REPORT ON THE RELATIONSHIP BETWEEN ANALYTICAL RESULTS, MEASUREMENT UNCERTAINTY, RECOVERY FACTORS AND THE PROVISIONS OF EU FOOD AND FEED LEGISLATION, WITH PARTICULAR REFERENCE TO COMMUNITY LEGISLATION CONCERNING

- CONTAINANTS IN FOOD (COUNCIL REGULATION (EEC) No 315/93 OF 8 FEBRUARY 1993 LAYING DOWN COMMUNITY PROCEDURES FOR CONTAMINANTS IN FOOD[1])

- UNDESIRABLE SUBSTANCES IN FEED (DIRECTIVE 2002/32/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL OF 7 MAY 2002 ON UNDESIRABLE SUBSTANCES IN ANIMAL FEED[2])

CONTENTS

1. INTRODUCTION
2. ISSUES INVOLVED
3. NUMBER OF SIGNIFICANT FIGURES TAKEN INTO ACCOUNT WHEN REPORTING RESULTS AND INTERPRETING THEM IN RELATION TO STATUTORY LIMITS
3.1 Introduction
3.2 Solution
4. REPORTING OF RESULTS WITH RESPECT TO THEIR MEASUREMENT UNCERTAINTY
4.1 Introduction
4.2 Reporting of results by food and feed control analysts
4.3 Consequences of reporting results in different ways
4.4 Solution
4.5 Procedures for estimating measurement uncertainty
4.6 Value of the measurement uncertainty
5. USE OF RECOVERY INFORMATION IN ANALYTICAL MEASUREMENT
6. OTHER LEGISLATION
7. RECOMMENDATIONS
8. LOOKING AHEAD

[1] OJ L 37, 13.2.1993, p. 1
[2] OJ L 140, 30.5.2002, p. 10

ANNEXES

ANNEX I: DIAGRAMMATIC ILLUSTRATION OF THE EFFECT OF MEASUREMENT UNCERTAINTY AND THE LIMIT

ANNEX II: PROCEDURES FOR THE ESTIMATION OF MEASUREMENT UNCERTAINTY

ANNEX II.1: ISO GUIDE TO THE EXPRESSION OF MEASUREMENT UNCERTAINTY

ANNEX II.2: EURACHEM GUIDE TO QUANTIFYING UNCERTAINTY IN ANALYTICAL MEASUREMENT

A. Component-by-component approach

B. Use of collaborative trial data

ANNEX II.3: USE OF COLLABORATIVE TRIAL: DATA – ISO 5725 CRITICAL DIFFERENCES

ANNEX II.4: DRAFT ISO TS 21748 – GUIDE TO THE USE OF REPEATABILITY, REPRODUCIBILITY AND ACCURACY ESTIMATES IN MEASUREMENT UNCERTAINTY ESTIMATION

ANNEX II.5: CONCEPT ESTABLISHED BY COMMISSION DECISION 2002/657/EC IMPLEMENTING COUNCIL DIRECTIVE 96/23/EC CONCERNING THE PERFORMANCE OF ANALYTICAL METHODS AND THE INTERPRETATION OF RESULTS

ANNEX II.6: AOAC INTERNATIONAL APPROACH

ANNEX II 7: INTERNAL QUALITY CONTROL APPROACH

ANNEX II.8: NMKL (NORDIC COMMITTEE ON FOOD ANALYSIS) APPROACH

ANNEX II.9: MICROBIOLOGICAL ANALYSES

ANNEX II.10: USEFUL GENERAL REFERENCES FOR MEASUREMENT UNCERTAINTY

ANNEX III PROCEDURES FOR ASSESSING RECOVERY

1. INTRODUCTION

It is now recognised that there are a number of actions that may be taken by those responsible for the enforcement of EU food and feed legislation which directly affect decisions as to whether a sample or batch from which a sample is taken is in compliance with an EU specification.

Before any specification is laid down in EU legislation, it must be understood that a specific parameter will depend on the procedures used to estimate it. In particular, an estimate of a specific value <u>may</u> be dependent upon the method of analysis used, but is <u>always</u> dependent on the method of sampling used to verify compliance with the specification. It is important for delegates in EU working groups to appreciate the influence that methods of analysis and sampling may have on the judgements that may be made with regard to the compliance of a batch with an EU specification. Without common and uniform criteria for methods of analysis and sampling procedures, and their common application and interpretation, different Member States will make different judgements as to whether a particular batch is in compliance with its EU specifications.

This report outlines the issues involved and makes recommendations and gives guidance to the enforcement authorities in Member States on procedures to be adopted to limit the possibility of Member States taking differing views as to whether a particular sample is in compliance with EU specifications.

This report focuses on analytical issues only. In particular, it looks at the role of analytical variability (normally known as "measurement uncertainty") in the interpretation of a specification.

The report is concerned with quantitative analytical results. Qualitative findings are also important but there are few, if any, internationally accepted recommendations and approaches for estimating their degree of uncertainty. However, these approaches are currently being developed.

This report is written in a form such that the complex issues involved can be readily appreciated by everyone. The report:

- sets out the issues;
- gives recommendations for consideration by the enforcement authorities in the Member States; and
- provides a series of technical annexes to help practitioners to estimate their measurement uncertainties.

It should be appreciated that the issues involved are real rather than hypothetical. Decisions have been taken by some Member States which in a given situation were contrary to those which would have been taken by other Member States in the same situation.

<u>IMPORTANT PRELIMINARY REMARK</u>: This report describes the relationship between analytical results, measurement uncertainty, recovery factors and provisions of EU food and feed legislation. However the <u>recommendations made in this report are for the time being only relevant to the application of Community legislation concerning contaminants in food</u> (Council Regulation 315/93 laying down Community procedures for contaminants in food) <u>and undesirable substances in feed</u> (Directive 2002/32/EC of the European Parliament and of the Council of 7 May 2002 on undesirable substances in animal feed).

2. ISSUES INVOLVED

There are a number of considerations which prevent the uniform implementation of legislative standards. In particular, problems concerning

1. the number of significant figures taken into account when reporting results and interpreting them in relation to statutory limits;

2. the treatment of analytical variability (or "measurement uncertainty") in the interpretation of a specification; and

3. the use of recovery correction when calculating and reporting an analytical result

are addressed in the report. The effect of different countries taking different approaches to each of these issues is described.

These aspects directly affect the interpretation of results in countries that apply EU legislation and so may be regarded as “food and feed controls”. Before 2003, there was no common interpretation of analytical results across the EU, with the result that different decisions may be taken after analysis of the "same" sample. Material for which there is a statutory limit of, for example, 4µg/kg for a contaminant, may be interpreted as containing 3µg/kg on analysis in one country but 10 µg/kg in another. This is because some countries correct analytical results for recovery whereas others do not; similarly, some take into account the measurement uncertainty associated with the analytical result while others do not.

It is essential that interpretation of analytical results be consistent if there is to be equivalence across the EU; without this consistency, there is no uniform interpretation of EU food and feed legislation. This is why provisions have been adopted in some EU Directives in order to ensure a uniform interpretation of analytical results.

It is stressed that this is not an analysis or sampling problem as such, but an administrative problem which has been highlighted as the result of recent activities in the analytical sector, most notably the development of international guidelines on the use of recovery factors when reporting analytical results, and various guides dealing with measurement uncertainty[3].

The differences described have also been identified as a result of the recently completed Scientific Cooperation Task 9.1[4].

As already mentioned, the recommendations are for the time being specifically aimed at the application of Community provisions concerning contaminants in food and undesirable substances in feed. Nevertheless, the above considerations may apply not only to the “contaminants" sector, but also to the additive, composition and microbiological aspects of food and feed analysis. Where appropriate, these aspects will also be addressed in this report.

---

[3] “Harmonised Guidelines for The Use of Recovery Information in Analytical Measurement”, Michael Thompson, Stephen L R Ellison, Ales Fajgelj, Paul Willetts and Roger Wood, Pure Appl. Chem., 1999, 71, 337 – 348

[4] Scientific Cooperation Task 9.1 on the “Preparation of a working document in support of the uniform interpretation of legislative standards and the laboratory quality standards prescribed under Directive 93/99/EEC”, downloadable from http://europa.eu.int/comm/food/fs/scoop/9.1_fr_en.pdf

3. NUMBER OF SIGNIFICANT FIGURES TAKEN INTO ACCOUNT WHEN REPORTING RESULTS AND INTERPRETING THEM IN RELATION TO STATUTORY LIMITS

3.1 Introduction

There are potential problems with the way in which maximum levels in legislation are interpreted by enforcement authorities and control analysts. This interpretation depends on the number of significant figures that are specified in any legislation, and therefore on the number of significant figures used when expressing an analytical result. Unless otherwise specified, some analysts will normally express analytical results using the same number of significant figures as prescribed in the relevant legislation. In order to avoid situations where analytical results are interpreted in relation to statutory limits in a non-harmonised manner, it is important to ensure that the statutory limit is uniform and consistent. This may be illustrated by an example:

Specification (independent of units)	Range within which a "satisfactory" result will lie
1	0 to 1.4
1.0	0 to 1.04
1.00	0 to 1.004

It is recognised that there are significant differences between maximum levels of 1 mg/kg, 1.0 mg/kg and 1.00 mg/kg, as illustrated above. It is essential that those setting the levels in legislation are fully aware of these differences. Officials involved in setting maximum levels may not be aware of the consequences of the form in which maximum levels are expressed, and should be fully aware of this when discussing maximum levels.

Recently issued legislation concerning maximum levels of contaminants (Commission Regulation (EC) No 466/2001) and undesirable substances (Directive 2002/32/EC of the European Parliament and of the Council of 7 May 2002 on undesirable substances in animal feed) and legislation on the methods of analysis to be used in official controls do not offer solutions to these problems. For example, Commission Directive 2001/22/EC on methods of analysis to be used in the control of e.g. lead, cadmium and mercury gives guidance on the expression of results only as regards the units to be used: these are the same as those in Commission Regulation 466/2001, but the latter Directive does not address the question of significant figures.

