Accred Qual Assur (1998) 3:231–236

Reports and notes on experiences with quality assurance, validation and accreditation

Paul Willetts
Roger Wood

Measurement uncertainty – a reliable concept in food analysis and for the use of recovery data?

Received: 4 November 1997
Accepted: 3 February 1998

Presented at: 2nd EURACHEM Workshop on Measurement Uncertainty in Chemical Analysis, Berlin, 29–30 September 1997

P. Willetts · R. Wood (✉)
Food Labelling and Standards Division, Ministry of Agriculture, Fisheries and Food, CSL Food Science Laboratory, Norwich Research Park, Colney, Norwich NR4 7UQ, UK
Tel.: +44-1603-259350
Fax: +44-1603-501123
e-mail: r.wood@fscii.maff.gov.uk

Abstract Steps which are taken to implement the concept of measurement uncertainty in analytical chemical laboratories should take full account of existing internationally agreed protocols for analytical quality assurance and reflect the needs of particular analytical sectors. For the food sector this may mean that for official purposes the use of the term measurement uncertainty is replaced by the term measurement reliability and that a quantitative estimation of this is made based on existing collaborative trial data. In many analytical sectors, the differing strategies currently followed for the determination and use of recovery information are an important cause of the non-comparability of analytical results. Guidelines which are being prepared for the estimation and use of recovery information in analytical measurement may provide a more unified approach which includes measurement uncertainty as a key concept in the use of recovery data.

Introduction

Recent years have seen the issue of the quality and reliability of data become of paramount importance in all analytical sectors. In order to address this matter, analysts from the different analytical sectors have worked together under the sponsorship of ISO, IUPAC and AOAC INTERNATIONAL, to produce International Harmonised Protocols on the subjects of the collaborative testing of analytical methods [1], proficiency testing [2] and the use of internal quality control in analytical chemistry laboratories [3]. In addition, the use of certified reference materials is increasingly being advocated with respect to the traceability of analytical data [4], and laboratory accreditation schemes are being widely implemented.

Each of these components of analytical quality assurance concerns a different aspect of data reliability, namely the external testing of methods and laboratory performance, internal data quality, trueness and the auditing of procedures and records. With the exception of the latter which is administratively based, in each case reliability is limited by either, or in many cases both, systematic or random experimental 'inaccuracies', quantities now being embraced by the concept of measurement uncertainty.

Requirements and initiatives in the food sector

In introducing this concept to the analytical chemical community there is a need to ensure that steps taken to implement measurement uncertainty are made in the context of the existing protocols and strategies in analytical quality assurance. Moreover, the needs and actions of particular analytical sectors must also be recognised. In the food sector a number of initiatives have

been advanced recently which affect specifically the issue of data quality, and therefore reliability, in this area of analysis. Firstly, in the EU, there is a tendency in the food analysis sector to not prescribe specific methods of analysis but to adopt a "criteria of methods approach" whereby analysts may use the method of their choice provided it meets certain prescribed quality criteria. This flexibility of approach, to take advantage of the developments of new techniques and procedures as they occur in analytical chemistry, clearly has consequences for the comparability and measurement uncertainty of reported data. Secondly, there is a requirement in the food sector, as set out in EC Directive 93/99, that methods of analysis for food control purposes should wherever possible be formally validated by collaborative trial [5]. Thirdly, there have been discussions on measurement uncertainty within the Codex Committee on Methods of Analysis and Sampling. The Report of the March 1997 Session of that Committee states that with regard to measurement uncertainty [6]:

1. The Committee will develop for Codex purposes an appropriate alternative term for measurement uncertainty, e.g. measurement reliability.
2. The precision of a method may be estimated through a method-performance study, or where this information is not available, through the use of internal quality control and method validation.
3. Consideration should be given as to whether it is necessary to undertake an additional formal evaluation of a method of analysis using the ISO approach [7] in addition to using information obtained through a collaborative trial.
4. Governments should advise accreditation agencies that for national and Codex purposes the measurement uncertainty result need not be calculated using the ISO approach [7] providing the laboratory is complying with the appropriate Codex principles.

Discussions are on-going in Codex. However, if these proposals are accepted, it is likely that the term 'measurement reliability' rather than measurement uncertainty will be adopted and that estimates of this will be made from collaborative trial data if such data are available. In a recent study, carried out in the UK, which compared 'top-down' (collaborative trial) and 'bottom-up' (ISO) approaches to the estimation of measurement uncertainty, it was concluded that for comparable matrix/analyte combinations these approaches gave not dissimilar results in the limited number of cases studied [8]. It should be noted that, in recognising the importance of the concept of measurement uncertainty in underpinning the reliability of analytical data, the Codex recommendations and discussions are in accordance with statements on uncertainty in ISO Guide 25 and EN 45001, which require accreditation agencies to ensure that measurement uncertainty estimations are carried out as part of the accreditation process [9].

