for an acceptable evaluation of individual results [23].

For some analytical problems, the highest level of traceability can also be reached by using a primary method. The EURACHEM guideline [1] gives a current definition of a primary method: "A primary method of measurement is a method having the highest metrological qualities, whose operation is completely described and understood in terms of SI units and whose results are accepted without reference to a standard of the same quality".

As already mentioned in an earlier section, methods accepted as having these properties are IDMS, coulometry, gravimetry, and titrimetry [8]. The results of these methods are usually directly traceable to the SI and, as a consequence, they should provide the smallest uncertainty achievable. However, for a large number of analytical problems, other techniques are used routinely. Nevertheless, primary methods are used for standardisation purposes by National Measurement Institutes. In order to reduce the uncertainty of the traceability, they can be used as reference methods to estimate the bias of a routine method. If primary methods and/or primary standards are used in the certification study for CRMs, the latter are considered secondary standards and have a higher traceability than other CRMs [18].

6. Conclusions

To compute uncertainty statements, it is possible and preferable that one should use as much as pos-

sible information obtained from method validation and other quality assurance procedures. In this way, the uncertainty expression becomes a natural extension of the validation of methods, which is, for analytical chemists, a much better understood concept than the component-by-component approach and therefore will be adopted much more easily. The first edition of the EURACHEM Guide for 'Quantifying Uncertainty in Analytical Measurement' [24] insisted very much on the component-by-component approach and, in this way, has caused unnecessary confusion. The second edition [1] follows closely a document prepared by IUPAC, AOAC, FAO and IAEA [6] and stresses to a much larger extent that "Measurement uncertainty must be integrated with its existing quality assurance measurements, with these measures themselves providing much of the information required to evaluate the measurement uncertainty". In its draft version [25], it goes on to say that "It attempts to correct the impression gained within the Analytical Community that it is only the so-called component-by-component approach that is acceptable...". The final version refers to the ISO guide 17025 [26], which also allows one the use of approaches for the uncertainty evaluation other than the component-by-component approach. The second edition of the EURACHEM guide is a very useful document – much more so than the first one. It is, however, somewhat unfortunate that the examples cited give much more room to the component-by-component approach than to those based on method validation data.

We ourselves consider that, when presenting the subject, it is better to start with equations in which sums of variances or sums of relative variances are given, rather than with the error propagation approach of Eq. 9. The latter approach certainly has its merits and is useful when going further into the subject, but it is preferable, when talking to analytical chemists, to start with concepts to which they are used.

Although the documents cited above form a good working basis, they suffer from one defect – that the term 'uncertainty' is used for many very different situations. It is our opinion that operational definitions of uncertainty are required so that a distinction is made according to the reference to which the result is traced and the sources of uncertainty that are included.

Acknowledgements

This work was supported by the Research contract No. NM/03/24 of the Belgian Government (The Prime Ministers Service – Federal Office for Scientific Technical and Cultural Affairs, Standardisation Program).