3.2 Solution

Legislation containing maximum levels should always look at how analytical results are to be expressed and interpreted. In general, when developing food and feed legislation, the number of significant figures to be laid down in the specifications must be appropriate to the specification under consideration. In cases where legislation already provides clear guidance on the number of significant figures to be specified (e.g. when maximum levels are expressed as 2.0, 3.0, and 0.40) then the analyst should report to the number of significant figures indicated in the specification. In other cases, and certainly in cases where this is appropriate for the precision of the result, the analyst should report to one more significant figure than is indicated in the specification, assuming that the analyst is using an appropriate method.

As a basic minimum, the following should be stated or considered when developing food and feed legislation:

- the units in which the results are to be expressed;
- the number of significant figures to be included in the reported result;
- the interpretation of an analytical result in relation to a statutory limit;
- the expected precision of the method of analysis likely to be used for the determination, and thus whether the number of significant figures being specified in legislation is “realistic”.

4. REPORTING OF RESULTS WITH RESPECT TO THEIR MEASUREMENT UNCERTAINTY

4.1 Introduction

All analytical results actually take the form "a ± 2u" or "a ± U" where "a" is the best estimate of the true value of the concentration of the measurand (the analytical result) and "u" is the standard uncertainty and "U" (equal to 2u) is the expanded uncertainty. "4u" is the range within which the true value is estimated, with a high probability, to fall. The value of "U" or "2u" is the value which is normally used and reported by analysts and is hereafter referred to as "measurement uncertainty" and may be estimated and expressed by analyst in a number of different ways.

With the introduction of ISO and other guidelines on uncertainty estimation, the accuracy available from analytical methods is increasingly characterised in terms of "measurement uncertainty", which takes into account both the "trueness" (average departure from a true value) and the "precision" (the degree to which successive results tally). The range within which "a" is likely to fall – i.e. the uncertainty in "a" – depends on the inherent "trueness" and precision of the analytical method as used in the laboratory.

Food control laboratories may be assumed to be "in control" by virtue of Council Directive 93/99/EEC of 29 October 1993 on the subject of additional measures concerning the official control of foodstuffs[5], which is currently in the process of being revised. The Directive requires of laboratories performing analyses for the official control of foodstuffs that they be formally accredited, participate in proficiency testing schemes, use internal quality-control procedures and use appropriately validated methods of analysis. These requirements have legal effect only for laboratories operating in the food sector. However, the draft feed and food control Regulation revises and extends these requirements to feed control laboratories as well.

4.2 Reporting of results by food and feed control analysts

The procedure adopted by some control analysts is to report samples as containing "not less than "a – 2u"" in situations where the statutory limit is a maximum permissible concentration. Enforcement action is only taken here when the analyst is sure that the specification has been exceeded. This is consistent with the requirement to prove ***beyond reasonable doubt*** that a limit has been exceeded if the case should come to court. This means that the effective enforcement level is, in some countries, not identical to the numerical value given in the EU legislation. The enforcement level thus equates with the maximum level and the expanded uncertainty.

Other control analysts may report and use the value "a" without taking into account any measurement uncertainty considerations.

[5] Official Journal of the European Communities, L 290, 24.11.1993, p. 14

4.3 Consequences of reporting results in different ways

There are potential problems with the reporting of results for which there is an EU specification.

This is best explained by means of an example:

Let us assume that there is an EU specification of 4 µg/kg for a substance being analysed. It could be expected that the analytical recovery is 100% and that the measurement uncertainty for the analysis will be of the order ± 44% of the analytical result, i.e. the analyst would, for nominal concentrations of 2, 3, 6 and 10 µg/kg, determine the following concentrations, including their uncertainties:

I. 10.0 ± 4.4 µg/kg
II. 6.0 ± 2.6 µg/kg
III. 3.0 ± 1.3 µg/kg, and
IV. 2.0 ± 0.9 µg/kg

Situation I
Here the level reported is above the EU specification and the true value lies in the range 5.6 to 14.4 µg/kg. All countries will state that the material is non-compliant with the EU specification.

Situation II
Here the level reported is above the statutory limit but the true value lies in the range 3.4 to 8.6 µg/kg. The level and its uncertainty would be reported.

Here some countries would report the sample as containing not less than 3.4 µg/kg of the analyte and, because it is not beyond reasonable doubt that the limit has been exceeded, no action will be taken.

However, other countries may take action on the 6.0 µg/kg result without taking uncertainty into account. For these countries, the material will be deemed to be non-compliant.

Nevertheless, for food hygiene legislation, a stricter view of risk may be taken, especially with respect to pathogens. Here there is more likelihood that some countries will deem the material to be non-compliant. There may therefore be a difference in interpretation when chemical and microbiological risks are being considered.

Situation III
Here the level reported is below the EU specification. Normally, countries would take the same view and accept the substance concerned. However, it should be noted that the reported value is "approaching the legal limit" and future samples will therefore be looked at more closely.

The situation described above for food hygiene also applies in this situation – i.e. when pathogens are being considered, some countries will take the view that even if there is the possibility, however remote, that the sample many contain such organisms, it should then be deemed to be non-compliant.

Situation IV
Here the level is below the EU specification with or without the uncertainty being taken into account. All countries would accept the material.

Conclusion

In situation II, there is the possibility that different countries will take make different decisions as to whether the material complies with an EU chemical specification.

In both situations II and III, there is the possibility that different countries will take make different decisions as to whether the material complies with an EU microbiological specification.

The concept is also shown in diagram form in Annex I with regard to quantification of chemicals.

4.4 Action

The various possible actions described make a crucial difference to the “enforcement” of EU provisions. Because the effect is so marked, Member States should be aware that there is the possibility of various countries “interpreting” the EU compliance with any standard in different ways. It is therefore recommended, when the setting of maximum levels (EU specification) is discussed, that this be done with full knowledge of those factors which affect the interpretation of the EU specification.

It is recommended that the measurement uncertainty be used when assessing compliance with a specification.

The situations described above (4.3) apply to maximum limits. However, similar considerations also apply to minimum limits in legislation. Thus the enforcement level equates to:

- the maximum level together with the uncertainty if a maximum value is specified in legislation; and
- the minimum level less the uncertainty if a minimum value is specified in legislation.

In practice, when considering a maximum value in legislation, the analyst will determine the analytical level and estimate the measurement uncertainty at that level. The value obtained by subtracting the uncertainty from the reported concentration, is used to assess compliance. Only if that value is greater than the maximum level in the legislation is it certain “beyond reasonable doubt” that the sample concentration of the analyte is greater than that required by the legislation.

4.5 Procedures for estimating measurement uncertainty

There are many procedures available for estimating the measurement uncertainty of a result. Some of the more common procedures are outlined in Annex II.

4.6 Value of the measurement uncertainty

There is concern that some laboratories under-estimate the size of their uncertainties and report unrealistically small uncertainties to their customers.

For chemical analyses, using the results from collaborative trials, it would not be unreasonable to expect that the (expanded) uncertainties reported by laboratories would be of the following orders[6]:

Concentration	Expanded uncertainty	Range of acceptable concentrations*
100g/100g	4%	96 to 104g/100g
10g/100g	5%	9.5 to 10.5g/100g
1g/100g	8%	0.92 to 1.08g/100g
1g/kg	11%	0.89 to 1.11g/kg
100mg/kg	16%	84 to 116mg/kg
10mg/kg	22%	7.8 to 12.2mg/kg
1mg/kg	32%	0.68 to 1.32mg/kg
< 100μg/kg	44%	56 to 144μg/kg

* this effectively means that values falling within these ranges may be regarded as being of the same analytical population.

[6] Derived from taking the Horwitz or modified Horwitz predicted σ_R values and doubling to obtain the equivalent expanded uncertainty.

For microbiological analyses, where it is frequently stated that results within the range of +/- 0.5 log units are acceptable, the range of actual counts with which this equates is frequently much larger than customers of analytical data appreciate (or require).

This is shown in the table below:

Count (absolute values)	**Count (log $_{10}$)**	**Expanded uncertainty**	**Range of acceptable counts in absolute values***
10 000 000	7	+/- 0.5	3 162 000 to 31 620 000
1 000 000	6	+/- 0.5	316 200 to 3 162 000
100 000	5	+/- 0.5	31 620 to 316 200
10 000	4	+/- 0.5	3 162 to 31 620
1 000	3	+/- 0.5	316 to 3 162
100	2	+/- 0.5	32 to 316
10	1	+/- 0.5	3 to 32

* this effectively means that values falling within these ranges may be regarded as being of the same analytical population.

However, for microbiological quantifications, the expanded uncertainties quoted in the table above may well be exceeded, particularly if any confirmation procedures are required in the analysis. Here it is not uncommon for the expanded uncertainties to be +/- 1 $\log_{10}$ unit.

The working group which discusses specifications in legislation, as well as any associated method performance criteria, should also look at the maximum measurement uncertainty which the laboratory estimates may be accepted as being fit-for-purpose.

It is important that customers for analytical data realise that analytical data are not exact; the above values indicate the extent of the uncertainty that could be expected.

5. USE OF RECOVERY INFORMATION IN ANALYTICAL MEASUREMENT

"Recovery" is the amount of material extracted for analysis as a fraction of the amount present. In most analytical work, not all the material is recovered (i.e. the "recovery" is less than 100%). For this reason, some analysts recommend correcting the result to compensate for low recovery.

The use of recovery information in analytical measurement is a difficult and contentious issue. The significance is best explained by means of an example.

In the field of mycotoxins, there is a limit of 4 µg/kg for total aflatoxin in nuts. Here the following situation may arise:

Country A will analyse a consignment and obtain a result of 3.5 µg/kg total aflatoxin using a method which, in the analytical run, has a recovery of 70%. Country A does not adjust for recovery corrections as a matter of policy and so the reported result will be 3.5µg/kg, which means that the sample will be deemed to comply with the 4 µg/kg limit.