Recovery and analyte losses

One aspect of analytical chemistry where, for all analytical sectors including the food sector, current practice continues to have important consequences in terms of the non-comparability and uncertainty or reliability of reported data, is that of the use of recovery information. This arises because of the different strategies for dealing with recovery assessment and the effect these may have on the variability of the analytical results reported.

Recovery studies are an essential component of quality assurance systems in analytical measurement. Their use, particularly in the trace analyte area, to assess the efficiency of the removal of the measurand from the sample matrix and its transfer prior to detection is widely quoted in the scientific literature. Although they thus provide an important indication of the reliability of these steps in the measurement process, there generally has been no consistent approach to the way in which recovery information is derived and used in analytical data. In particular, in the case of recovery factors calculated and applied to analytical data to correct for displacement or bias, the absence of accepted strategies for the determination and use of these factors has meant that it frequently has been difficult to make comparisons between analytical results produced in different laboratories or verify the suitability of that data for the intended purpose. This is particularly marked in the case of complex matrices, such as foodstuffs, where the difficulties of completely extracting the analyte are most pronounced. Quite commonly in such procedures a substantial proportion of the analyte remains in the matrix after extraction, so that the transfer is incomplete, and the subsequent measurement is lower than the true concentration in the original test material. If no compensation for these losses is made, then markedly discrepant results may be obtained by different laboratories. Even greater discrepancies are likely to arise if some laboratories compensate for losses and others do not. These considerations are especially important in legislative/enforcement situations where for instance the difference between applying or not applying a recovery factor to correct for the incomplete removal of the analyte may mean respectively that a legislative limit is exceeded or that a result is in compliance with the limit.

Recovery correction factors

Thus, where an estimate of the *true concentration* is required, there is a compelling case for including a compensation for losses in the calculation of the reported analytical result, provided that the correction factor can be estimated reliably. In the case of an empirical method, where the measurand is defined in terms of the method used and no attempt is being made to estimate the amount of analyte actually present in the sample matrix, the question whether or not a correction is applied is a matter for the definition of the empirical method.

The four most common approaches which typically have been taken by analysts in respect of the application of recovery factors are shown in Table 1.

Table 1 Typical approaches relating to the application of recovery factors

a	The reporting of an analytical result without correcting for bias by the application of a recovery factor, no accompanying statement being given of the level of recovery achieved
b	The reporting of an analytical result without correcting for bias by the application of a recovery factor, together with a statement of the level of recovery achieved
c	The reporting of an analytical result corrected for bias by the application of a recovery factor, without an accompanying statement of the level of recovery
d	The reporting of an analytical result corrected for bias by the application of a recovery factor, together with a statement of the level of recovery achieved

Reference materials and spiking experiments

Quite apart from the variation which can arise from laboratories adopting different practices in respect of whether a correction factor is applied or is not applied to an analytical result, a further aspect which can hinder data comparison is the fact that 'recovery' information may be derived either from the inclusion of reference materials or the use of spiked samples.

In the case of reference materials, the analyte is usually integrated or incorporated into the matrix, whereas in the case of spiked samples the analyte is merely added to the matrix. Potentially different information relating to the behaviour of the native analyte to be measured may be derived from each type of recovery measurement. Moreover, the regularity and pattern of use of these recovery materials may affect the recovery information produced. In the case of spiking, for example, the different ways in which the recovery factor may be determined include those shown in Table 2.

Table 2 Examples of ways in which the recovery factor may be determined with spiking

a	Basing a recovery correction factor on the recovery of the analyte from a spiked sample in the batch
b	Basing a recovery correction factor on the mean value obtained for the recovery of the analyte spiked into a sample in each of a number of batches
c	Basing a recovery correction factor on the recovery of a chemically similar internal standard added to the test material
d	Basing a recovery correction factor on the recovery of an isotopic form of the analyte added as internal standard to the test material

Each of these approaches differs in the representativeness it provides of the actual extraction of the analyte itself, the basis of the representation being different in each case. While it is generally agreed that, of these four alternatives, the use of an isotopic internal standard is the preferred approach since the recovery of the auxiliary analyte equates most closely to being 'fully equivalent' to that of the target analyte, this option is often not possible. As a consequence one of the other alternatives is often followed in spiking experiments.

When a reference material is used rather than spiking, then it will be included at a different position in the batch to the test material itself. In this respect the use of a reference material is akin to options a or b for spiking (see Table 2).

Consideration of these different strategies has led analytical chemists to recognise the desirability of using a more uniform approach when dealing with the topic of recovery measurements in order to facilitate the comparability of data.