References

[1] EURACHEM/CITAC Guide, Quantifying Uncertainty in Analytical Measurement, 2nd edition, 2000.
[2] Guide to the Expression of Uncertainty in Measurement, ISO, Geneva, 1995.
[3] International Vocabulary of Basic and General Terms in Metrology (VIM), ISO, Geneva, 2nd edition, 1993.
[4] S. Ellison, W. Wegscheider, A. Williams, Anal. Chem. 69 (1997) 607A.
[5] D.L. Massart, B.G.M. Vandeginste, L.M.C. Buydens, S. de Jong, P.J. Lewi, J. Smeyers-Verbeke, Handbook of Chemometrics and Qualimetrics, Part A, Elsevier, Amsterdam, 1997, pp. 56, 97, 395.
[6] Report on the FAO, IAEA, AOAC Int., IUPAC, International Workshop on Principles and Practices of Method Validation, 4–6 November 1999, Budapest.
[7] NMKL, Procedure No. 5, Estimation and Expression of Measurement Uncertainty in Chemical Analysis, Nordic Committee on Food Analysis, 1997.
[8] R. Kaarls, T.J. Quinn, The Comité Consultatif pour la Quantité de Matière: a Brief Review of its Origin and Present Activities, Metrologia 34 (1997) 1, cited in BCR Report, EUR 18405 EN, Community Bureau of Reference, 1998.
[9] A. Dobney, H. Klinkenberg, F. Souren, W. Van Borm, Anal. Chim. Acta 420 (2000) 89.
[10] R. Wood, A. Nilsson, H. Wallin, Quality in the Food Analysis Laboratory, Royal Society of Chemistry Monographs, Cambridge, 1998.
[11] G.T. Wernimont, Use of Statistics to Develop and Evaluate Analytical Methods, AOAC, Arlington, VA, 1985, p. 104.
[12] W. Horwitz, L.R. Kamps, K.W. Boyer, J. Assoc. Off. Anal. Chem. 63 (1980) 1344.
[13] International Standard, Accuracy (Trueness and Precision) of Measurement Methods and Results, ISO 5725-6-1994, Geneva, 1994, especially paragraph 8.1.
[14] Harmonised Guidelines for the Use of Recovery Information in Analytical Measurement, Technical Report Resulting from the Symposium on Harmonization of Quality Assurance systems for Analytical Laboratories, IUPAC, ISO, AOAC Int. and EURACHEM, Orlando, FL, 4–5 September 1996.
[15] S. Kuttatharmmakul, D.L. Massart, J. Smeyers-Verbeke, Anal. Chim. Acta 391 (1999) 203.
[16] BCR Report, EUR 18405 EN, Community Bureau of Reference, 1998.

[17] V.J. Barwick, S.L.R. Ellison, Analyst 124 (1999) 981.
[18] R. Walker, I. Lumley, Trends Anal. Chem. 18 (1999) 594.
[19] M. Valcárcel, A. Rios, Trends Anal. Chem. 18 (1999) 68.
[20] ISO Guide 35:1989 (E), Certification of Reference Materials – General and Statistical Principles, ISO, Geneva, 1989.
[21] R. Walker, The selection and use of reference materials: some examples produced by LGC, in: A. Fajelj, M. Parkany (Editors), The Use of Matrix Reference Materials in Environmental Analytical Processes, Royal Society of Chemistry, Cambridge, 1999, p. 155.
[22] L. Jorhem, Fresenius' J. Anal. Chem. 360 (1998) 370.
[23] J. Pauwels, How to use matrix certified reference material, in: A. Fajelj, M. Parkany (Editors), The Use of Matrix Reference Materials in Environmental Analytical Processes, Royal Society of Chemistry, Cambridge, 1999, p. 41.
[24] Quantifying Uncertainty in Analytical Measurement; published on behalf of EURACHEM by the LGC, London, 1995.
[25] EURACHEM/CITAC Guide, Quantifying Uncertainty in Analytical Measurement, 2nd editionn, Draft, June 1999.
[26] ISO/IEC Standard 17025, General Requirements for the Competence of Testing and Calibration Laboratories, ISO, Geneva, 1999.

Edelgard Hund graduated in chemistry from the University of Regensburg, Germany, in 1997. She is now preparing a Ph.D. thesis in analytical chemistry at the Vrije Universiteit Brussel (VUB), Brussels, Belgium, in the laboratory of analytical chemistry of the Pharmaceutical Institute. Professor Johanna (An) Smeyers-Verbeke obtained her Ph.D. in chemistry in 1977 at the VUB. She is now responsible for parts of the courses in Analytical Chemistry in the Pharmaceutical Institute of the VUB. Professor Désiré Luc Massart is Head of Department of the Laboratory of Analytical Chemistry of the Pharmaceutical Institute of the VUB. He obtained his Ph.D. in chemistry at the University of Ghent, and in 1968 moved to the University of Brussels where he is mainly interested in chemometrics.