Country B, however, uses recovery corrections as a matter of policy. That country could analyse the "same" sample using the "same" methodology and obtain the "same" analytical result, but will report not 3.5 but 5 µg/kg on a recovered basis. Here there is the possibility that because the 5 µg/kg level is greater than the EU limit of 4 µg/kg limit for total aflatoxin, the country concerned may deem the sample not to be in compliance with the EU limit.

Many of these issues are also addressed in the "Harmonised Guidelines for the Use of Recovery Information in Analytical Measurement", published by IUPAC[7]. This document provides information on how the recovery rates should be derived and encourages the use of certified reference materials where these are available. Procedures for assessing recovery are outlined in Annex III.

The EU has discussed the harmonisation of reporting of test results corrected for recovery factors in various working groups. As early as in 1998, it was agreed that, as regards the control of aflatoxins in foodstuffs, the recovery should be estimated and the results reported with or without adjustments; nevertheless, analysts should indicate the reporting method and the level of recovery[8]. More recently, in connection with the control of the levels of patulin and aflatoxin in foodstuffs, it has been specified in legislation that the analytical result adjusted for recovery should be used for checking compliance[9].

---

[7] "Harmonised Guidelines For The Use Of Recovery Information In Analytical Measurement", Michael Thompson, Stephen L R Ellison, Ales Fajgelj, Paul Willetts and Roger Wood, Pure Appl. Chem., 1999, 71, 337 – 348

[8] Commission Directive 98/53/EC of 16 July 1998 laying down the sampling methods and the methods of analysis for the official control of the levels of contaminants in foodstuffs – Annex II point 4.4. Recovery calculation, OJ L 201, 17.7.1998, p. 93.

[9] - Commission Directive 2003/78/EC of 11 August 2003 laying down the sampling methods and the methods of analysis for the official control of the levels of patulin in foodstuffs – Annex I point 5 Compliance of the lot of the sublot with the specification, *OJ L* 203, 12.8.2003, p. 40.

- Commission Directive 2003/121/EC of 15 December 2003 amending Directive 98/53/EC laying down the sampling methods and the methods of analysis for the official control of the levels for certain contaminants in foodstuffs *OJ L* 332, 19.12.2003, p. 39.

It is recommended that:

- each working group should address the issue on a case-by-case basis; in the case of contaminants in food and undesirable substances in feed, the provision that the analytical result has to be corrected for recovery in order to determine compliance should be considered for inclusion in the specific legislation;
- laboratories involved in official controls should determine their own recovery rates and ensure that these then meet the requirements for recovery where these are stipulated in the legislative criteria for acceptable methods of analysis.

6. OTHER LEGISLATION

There are a number of Commission directives, decisions and regulations which make reference to analytical tolerance, compliance, interpretation of results, etc. Examples of these are in the food contact materials legislation, legislation on the circulation of compound feedingstuffs, legislation on residues in live animals and animal products. None of these make specific reference to the now internationally accepted terminology of measurement uncertainty. In some of these fields, it should be appreciated that the approaches developed are unique to the sectors concerned.

7. RECOMMENDATIONS

- In developing food and feed legislation, the number of significant figures in their specifications under consideration should be laid down. In cases where legislation already provides clear guidance on the number of significant figures to be specified, then the analyst should comply with this requirement. In the other cases in general, and certainly in cases where this is appropriate for the accuracy of the result, the analyst should report to one significant figure more than is laid down in the specification, or rounded to one more significant figure.

- Enforcement authorities should use the measurement uncertainty associated with an analytical result when deciding whether or not a result falls within the specification for food and feed control purposes. The way that measurement uncertainty is to be used by enforcement authorities must be taken into account when analytical specifications are discussed. In practice, the analyst will determine the analytical level and estimate the measurement uncertainty at that level. The value obtained by subtracting the uncertainty from the reported concentration is used to assess compliance. Only if that value is greater than the maximum level in legislation, it is sure “beyond reasonable doubt” that the sample concentration of the analyte is greater than that prescribed by legislation.

- The working group which discusses specifications in legislation and any related performance criteria should also discuss the maximum measurement uncertainty which may be accepted as being fit-for-purpose.

- Finally, in developing food and feed legislation, it should be specified whether analytical results are to be reported on a recovery- or non-recovery-corrected basis case by case, whether recovery should also be indicated, and whether any minimum and/or maximum recovery is deemed acceptable.

Although each of the above recommendations involves a number of scientific considerations, it is of prime importance that all Member States adopt the same approach so that EU legislation can be applied consistently.

8. LOOKING AHEAD

The approach has already been specified in a number of EU Regulations[10] and may be expected to be adopted in others. However, it is recognised that this is a developing area, and consequently this report will have to be regularly updated in the light of developments within the scientific and enforcement communities.

It is acknowledged that other important issues having an impact on compliance also need to be considered on a regular basis in the light of current discussions at international level.

---

[10] See footnote 9. Similar provisions also under discussion for the control of tin, dioxins and ochratoxin A in certain foodstuffs

ANNEX I: DIAGRAMMATIC ILLUSTRATION OF THE EFFECT OF MEASUREMENT UNCERTAINTY AND THE LIMIT

The diagram below illustrates four different situations:

Situation I

The analytical result together with the measurement uncertainty exceeds the maximum level. All enforcement authorities would consider the sample to be non-compliant with the specification.

Situation II

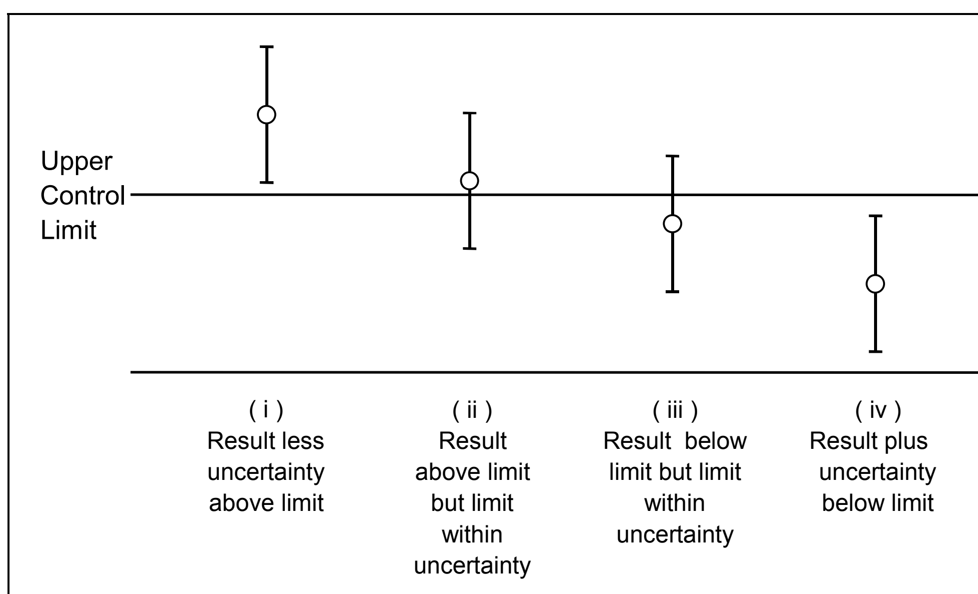
The analytical result exceeds the maximum level by less than the measurement uncertainty. Some enforcement authorities would accept the sample as being compliant with the specification if they take account of the measurement uncertainty. Others would ignore the measurement uncertainty and refuse to accept the sample.

Situation III

The analytical result is below the maximum level by less than the measurement uncertainty. In general, enforcement authorities would consider the sample to be compliant with the specification, but would probably be wary of future samples.

Situation IV

The analytical result is below the maximum value by an amount greater than the measurement uncertainty. All enforcement authorities would consider the sample to be compliant without any hesitation.



ANNEX II: PROCEDURES FOR THE ESTIMATION OF MEASUREMENT UNCERTAINTY

INTRODUCTION

It is important to recognise that the measurement uncertainty reported by a control laboratory is associated with an analytical result.

Within the EU, there are now a number of formal quality assurance measures which have to be implemented by food and feed[11] control laboratories. In particular, such laboratories have to be:

- accredited to an internationally recognised standard; such accreditation is aided by the use of internal quality control procedures;
- participate in proficiency schemes; and
- use validated methods.

It is essential that the information provided as a result of these requirements being applied is used by laboratories when estimating their measurement uncertainty in order to avoid their carrying out unnecessary work. In the food and feed sector, where great emphasis is being placed on the use of "fully validated" methods of analysis (i.e. methods which have been validated through collaborative trials), information obtained from such trials can be used in many scenarios.

In addition, information derived from internal quality-control procedures may also be used to estimate uncertainties in some scenarios.

For the analyst, it is important that no unnecessary duplication of existing work takes place.

[11] There is a current proposal to replace the current Council Directive 93/99/EEC of 29 October 1993 on the subject of additional measures concerning the official control of foodstuffs by a regulation on the official control of feed and food. The requirements for analytical laboratories are maintained, but updated as necessary. However it is proposed that they now apply to feed as well as food control laboratories.

RECOMMENDED PROCEDURES FOR THE ESTIMATION OF MEASUREMENT UNCERTAINTY

It is recommended that food and feed control laboratories use information derived from the following procedures to help in estimating the measurement uncertainty of their results:

Annex II.1 ISO guide to the expression of measurement uncertainty

Annex II.2 EURACHEM Guide to quantifying uncertainty in analytical measurement:

A. component-by-component approach

B. use of collaborative trial data

Annex II.3 Use of collaborative trial: data – ISO 5725 critical differences

Annex II.4 Draft ISO TS 21748 – Guide to the Use of Repeatability, Reproducibility and Trueness Estimates in Measurement Uncertainty Estimation

Annex II.5 Concept established by Commission Decision 2002/657/EC implementing Council Directive 96/23/EC concerning the performance of analytical methods and the interpretation of results

Annex II.6 AOAC INTERNATIONAL approach

Annex II.7 Internal quality control approach

Annex II.8 NMKL (Nordic Committee on Food Analysis) approach

Annex II.9 Microbiological Analyses

Annex II.10 Useful references

This information is outlined in the following sections of the Annex.