Guidelines for using recovery information

Following the circulation to a broad cross section of the analytical community world-wide of a questionnaire on the determination and use of recovery measurements in 1995, background information was obtained which enabled further consideration to be given to the role of recovery studies in chemical analysis [10]. The main questions addressed the issues shown in Table 3.

As expected, the differing answers given to the questions posed revealed considerable variation in the ways in which analysts deal with recovery measurements. In particular, the question on measurement uncertainty itself produced more differences than any of the other questions, perhaps suggesting a lack of appreciation of either the need for or the means of calculating this value. The findings of this survey were presented at the

Table 3 Outline of questions included in the recovery factors questionnaire

Question number	Question	Question number	Question
1	Meaning of recovery	12	Recovery of analyte and internal standard
2.1	Purpose/use of recovery measurements	13	Spiking procedure
2.2	Reporting results	14	Blank material
3	Recovery frequency in time	15	Spiking level
4	Recovery frequency in batch	16	Spiking concentration
5	Recovery level	17	Carrier solvent for analyte
6	Acceptable recovery levels	20	Recovery solution
7	Assessment of acceptable recovery	21	Time of sample preparation
9	Multi-analyte determinations	23	Precision
10	Procedure for the determination of recovery	24	Measurement uncertainty
11	Matrices used		

Symposium on Harmonisation of Quality Assurance Systems in Chemical Analysis held in Orlando, USA, in 1996. From the deliberations of that meeting, harmonised guidelines for the use of recovery information in analytical measurement are being prepared under the sponsorship of IUPAC, ISO and AOAC INTERNATIONAL [11]. The guidelines, the main points of which are summarised in Table 4, refer to uncertainty as being a key concept in formulating an approach to the estimation and use of recovery information.

Uncertainty and recovery correction

Although the estimation of uncertainty in recovery has yet to be studied in detail, the guidelines list some sources of the uncertainty in measured recovery (Table 5) and include a treatment which considers the uncertainty estimation in cases of incomplete recovery where either a correction is or is not applied to an analytical result [12]. In this treatment, the difference in the measured recovery (R) from the value of unity, representing total recovery, is compared to the uncertainty in the determination of R.

The comparison is made using a significance test to assess whether $|R-1|$ is greater than the uncertainty

Table 4 A summary of guidelines for the use of recovery information

1. A distinction is recognised between:
   surrogate recovery (recovery of a pure compound or element specifically added to the test portion or test material as a spike – sometimes called "marginal recovery")
   the recovery of native analyte incorporated into the test material by natural processes and manufacturing procedures – sometimes called "incurred analyte".
2. It is recognised that there is a dual role for recovery determinations in analytical measurement, that is, (a) for quality control purposes and (b) for deriving values for recovery factors. In the latter application, more extensive and detailed data are required.
3. Variable practice in handling recovery information is an important cause of the non-equivalence of data. To mitigate its effects, in general, results should be corrected for recovery, unless there are overriding reasons for not doing so. Such reasons would include the situation where a limit (statutory or contractual) has been established using uncorrected data, or where recoveries are close to unity.
4. It is of over-riding importance that (a) all data, when reported, should be clearly identified as to whether or not a recovery correction has been applied and (b) if a recovery correction has been applied, the amount of the correction and the method by which it was derived should be included with the report. This will promote direct comparability of data sets. Thus, in all situations, correction functions should be established based on appropriate statistical considerations, documented, archived and available to the client.
5. Recovery values should always be established as part of method validation, whether or not recoveries are reported or results are corrected, so that measured values can be converted to corrected values and vice versa.
6. When the use of a recovery factor is justified, the method of calculation should be given in the method.
7. IQC control charts for recovery should be established during method validation and used in all routine analysis. Runs giving recovery values outside the control range should be considered for re-analysis in the context of acceptable variation, or the results should be reported as semi-quantitative.
8. Uncertainty is a key concept in formulating an approach to the estimation and use of recovery information. Although there are substantive practical points in the estimation of uncertainty that remain to be settled, the principle of uncertainty is an invaluable tool in conceptualising recovery issues.

Table 5 Sources of uncertainty in recovery estimation

1	Repeatability of the recovery experiment
2	Uncertainties in reference material values
3	Uncertainties in added spike quantity
4	Poor representation of native analyte by the added spike
5	Poor or restricted match between experimental matrix and the full range of sample matrices encountered
6	Effect of analyte/spike level on recovery and imperfect match of spike or reference material analyte level and analyte level in samples

(u_R) in the determination of R, at some level of confidence. The significance test takes the form

$|R-1|/u_R > t$: R differs significantly from 1
$|R-1|/u_R \leq t$: R does not differ significantly from 1

where t is a critical value based either on a 'coverage factor' allowing for practical significance or, where the test is entirely statistical, $t_{(\alpha/2,\ n-1)}$, being the relevant value of Student's t for a level of confidence $1-\alpha$.