There is no “hierarchy” of procedures given to the sections. They are considered to be equally valid. However, the procedure that an individual laboratory uses will be considered appropriate by its accreditation agency as part of its 17025 accreditation.

It is recognised that further procedures for the estimation of measurement uncertainty are being developed and that, in this evolving situation, further recommendations will be made regarding acceptable procedures. It is expected that procedures will be developed based on results obtained from participation in proficiency-testing schemes, for example.

ANNEX II.1: ISO GUIDE TO THE EXPRESSION OF MEASUREMENT UNCERTAINTY

In 1993, ISO published the “Guide to the Expression of Uncertainty in Measurement”[12] in collaboration with BIPM, IEC, IFCC, IUPAC and OIML. The Guide lays down general rules for the expression and evaluation of measurement uncertainty across a wide range of chemical measurements. Also included in the Guide are examples of how the concepts described can be applied in practice. The Guide also introduces the idea of uncertainty and distinguishes it from "error", followed by a description of the steps involved in the evaluation of uncertainty.

The Guide may be applied to:

- quality control and quality assurance in manufacturing industries;
- testing for regulatory compliance;
- testing the use of agreed method;
- calibration of standards and equipment;
- development and certification of reference materials;
- research and development; and
- both empirical and rational methods.

The Guide places emphasis on the component-by-component approach, in which the method is dissected and incremental calculations of uncertainty are made and eventually added up to provide a combined uncertainty. There has been some criticism of the practicability of this approach. Much of the work to date regarding MU has been theoretical in nature and the amount of supporting analytical data has been limited. This has caused concern to analytical chemists, especially in the food sector, where they are already required by legislation to have some estimate of the “variability” of their results, mainly as a result of being required to use methods which have been assessed in a collaborative trial.

The evaluation of the measurement uncertainty for a method requires the analyst to look closely at all the possible sources of uncertainty in the method concerned, which may involve a considerable amount of effort which should not, however, be disproportionate. In practice, an initial study will usually identify the major source of uncertainty associated with the method; this will be the dominating influence on the total uncertainty. It is thus possible to make a good estimate of the uncertainty for a method as a whole by concentrating on the major sources of uncertainty inherent in it.. Once the measurement uncertainty has been estimated for a certain method in a particular laboratory, this estimate can be applied to subsequent results, provided that they are carried out in the same laboratory using the same method and equipment – always assuming, of course, that the quality control data justifies this course of action.

[12] “Guide to the Expression of Uncertainty in Measurement”, ISO, Geneva, 1993.

ANNEX II.2: EURACHEM GUIDE TO QUANTIFYING UNCERTAINTY IN ANALYTICAL MEASUREMENT

BACKGROUND

EURACHEM recently issued the second edition of its guide to quantifying uncertainty in analytical measurement[13], which is available as a download from www.measurementuncertainty.org

The EURACHEM guide is a protocol which establishes general rules for the evaluation and expression of uncertainty in quantitative chemical analysis based on the approach laid down in the ISO Guide. It is applicable at all levels of accuracy and in all fields, including quality control in manufacturing, testing for regulatory compliance, calibration, certification of reference materials, and research and development.

The Guide assumes that the evaluation of uncertainty requires the analyst to look closely at all the possible sources of uncertainty. It recognises that, although a detailed study of this kind may require a considerable effort, it is essential that the effort expended should not be disproportionate. It suggests that in practice a preliminary study will quickly identify the most significant sources of uncertainty, and as the examples showed, the value obtained for the total uncertainty is almost entirely determined by the major contributory factors.. It recommends that a good estimate can be made by concentrating effort on the main factors and that, once evaluated for a given method applied in a particular laboratory, the uncertainty estimate obtained may be reliably applied to subsequent results obtained by the method in the same laboratory, provided that this is justified by the relevant quality control data. No further effort should be necessary unless the method itself or the equipment used is changed, in which case the estimate would be reviewed as part of the normal revalidation.

Chapters 1 and 2 of the Guide deal with the scope and the concept of uncertainty. Chapter 3, entitled Analytical Measurement and Uncertainty, covers the process of method validation and conduct of experimental studies to determine method performance and their relationship to uncertainty estimates. There is also a new section on traceability. The chapter on uncertainty estimation in the previous guide has been considerably expanded and split into four separate sections dealing with the four steps involved. Step 1 deals with the specification of the measurand; Step 2 with identifying the uncertainty sources; Step 3, which has been considerably expanded to cover the use of existing method validation data, deals with quantifying the uncertainty; and Step 4 covers the calculation of the combined uncertainty. The examples have been completely revised and new ones added. They are now all in a standard format, which follow the four steps described above. They all utilise the cause and effect diagram as an aid to identifying the sources of uncertainty and to ensuring that all significant sources are included in the evaluation. In addition, a website has been set up (www.measurementuncertainty.org) which contains an indexed HTML version of the Guide. This site hosts a discussion forum on the application of the guide, and has a section for the publication of additional examples.

---

[13] A Williams, S L R Ellison, M Roesslein (eds.), *Quantifying uncertainty in analytical measurement*, available as QUAM2000-p1.pdf., 2000, EURACHEM Secretariat, www.measurementuncertainty.org

Of particular interest to food and feed analysts are the changes since the first edition of the Guide dealing with the use of method performance data and in particular the use of method validation data, from both collaborative validation studies and from in-house studies. There are new sections dealing with the use of method performance data which show that in many cases such data provides all or nearly all, the information required to evaluate uncertainty. These new sections are of particular interest to food and feed analysts, who frequently use methods of analysis which are “fully validated” through collaborative trial. An important aspect is the use of cause and effect diagrams as an aid in both method validation and uncertainty evaluation. By using these diagrams, it is possible to determine whether there are any components of uncertainty that are not covered by the validation data. In most cases, a good validation study will provide all of the necessary data, and it is possible to justify the use of an appropriate statistic, such as S_R, to determine the uncertainty.

A. COMPONENT-BY-COMPONENT APPROACH

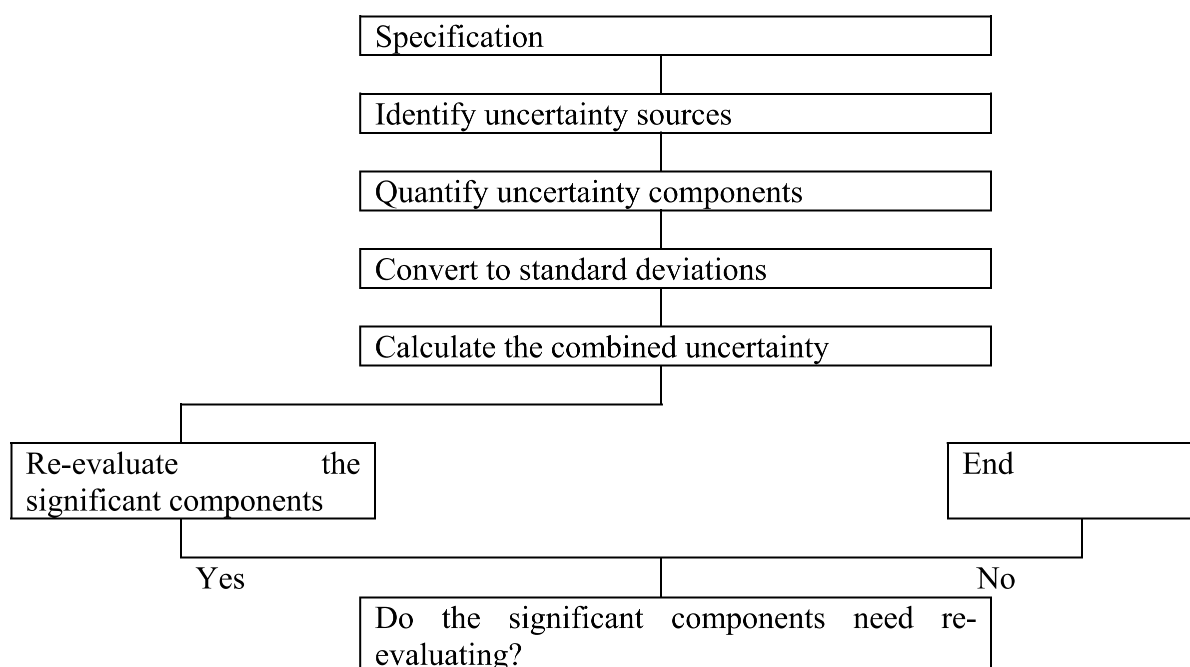
The EURACHEM guide to quantifying uncertainty in analytical measurement is a protocol which establishes general rules for the evaluation and expression of uncertainty in quantitative chemical analysis based on the approach laid down in the ISO guide. It is applicable at all levels of accuracy and in all fields, including quality control in manufacturing, testing for regulatory compliance, calibration, certification of reference materials, and research and development.

Uncertainty

The word "uncertainty", when used outside the world of science, indicates doubt. Thus uncertainty of measurement could be understood to mean that the analyst is unsure about the validity and exactness of his result. In the EURACHEM Guide, the definition of "uncertainty" is: “a parameter associated with the result of a measurement that characterises the dispersion of the values that could reasonably be attributed to the measurand.”

The Uncertainty Estimation Process

The estimation process is outlined in the EURACHEM Guide and involves the steps shown in the diagram below.



where:

Specification:	Write down a clear statement of what is being measured and the relationship between it and the parameters on which it depends.
Identify uncertainty sources:	List sources of uncertainty for each part of the process or each parameter. This is best achieved by breaking down a measurement process into a “cause-and-effect” diagram.
Quantify uncertainty components:	Estimate the size of each uncertainty. At this stage, approximate values suffice; significant values can be refined in subsequent stages.
Convert to standard deviations:	Express each component as a standard deviation.
Calculate the combined uncertainty:	Combine the uncertainty components, either using a spreadsheet method or algebraically. Identify significant components.

The final stage is to calculate the expanded uncertainty. This is achieved by multiplying the combined standard uncertainty by a coverage factor k. The coverage factor is chosen after considering a number of issues, such as the level of confidence required and any knowledge of underlying distributions. For most purposes a coverage factor of 2 is chosen which gives a level of confidence of approximately 95%.

Reporting uncertainty

The information required when reporting the result of a measurement ultimately depends on the intended use, but should contain enough information for the result to be re-evaluated if new data becomes available. A complete report should include a description of the methods used to calculate the result and its uncertainty, the values and sources of all corrections and constants used in the result calculations and uncertainty analysis, and a list of all the components of uncertainty with full documentation on how each was evaluated. The data and analysis should be presented in such a way that it can be easily followed and, if necessary, repeated. Unless otherwise required, the result should be reported together with the expanded uncertainty U.

B. USE OF COLLABORATIVE TRIAL DATA

Section 7.6.1 of the Second Edition of the EURACHEM Guide explicitly states:

"A collaborative study carried out to validate a published method, for example according to the AOAC/IUPAC protocol or ISO 5725 Standard, is a valuable source of data to support an uncertainty estimate. The data typically include estimates of reproducibility standard deviation, s_R, for several levels of response, a linear estimate of the dependence of s_R on level of response, and may include an estimate of bias based on CRM studies. How this data can be utilised depends on the factors taken into account when the study was carried out. During the 'reconciliation' stage indicated above, it is necessary to identify any sources of uncertainty that are not covered by the collaborative study data. The sources which may need particular consideration are:

- Sampling: Collaborative studies rarely include a sampling step. If the method used in-house involves sub-sampling, or the measurand (see Specification) is estimating a bulk property from a small sample, then the effects of sampling should be investigated and their effects included.
- Pre-treatment: In most studies, samples are homogenised, and may additionally be stabilised, before distribution. It may be necessary to investigate and add the effects of the particular pre-treatment procedures applied in-house.
- Method bias: Method bias is often examined prior to or during interlaboratory study, where possible by comparison with reference methods or materials. Where the bias itself, the uncertainty in the reference values used, and the precision associated with the bias check, are all small compared to s_R, no additional allowance need be made for bias uncertainty. Otherwise, it will be necessary to make additional allowances.
- Variation in conditions: Laboratories participating in a study may tend towards the means of allowed ranges of experimental conditions, resulting in an underestimate of the range of results possible within the method definition. Where such effects have been investigated and shown to be insignificant across their full permitted range, however, no further allowance is required.
- Changes in sample matrix: The uncertainty arising from matrix compositions or levels of interferents outside the range covered by the study will need to be considered.

Each significant source of uncertainty not covered by the collaborative study data should be evaluated in the form of a standard uncertainty and combined with the reproducibility standard deviation s_R in the usual way.

For methods operating within their defined scope, when the reconciliation stage shows that all the identified sources have been included in the validation study or when the contributions from any remaining sources have been shown to be negligible, then the reproducibility standard deviation S_R, adjusted for concentration if necessary, may be used as the combined standard uncertainty."

ANNEX II.3: USE OF COLLABORATIVE TRIAL: DATA – ISO 5725 CRITICAL DIFFERENCES

Note: this procedure is abstracted from the ISO 5725 Standard[14]. It was originally developed in 1981, i.e. before the concept of measurement uncertainty became formally recognised. It presumes that laboratories are operating at the same level as those which participated in the original collaborative trial to validate the method.

Contractor laboratories frequently have available an appropriate method of analysis which has been fully validated through a collaborative trial. The collaborative trial will provide information on the analytical performance of the method, particularly the precision as expressed as the repeatability (within the laboratory) and reproducibility (within and between laboratories) characteristics of the method. These values can be used to obtain a measurement uncertainty though the estimation of what is known as the "critical differences".

The arithmetic mean of the two single analyses obtained under repeatability conditions is compared to the (legislative or contractual) limit after calculation of the critical difference, as calculated below for the analytical result.

The critical difference for the analytical result is calculated using the formula given below:

$$CrD_{95}(|\overline{Y} - \overline{m}_o|) = \frac{0.84}{\sqrt{2}} \sqrt{R^2 - r^2 \frac{n-1}{n}}$$

where:

CrD_{95}	is the critical difference at the 95% probability value;
$\overline{Y}$	is the arithmetic mean of the results obtained;
m_o	is the (statutory/contractual etc.) limit;
n	is the number of analyses per sample;
R	is the reproducibility of the method at the relevant concentration; and
r	is the repeatability of the method at the relevant concentration.

If the difference between the (arithmetic mean) analytical result and the limit value is greater than the critical difference as calculated above, then it may be assumed that the sample which has been analysed does not fulfil the statutory or contractual requirements.

The values of r and R may have to be determined by interpolation so as to obtain the values which would apply at the limit concentration/value.

If it is to be expected that most samples comply with the statutory or contractual limit, then the final analytical results may be expected to be less than $[m_o + CrD_{95}(|\overline{Y} - m_o|)]$ if the limit is a maximum; or greater than $[m_o + CrD_{95}(|\overline{Y} - m_o|)]$ if the limit is a minimum and m_o is the given limit value.

This procedure has been employed in the EU's milk market support regime.

[14] "Precision of Test Methods", Geneva, 1994, ISO 5725, Previous editions were issued in 1981 and 1986.

ANNEX II.4: DRAFT ISO TS 21748 - GUIDE TO THE USE OF REPEATABILITY, REPRODUCIBILITY AND TRUENESS ESTIMATES IN MEASUREMENT UNCERTAINTY ESTIMATION

INTRODUCTION

The introduction to the ISO draft Guide[15] is reproduced below and demonstrates its scope:

"Knowledge of the uncertainty of measurement results is essential to the interpretation of the results. Without quantitative assessments of uncertainty, it is impossible to decide whether observed differences between results reflect more than experimental variability, whether test items comply with specifications, or whether laws based on limits have been broken. Without information on uncertainty, there is a real risk of either over- or under-interpretation of results. Incorrect decisions taken on such a basis may result in unnecessary expenditure in industry, incorrect prosecution in law, or adverse health or social consequences.

Laboratories operating under ISO 17025 accreditation and related systems are accordingly required to evaluate measurement uncertainty for measurement and test results and report the uncertainty where relevant. The *Guide to the expression of uncertainty in measurement* (GUM), published by ISO, is a widely adopted standard approach, but applies poorly in the absence of a comprehensive model of the measurement process. A very wide range of standard test methods are, however, subjected to collaborative study according to Part 2 of ISO 5725:1994. The present technical specification, TS 21748, provides an appropriate and economic methodology for estimating uncertainty for the results of these methods which complies fully with the relevant BIPM principles whilst taking advantage of method performance data obtained by collaborative study.

The general approach used in this technical specification requires that:

Estimates of the repeatability, reproducibility and trueness of the method in use, obtained by collaborative study as described in Part 2 of ISO 5725:1994, are available from published information about the test method in use. These provide estimates of within- and between-laboratory components of variance, together with an estimate of uncertainty associated with the trueness of the method.

The laboratory confirms that its implementation of the test method is consistent with the established performance of the test method, by checking its own bias and precision. This confirms that the published data are applicable to the results obtained by the laboratory.

Any influences on the measurement results which were not adequately covered by the collaborative study are identified and the variance in results that could arise from these effects is quantified.

An uncertainty estimate is made by combining the relevant variance estimates in the manner prescribed by the GUM.

The dispersion of results obtained in a collaborative exercise may also usefully be compared with measurement uncertainty estimates obtained via GUM procedures as a test of full understanding of the method. Such comparisons will be more effective given a consistent methodology for estimating the same parameter using collaborative study data."

---

[15] Draft ISO TS 21748 – "Guide to the Use of Repeatability, Reproducibility and Trueness Estimates in Measurement Uncertainty Estimation", Geneva, 2003

SCOPE OF TS 21748

The Guide gives guidance on:

- evaluation of measurement uncertainties using data obtained from studies conducted in accordance with ISO 5725-2:1994;

- comparison of collaborative trial results with measurement uncertainty (MU) obtained using formal principles of uncertainty propagation.

It is recognised that ISO 5725-3:1994 provides additional models for studies of intermediate precision. While the same general approach may be applied to the use of such extended models, uncertainty evaluation using these models is not incorporated into the present document.

The Guide does not describe the application of repeatability data in the absence of reproducibility data.

The Guide is applicable in all measurement and test fields where an uncertainty associated with a result has to be determined.

The Guide assumes that recognised, non-negligible systematic effects are corrected, either by applying a numerical correction as part of the method of measurement, or by investigation and removal of the cause of an effect.

The recommendations in the Guide are primarily for guidance. It is acknowledged that although they constitute a valid approach to the evaluation of uncertainty for many purposes, other suitable approaches may also be adopted.

In general, references to measurement results, methods and processes in this document should be understood to also apply to test results, methods and processes.

ANNEX II.5: CONCEPT SET BY COMMISSION DECISION 2002/657/EC IMPLEMENTING COUNCIL DIRECTIVE 96/23/EC CONCERNING THE PERFORMANCE OF ANALYTICAL METHODS AND THE INTERPRETATION OF RESULTS

INTRODUCTION

Council Directive 96/23/EC of 29 April 1996 on measures to monitor certain substances and residues thereof in live animals and animal products[16] provides for measures to monitor substances and groups of residues listed in the Annex to the Directive. Provisions on the implementation of this Directive concerning the performance of analytical methods and the interpretation of results in this sector is given in Commission Decision 2002/657/EC of 12 August 2002 implementing Council Directive 96/23/EC concerning the performance of analytical methods and the interpretation of results[17]. This Decision stipulates procedures that may be followed in order to demonstrate that a specific method can be used to enforce the legislation. One of the basic concepts applied in this Directive refers to the calculation of "CCα" and "CCβ". These acronyms mean "critical concentration" characterised by a defined value for α and β using statistical terminology from hypotheses-testing. This concept, when applied to the characterisation of analytical methods, is explained below.