Following this assessment, for a situation where incomplete recovery is achieved, four cases can be distinguished, chiefly differentiated by the use made of the recovery R.

(a) R is not significantly different from 1. No correction is applied.
(b) R is significantly different from 1 and a correction for R is applied.
(c) R is significantly different from 1 but, for operational reasons, no correction for R is applied
(d) An empirical method is in use. R is arbitrarily regarded as unity and u_R as zero. (Although there is obviously some variation in recovery in repeated or reproduced results, that variation is subsumed in the directly estimated precision of the method.)

In the first case, where R is not significantly different from 1, the recovery can be viewed as being equal to unity, no correction being applied. There is still an uncertainty, u_R, about the recovery that contributes to the overall uncertainty of the analytical result.

In the cases where R is significantly different from 1, the loss of analyte occurring in the analytical procedure is taken into account, and two uncertainties need to be considered separately. First, there are the uncertainties associated only with the determination, namely those due to gravimetric, volumetric, instrumental, and calibration errors. That relative uncertainty u_x/x will be low unless the concentration of the analyte is close to the detection limit. Second, there is the uncertainty u_R on the estimated recovery R. Here the relative uncertainty u_R/R is likely to be somewhat greater. If the raw result is corrected for recovery, we have $x_{\text{corr}} = x/R$ (*i.e.*, the correction factor is $1/R$). The relative uncertainty on x_{corr} is given by

$$\frac{u_{\text{corr}}}{x_{\text{corr}}} = \sqrt{\left(\left(\frac{u_x}{x}\right)^2 + \left(\frac{u_R}{R}\right)^2\right)}$$

which is necessarily greater than u_x/x and may be considerably greater. Hence correction for recovery seems at first sight to degrade, perhaps substantially, the reliability of the measurement.

It is stated that such a perception is incorrect. Only if the method is regarded as empirical, and this has drawbacks in relation to comparability as already discussed, is u_x the appropriate uncertainty. If the method were taken as rational, and the bias due to loss of analyte were not corrected, a realistic estimate u_x would have to include a term describing the bias. Hence u_x/x would be at least comparable with, and may be even greater than, $u_{\text{corr}}/x_{\text{corr}}$.

These approaches to the estimation of the uncertainty of a recovery are necessarily tentative. Nevertheless, the following important principles of relevance to the conduct of recovery experiments are demonstrated.

(a) The recovery and its standard uncertainty may both depend on the concentration of the analyte. This may entail studies at several concentration levels.
(b) The main recovery study should involve the whole range of matrices that are included in the category for which the method is being validated. If the category is strict (e.g., bovine liver) a number of different specimens of that type should be studied so as to represent variations likely to be encountered in practice (e.g., sex, age, breed, time of storage etc.). Probably a minimum of ten diverse matrices are required for recovery estimation. The standard deviation of the recovery over these matrices is taken as the main part of the standard uncertainty of the recovery.
(c) If there are grounds to suspect that a proportion of the native analyte is not extracted, then a recovery estimated by a surrogate will be biased. That bias should be estimated and included in the uncertainty budget.
(d) If a method is used outside the matrix scope of its validation, there is a matrix mismatch between the recovery experiments at validation time and the test material at analysis time. This could result in extra uncertainty in the recovery value. There may be problems in estimating this extra uncertainty. It would probably be preferable to estimate the recovery in the new matrix, and its uncertainty, in a separate experiment.

References

1. Horwitz W (1988) Pure Appl Chem 60:855–864
2. Thompson M, Wood R (1993) Pure Appl Chem 65:2123–2144 [Also published in J. AOAC Int (1993) 76:926–940]
3. Thompson M, Wood R (1995) Pure Appl Chem 67:649–666
4. Thompson M (1996) Analyst 121:285–288
5. Official Journal of the European Communities (1993) L290/14 Council Directive 93/99/EEC
6. Codex Alimentarius Commission (1997) Codex Committee on Methods of Analysis and Sampling 21st Session

7. ISO (1993) Guide to the Expression of Uncertainty in Measurement, Geneva
8. Brereton P, Anderson S, Willetts P, Ellison S, Barwick V, Thompson M, Wood R (1997) CSL Report FD 96/103
9. Draft - ISO/IEC Guide 25 (1996) General requirements for the competence of testing and calibration laboratories
10. Willetts P, Anderson S, Wood R (1998) CSL Report FD 97/65
11. International Union of Pure and Applied Chemistry (1997) Draft Harmonised Guidelines for the use of Recovery Information in Analytical Measurement
12. Ellison SLR, Williams A (1996) In: Parkany M (ed) Proceedings of the Seventh International Symposium on the Harmonisation of Quality Assurance Systems in Chemical Analysis. Royal Society of Chemistry, London