CONCEPT

A laboratory will determine the concentration of a specific contaminant in a sample. Based on the results, a decision is taken as to whether the measured value is:

1. above a limit indicating non-compliance or
2. below the limit.

Since the measured value has an analytical error, wrong decisions can be taken: a sample containing the analyte below the limit can wrongly be considered as non-compliant since the *measured* value is above the limit. In this Directive, the Greek letter α is used to indicate the probability that this error will occur. Similarly, a sample containing the (true concentration of the) analyte *above* the limit is wrongly classified as compliant since the *measured* value is below the limit. This kind of error also has a certain probability which is expressed as the Greek letter β. The values CCα and CCβ indicate the concentrations at which such errors will occur with a defined probability. Based on the known values of CCα and CCβ, laboratories may evaluate the significance of their results. Thus, in the areas of analysis to which Commission Decision 2002/657/EC applies, the values are considered important performance characteristics that need to be experimentally evaluated.

Commission Decision 2002/657/EC describes various cases and different ways of calculating CCα and CCβ. The Annex describes the situation in which CCα and CCβ are calculated for a method that determines a substance for which a permitted level exists.. The situation would be different when dealing with substances *without* permitted levels.

[16] OJ L 125, 23.5.1996, p. 10
[17] OJ L 221, 17.8.2002, p. 8

CALCULATION

Where a permitted level is prescribed, the measured concentration must be evaluated in order to establish whether the actual but unknown concentration of the analyte is above the permitted level or not. First consider the situation in which the sample contains the analyte with a concentration just at the level provided for by legislation. Because of the analytical error (measurement uncertainty), the probability of the measured value being below or above the limit would be 50% in each case. In consequence, all cases, in which the measured result is above the legislative level, would lead to *false positive* decisions when not taking into account the uncertainty of the result. In order to be certain that the *measured* result demonstrates that the true concentration is above the permitted or statutory limit, Commission Decision 2002/657/EC uses the CCα concept. This concentration, greater than the statutory limit or permitted level, is the lowest *measured* concentration at which it is certain, with a given probability, that the *true* concentration is above the permitted level. Thus CCα is a *decision limit* and the risk that the true value is below the permitted limit is characterised by α. A typical value for α is 5 % indicating that the probability of a false positive result is 5 %.

EVALUATION OF CCα

The CCα concentration of a method may be established by spiking 20 blank materials with the target analyte at the permitted level or statutory limit and calculating the mean value of the 20 analyses along with the standard deviation of the results. CCα is equal to the mean value *plus* 1.64 times the standard deviation when accepting the probability of an α–error of 5%.

The development of the CCα enables the method to not give false positive results by focusing on samples that contain the target analyte with a concentration *below* or at the permitted level.

EVALUATION OF CCβ

In contrast, when discussing CCβ, the reverse situation is considered, i.e. samples that (truly) contain the target analyte *above* the permitted level even though they analyse at below the permitted limit. Here it is important that the method clearly indicate non-compliance of the samples. However, the capacity of the method to prove non-compliance depends on the true concentration of the target analyte. Let us again consider the example used above. Now it is assumed that the true concentration is equal to CCα. A sample containing the analyte at this concentration should be considered as non-compliant but the probability that the measured concentration is below CCα is 50 % thereby leading to *false negative* results. It may be concluded that the method's suitability for detecting non-compliance is not sufficient when the true concentration is equal to the decision limit. Indeed, only if the true concentration of the sample is *above* CCα is the method capable of proving non-compliance, assuming a low rate of false negative results. In particular, it is of interest to establish the critical concentration at which the probability of *false negative results,* for instance, is below 5%, and this concentration is therefore called the detection capability, CCβ.

This value can be determined by spiking 20 blank materials with the target analyte at a concentration equal to CCα and calculating the mean value of the 20 analyses along with the standard deviation of the results. CCβ is equal to the mean value *plus* 1.64 times the standard deviation when accepting the probability of a β–error of 5%.

COMMENT

It is important to realise that CCα and CCβ are statistically derived measures aimed at limiting the risk that a compliant sample is wrongly classified as non-compliant and at indicating the concentration, above the permitted limit, at which the method can demonstrate non-compliance, assuming that the rate of false negatives is sufficiently low. Moreover, these limits should not be confused with other performance characteristics, such as limit of quantification. In fact, though CCα and CCβ are *above* the permitted level, Commission Decision 2002/657/EC also requires that the method show sufficient trueness and repeatability at concentrations *below* the permitted level.

For normal enforcement analyses, however, the CCβ concept is not used, and a concentration deemed to be below the statutory limit is accepted as such in all scenarios.

ANNEX II.6: AOAC INTERNATIONAL APPROACH

The following paper, which was recently published in the *Journal of AOAC INTERNATIONAL*[18] sets out the AOAC INTERNATIONAL view on measurement uncertainty. It is an attempt to explain the concept of "uncertainty" as it is being widely discussed and used in the analytical community.

"The idea is very simple – what variability can one expect from one's measurements. But the concept was introduced initially into the analytical laboratory from metrology, which required an examination of all possible sources of error, adding them vectorially, and expanding the resulting total error statistically to arrive at a result with an attached 95% probability statement. Analytical chemists, however, had long ago realized that by performing an interlaboratory study on a standard method using a group of typical laboratories analysing a set of typical matrixes, they could reproduce almost all the uncertainty that nature could create. This practical aspect is now being incorporated into the discussion of uncertainty.

The official definition of measurement uncertainty (from the NIST Web site http://physics.nist.gov/cuu/Uncertainty/glossary.html) is:

- *Uncertainty (of measurement)*: parameter, associated with the result of a measurement that characterizes the dispersion of the values that could reasonably be attributed to the measurand.

- The parameter may be, for example, a standard deviation (or a given multiple of it), or the half-width of an interval having a stated level of confidence.

- Uncertainty of measurement comprises, in general, many components. Some of these components may be evaluated from the statistical distribution of the results of a series of measurements and can be characterized by experimental standard deviations. The other components, which also can be characterized by standard deviations, are evaluated from assumed probability distributions based on experience or other information.

- It is understood that the result of the measurement is the best estimate of the value of the measurand, and that all components of uncertainty, including those arising from systematic effects, such as components associated with corrections and reference standards, contribute to the dispersion.

Considerable confusion about this term will be swept away immediately if you note that *the term "UNCERTAINTY" is attached to a RESULT, not to a method*; i.e., *measurement uncertainty* is being discussed, not *method uncertainty*. We will see how the method gets into the discussion later.

The introductory chapter to practically every textbook of quantitative analysis discusses the variability of analytical results and often advises reporting results in terms of the mean of a series of replicates and an interval within which you expect most (i.e. 95%) of your future results to fall if future analyses were conducted in an identical manner. However, the economics of chemical analysis dictates that only a few analyses are conducted on a test sample ("the results are usually good enough for government work"), so this theoretical admonition has been largely ignored until recently. Now, for accreditation purposes, laboratories are required to attach a statement of *measurement uncertainty* to their analytical results.

[18] W. Horwitz, 2003, The Certainty of Uncertainty Journal of AOAC INTERNATIONAL, **86**, 109-111

To obtain that halo of uncertainty surrounding your reported result, you have essentially four options:

(1) The option of calculating the equivalent of a confidence interval from the "*t*" factor applied to the standard deviation of replicates.

(2) The theoretical "bottom-up" approach recommended by the bible on uncertainty, rubber stamped by nine international organizations[19].

(3) The practical "top-down" approach from the relative standard deviation derived from an interlaboratory study by the Harmonized IUPAC/AOAC protocol[20] or ISO 5725[21].

(4) The estimate obtained by applying the Horwitz formula relating the relative standard deviation to concentration, as a mass fraction, $RSD_R = 2C^{(-0.15)}$, which is based upon a review of over 10 000 interlaboratory results, primarily published in the *Journal of AOAC INTERNATIONAL*.

[Alternative formula are: $\sigma_H = 0.02c^{0.8495}$ and $RSD_R = 2^{(1-0.5\log C)}$]

Option 1

Run sufficient replicates on the specific test sample under consideration to obtain a fairly good idea of how the results will scatter in routine work. If you manufacture a product to a specification of 20% fat day in and day out, with the help of a statistician, you would soon be able to know the typical uncertainty of the fat content of the product, of the sampling, and of the analysis. But if you are called upon to provide an *estimate of uncertainty* from a set of duplicates from a material you will never see again, you will have to multiply the standard deviation calculated from that pair of results by a factor of 12! Such an estimate is essentially useless because experience shows that future analysis from even a moderately experienced analyst will rarely approach the expected extreme.

Incidentally, running more replicates will not change the "true value" of the mean or of the standard deviation. More replicates provide more confidence in the interval estimate bracketing the true concentration and the true standard deviation.

---

[19] "Guide to the Expression of Uncertainty in Measurement", ISO, Geneva, 1993.

[20] W. Horwitz, 1995. "Protocol for the Design, Conduct and Interpretation of Method Performance Studies", *Pure Appl. Chem.,* 1995, **67**, 331-343

[21] "Precision of Test Methods", Geneva, 1994, ISO 5725, previous editions were published in 1981 and 1986.

Option 2

Sit down and think about everything that might possibly affect the result and estimate the expected variation that each factor will contribute to the final value. These will include uncertainties, expressed as standard deviations, from:

- standard weight corrections;
- buoyancy corrections (temperature, pressure);
- volumetric flask corrections (calibration, temperature);
- pipette volume corrections (calibration, temperature);
- reference material content uncertainty;
- concentration of calibrant uncertainty;
- signal measurement uncertainty;
- time measurement uncertainty;
- extraction variability (volume, temperature, and solubility effects);
- reaction or separation variability;
- effect of interferences which may or may not be present.

When you have thought of everything that might possibly influence your reaction, separation, and measurement, and assigned a standard deviation to each factor, calculate the square root of the linear combination of the variances to obtain the final standard deviation that you attach to your measurement as the measurement uncertainty. Then multiply this final standard deviation by a coverage factor (k) of 2 to ensure a probability of 95%, i.e., only a 5% chance that the true value lays outside the expanded uncertainty limits. Incidentally, do not forget lot and analytical sampling, which are unique for every lot and which, therefore, require individual estimation by replication of these components for completeness. "Practical" examples can be found in a EURACHEM guide[22].

This is known as the bottom-up approach. You can come back later and add in those factors that you initially overlooked or which are pointed out to you by your colleagues or by your friendly assessor months after the report has been delivered and forgotten.

This absurd and budget-busting approach (for analytical chemistry) arose from metrological chemists taking over in entirety the concepts developed by metrologists for physical processes measured with 5-9 significant figures (gravitational constant, speed of light, etc.) and applying them to analytical chemistry measurements with 2 or 3 significant figures. This approach also ignores the fact that some chemical methods are influenced by numerous factors, some positive and some negative, that tend to cancel out, and that often other chemical methods are influenced by a few factors that overwhelm the weight and volume uncertainty calculations presented in the published examples.

[22] A Williams, S L R Ellison, M Roesslein (eds.), *Quantifying uncertainty in analytical measurement*, available as QUAM2000-p1.pdf., 2000, EURACHEM Secretariat, www.measurementuncertainty.org

Option 3

The approach, which is becoming generally accepted in Europe, is to conduct an interlaboratory study utilizing the Harmonized IUPAC/AOAC or ISO 5725 protocol (which utilizes an identical statistical model except for outlier removal). The protocols require a sample of at least 8 typical laboratories analysing a minimum set of 5 matrices covering the range of materials of interest. Then relate the standard deviation among laboratories (S_R) as being proportional to measurement uncertainty. This is known as the top-down approach. By utilizing a sample of presumably typical laboratories operating in different environments on at least 5 materials covering the range of interest, it is very likely that most of the potential error factors that are likely to be encountered in practice will have been introduced. Therefore, if we equate this S_R to measurement uncertainty and call it standard measurement uncertainty (standard uncertainty for short), we are at least about 70% certain that our result plus and minus S_R will encompass the "true" value. If we multiply S_R by a coverage factor of 2, we obtain the "expanded measurement uncertainty" (expanded uncertainty for short); we are now at least 95% certain that our result plus and minus $2S_R$ will encompass the "true" value.

When using this collaborative study approach, which results in a "standard method" as used by ISO 17025, be sure that all of the important variables are specified or understood (*see Definition of Terms and Explanatory Notes* section of the *Official Methods of Analysis of AOAC INTERNATIONAL*) with assigned limits. Weights are assumed to be within ±10% (but use the actual weight for calculations), volumetric glassware are assumed to have their assigned volume with negligible uncertainty when used with instrumental methods (but not when used in titrations), graduates are assumed to deliver the volume read from their scale, temperatures are set to be within ±2°C, pHs are within ±0.05 unit, times are followed to within 5%, and instrument scales, dials, and markers are estimated to their finest degree, then Clause 5.4.6.2 Note 2 in ISO 17025 reading, "In those cases where a well-recognized test method specifies limits to the values of the major sources of uncertainty of measurement and specify the form of presentation of the calculated results, the laboratory is considered to have satisfied this clause by following the test method and reporting instructions." Under such conditions, S_R derived from the supporting collaborative study in the same units as the reported result with the accompanying number of significant figures, usually 2 or 3, may be used as the standard uncertainty, assuming the laboratory has demonstrated that it operates within the performance limits for that method.

Option 4 or 0

As a last resort, or even before you start any analyses, you can make a rough calculation to determine if the expected uncertainty at the expected concentration will be fit for the intended purpose. Apply the Horwitz formula (or a suitably adjusted version of the Horwitz formula to account for special circumstances such as a single laboratory) to the anticipated concentration to obtain a within-laboratory S_r and multiply it by 2 to obtain the expanded uncertainty. The Horwitz formula as initially applied to among-laboratory reproducibility parameters in %, and with C expressed as a mass fraction, is

$$RSD_R \text{ (in \%)} = 2C^{(-0.15)}$$

or as a standard deviation

$$S_R = 0.02C^{(0.85)}$$

To apply to within-laboratory repeatability parameters, divide by 2 and equate this to estimated standard uncertainty:

$$S_r = 0.01C^{(0.85)}$$

To obtain the expanded (repeatability) uncertainty, multiply by 2:

$$S_r = 0.02C^{(0.85)}$$

For example, if we are dealing with a pure compendial material, C expressed as a mass fraction is 1, so the anticipated expanded uncertainty, $2S_r$, is 0.04 or 4%. This is interpreted as 95% of anticipated results will fall between 96 and 104%. You can “improve” your uncertainty by running independent replicates. “Independent” means as a minimum "non-simultaneous" but again economics would not permit it, so the improvement would be considerably less than theoretical.

Summary: The Horwitz formula will tell you if your anticipated uncertainty is such that you will be within the limits of the ballpark with a typical method. The maximum spread obtained by the top-down approach will encompass the “true value” in almost all practical cases. It is usually easier to let nature slip in all the unanticipatable tricks that can befall even the most careful analysts than to valiantly attempt to foresee them beforehand by the budget approach. This is how the uncertainty of the method becomes entangled with the uncertainty of the measurement.

Note 1: Some of these “unanticipatable tricks” are chaotic, like dropping the thermometer or missing a decimal point. They are not subject to statistical description. Such adventitious flaws are handled by quality control but they cannot be predicted in any quantitative way. Such flaws are not intrinsic to the method.

Note 2: The uncertainty of a method, its bias and variability, is revealed by the spread of the individual measurements, i.e., by the average and standard deviation of the set of measurements. The theory envisions that an infinite set of concentration estimates is obtained for each true concentration but the hapless finite chemist is forced just to take a sampling from this infinite set at the given concentration, usually just one or two estimates. Outlier tests are applied to remove clearly extrinsic interferences with the proper application of the chemical method. Note also that the uncertainty components, both bias and variability, are functions of the true concentration, though variability is usually observed to be more concentration-dependent than the bias.

If a method is to be corrected for recovery (bias) the method will usually so indicate. Many regulatory methods do not require such a correction because the specification (tolerance) was established by the same method so the recovery is "built into" the specification.

Note 3: The analytical chemist usually ignores sampling uncertainty primarily because typically little or no information accompanies the laboratory sample as to whether or not the laboratory sample truly reflects the lot. It is usually left to "management" to coordinate the analytical information with the sampling information. However, if the sample has been collected according to statistical principles (a process that usually requires a very large number of increments) and if these increments have been analysed to provide the basis for an estimate of sampling uncertainty, then propagation of error considerations can provide an overall "sampling + analysis" uncertainty.

Note 4: We have deliberately omitted mentioning the problem of expressing measurement and method uncertainties of microbiological examinations where the target analyte is intentionally diluted to the point of producing "true" false positives and "true" false negatives for comparison of the results from a test method to those from a reference method."

ANNEX II.7: INTERNAL QUALITY CONTROL APPROACH

Accredited laboratories are required to have introduced acceptable internal quality control procedures. In the food sector, the use of the international harmonised guidelines has been recommended by the Codex Alimentarius Commission.

From the use of quality control procedures it is possible to devise, within the laboratory, estimates of repeatability and reproducibility by taking the standard deviation used in the Shewhart Charts set up upon the introduction of the internal quality control procedures. The value here can be multiplied by 1.6 to calculate the appropriate value of reproducibility and then used as the value of σ_R in the same way as described previously.

This procedure has been used within the Netherlands Food Inspection Service (*Keuringsdienst van Waren*).

ANNEX II.8: NMKL (NORDIC COMMITTEE ON FOOD ANALYSIS) APPROACH

The NMKL Procedure No. 5 (1997) – "Estimation and expression of measurement uncertainty in chemical analysis"[23] comprises guidelines for how to estimate and present the uncertainty of a given analytical result. The procedure recognises the existence of documents on measurement uncertainty based on the principle that the uncertainty is estimated for each step of a method, following which the uncertainties are combined in an error budget. However, at the time of drafting the procedure, it was decided to base it on less time-consuming models for the evaluation of measurement uncertainty, e.g. using internal reproducibility.

This existing NMKL procedure is currently under revision. The central point of the revised document will be moved towards an estimation of measurement uncertainty based on an error budget. While taking the GUM-document into account, the revised procedure will also be based on an important part of the new EURACHEM-document, which presents the possibility of intensive use of experimental data. This is in agreement with the new standard ISO/IEC 17025, which draws attention to the relevance of taking experience and validation data into account during estimation of measurement uncertainty. It is evident that use of experimental data can simplify the estimation of the total measurement uncertainty. However, it is still of importance to identify all sources which may contribute to the uncertainty of the method and result. A detailed evaluation of all steps of a method may provide the chemist with important information on where to find the major source of error and subsequently improve knowledge of how to minimise the total error of a method.

The aim of the revised procedure is to describe, simply and succinctly, how to perform a good estimate of measurement uncertainty using, among other things, data obtained through validation or other types of quality assurance checks.

---

[23] "Estimation and Expression of Measurement Uncertainty in Chemical Analysis", NMKL Secretariat, Finland, 1997, NMKL Procedure No. 5.

ANNEX II.9: MICROBIOLOGICAL ANALYSES

INTRODUCTION

ISO/TC 34/SC 9 "*Food products – Microbiology*" dealt with the topic of measurement uncertainty for microbiological analyses at its last meeting (Bangkok, 2-4 December 2002).

It agreed to define a general approach for quantitative determinations (counting and alternative quantitative methods). Despite the clear need to define an approach for qualitative determinations (presence/absence tests), the subject was regarded as not having been considered in sufficient detail to enable a harmonised approach to be adopted, and the issue would be considered again at the 2004 meeting.

The general approach agreed on at the meeting for quantitative determinations is to be described in an ISO Technical Specification entitled "*Microbiology of food and animal feeding stuffs – Guide to the expression of measurement uncertainty for quantitative determinations*". The first draft has been circulated by ISO for comment.

This approach prescribes one (or several) value(s) of measurement uncertainty per target micro-organism associated with a particular laboratory's results. It follows a global or "top-down" approach, based on the **standard-deviation of reproducibility on the final result of the analysis**. ISO/TC 34/SC 9 considered that the "step-by-step" approach does not apply satisfactorily to the case of the microbiological analysis of food, where it is difficult to build a genuinely comprehensive model of the measurement process, and where it is easy to forget a non-negligible source of uncertainty. Thus, there is a high risk of underestimation of MU. Moreover, a step-by-step approach would place a heavier burden on laboratories than the global approach selected.

Three options, in decreasing order of preference, are available to experimentally determine this standard deviation of reproducibility:

- 1st option: intralaboratory standard deviation of reproducibility, established by the laboratory itself;
- 2nd option: interlaboratory standard deviation of reproducibility, established through an interlaboratory trial for the validation of the method used;
- 3rd option: interlaboratory standard deviation of reproducibility, established through an interlaboratory proficiency trial in which the laboratory has taken part.

The ISO/TS would detail two experimental protocols for option 1, as well as the conditions for using options 2 and 3. In particular, the laboratory must demonstrate that its bias and precision are compatible with those derived from the interlaboratory trial, and the samples used for the trial should represent the samples used routinely by the laboratory in terms of matrices, strains of micro-organism, background flora and level of contamination.

INTERPRETATION OF RESULTS WITH REGARD TO MEASUREMENT UNCERTAINTY

In the existing Community legislation, the measurement uncertainties linked to microbiological criteria have been addressed in Council Directive 94/65/EEC on minced meat and meat preparations and in Commission Decision 93/51/EEC on cooked crustaceans and molluscan shellfish. In these provisions, the traditional '3x limit' approach has been used in order to take into account the effects of measurement uncertainties. As regards other microbiological criteria in the Community legislation in force, the measurement uncertainties have not been addressed.

The Health and Consumer Protection DG of the European Commission is currently in the process of revising the microbiological criteria in Community legislation and is preparing a draft Commission regulation on microbiological criteria as well as a discussion paper on the strategy to set microbiological criteria. These draft documents have been discussed with experts from the Member States on several occasions. One of the issues still under discussion is whether the measurement uncertainty should be taken into account for interpreting the compliance of test results with statutory limits (quantitative limits in microbiological criteria) and, if so, in what way.

SPECIFIC REFERENCES – MICROBIOLOGICAL SECTOR

AAG (Accreditation Advisory Group of the Institute of Food Science Technology) 2000 Guideline no. 13. "Uncertainty of Measurement in Food Microbiology by Analysis of Variance."

Andrews, W.H. 1996 Validation of modern methods in food microbiology by AOAC International, collaborative study. Food Control, **7**: 19-29

Andrews, W.H. 1997 New trends in food microbiology: An AOAC international perspective. Journal of AOAC International, **80**: 908

Fuentes-Arderiu, X. Uncertainty of measurement in clinical microbiology. Journal of the International Federation of Clinical Chemistry and Laboratory Medicine, **13** (4): http://www.ifcc.org/ejifcc/vol113no4/130401006.htm

IAHA 2001 Guidelines for Uncertainty Estimation, Revision 0.

In't Veld, P.H. 1998 The use of reference materials in quality assurance programmes in food microbiology laboratories. International Journal of Food Microbiology, **45**: 35-41

ISO 2001 Milk and milk products – Quality control in the microbiological laboratory – Part 1: Analyst performance assessment for colony counts ISO/CD 14461-1| IDF 169-1

ISO 2001 Milk and milk products – Quality control in the microbiological laboratory – Part 2: Determination of the reliability of colony counts of parallel plates and subsequent dilution steps ISO/CD 14461-2|IDF 169-2

ISO 2002 Statistics of analytical data. Protocol for the establishment of precision characteristics of microbiological quantitative methods by interlaboratory studies, 1st draft, ISO/TC 34/SC 9 N 543

ISO 2002 Microbiology – Determination of measurement uncertainty, proposal, ISO/TC 34/SC 9 N 558

Lightfoot, N. F. & Maier, E. A. (eds) 1998 Microbiological Analysis of Food and Water: Guidelines for Quality Assurance. Elsevier, Amsterdam

Miemi, R. M. & Niemelä, S. I. 2001 Measurement uncertainty in microbiological cultivation methods. Accreditation and quality assurance, 6: 372-375

Niemelä, S.I, 2002 Uncertainty of quantitative determinations derived by cultivation of micro organisms. 2nd edition, Centre for Metrology and Accreditation, Advisory Commission for Metrology, Chemistry Section, Expert Group for Microbiology, Helsinki, Finland, Publication J3/2002.

NMKL (Nordic Committee on Food Analysis) 1989 Report no. 9 Handbook for Microbiological Laboratories. Introduction to Internal Quality Control of Analytical Work.

NMKL 1994 Report no. 5, 2nd edition Quality assurance guidelines – for microbiological laboratories.

NMKL Procedure No 8 (1999) Measurement of Uncertainty in Microbiological Examination of Foods

Peterz, M. 1992 Laboratory performance in a food microbiology proficiency testing scheme. Journal of Applied Bacteriology, **73**: 210-216.

ANNEX II.10. USEFUL GENERAL REFERENCES FOR MEASUREMENT UNCERTAINTY

Analytical Methods Committee of the Royal Society of Chemistry “Uncertainty of Measurement - Implications of its Use in Analytical Science”, Analyst, 1995, **120 (9)**, 2303-2308.

Dveyrin, Z., Ben-David, H., Mates, A. 2001 Proficiency testing as tool for ISO 17025 implementation in National Public Health Laboratory: a mean for improving efficiency. Accreditation & Quality Assurance, **6**: 190-194

EURACHEM Guidance Document No. 1/WELAC Guidance No. WGD 2: “Accreditation for Chemical Laboratories: Guidance on the Interpretation of the EN 45000 series of Standards and ISO/IEC Guide 25”

ISO (2nd ed., 1993) VIM “International Vocabulary of Basic and General Terms in Metrology”. Geneva.

NMKL Procedure no. 3 (1996) Control charts and control samples in the internal quality control in chemical food laboratories

NIST Technical note 1297 (1994 Edition): “Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results”

Örnemark, U., Boley, N., Saeed, K., van Berkel, P.M., Schmidt, R., Noble, M., Mäkinen, I., Keinänen, M., Uldall, A., Steensland, H., Van der Veen, A., Tholen, D. W., Golze, M., Christensen, J.M., De Bièvre, P., De Leer, W. B (ed). 2001 Proficiency testing in analytical chemistry, microbiology, and laboratory medicine – working group discussions on current status, problems, and future directions. Accreditation & Quality Assurance, **6**: 140-146.

UKAS (United Kingdom Accreditation Service) 2000 The Expression of Uncertainty in Testing Edition 1, UKAS Publication ref: LAB 12

ANNEX III: PROCEDURES FOR ASSESSING RECOVERY[24]

INTRODUCTION

The estimation and use of recovery is an area where practice differs among analytical chemists. The variations in practice are most obvious in the determination of analytes such as veterinary drug residues and pesticide residues in complex matrices, such as foodstuffs and in environmental analysis. Typically, such methods of analysis rely on transferring the analyte from the complex matrix into a much simpler solution that is used to present the analyte for instrumental determination. However, the transfer procedure results in loss of analyte. 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 gives a value lower than the true concentration in the original test material. If no compensation for these losses is made, significantly discrepant results may be obtained by different laboratories. Even greater discrepancies arise if some laboratories compensate for losses and others do not.

Recovery studies are clearly an essential component of the validation and use of all analytical methods. It is important that all concerned with the production and interpretation of analytical results are aware of the problems and the basis on which the result is being reported. At present, however, there is no single well-defined approach to estimating, expressing and applying recovery information. The most important inconsistency in analytical practice concerns the correction of a raw measurement, which can (in principle) eliminate the low bias due to loss of analyte. The difficulties involved in reliably estimating the correction factor deter practitioners in some sectors of analysis from applying such corrections.

In the absence of consistent strategies for the estimation and use of recovery information, it is difficult to make valid comparisons between results produced in different laboratories or to verify the suitability of those data for the intended purpose. This lack of transparency can have important consequences in the interpretation of data. For example in the context of enforcement analysis, the difference between applying or not applying a correction factor to analytical data can mean, respectively, that a legislative limit is exceeded or that a result is in compliance with the limit. Thus, where an estimate of the *true concentration* is required, there is a compelling case for compensation for losses in the calculation of reported analytical result.

---

[24] Extract from « Harmonised guidelines for the use of recovery information in analytical measurement (Technical report)» by *M. Thompson, S.L.R. Ellison, A. Fajgelj, P. Willetts and R. Wood*.

Resulting from the Symposium on Harmonisation of Quality Assurance
Systems for Analytical Laboratories, Orlando, USA, 4–5 September 1996
held under the sponsorship of IUPAC, ISO and AOAC INTERNATIONAL. Published in *Pure and Applied Chemistry,* Volume 71, No 2, pp. 337-348, 1999.

Codex Alimentarius Commission adopted at its 24th session held in Geneva from 2 to 7 July 2001, the adopted the IUPAC Guidelines for the use of recovery information in Analytical Measurement by reference for the purposes of Codex (ALINORM 01/41, § 196.

DEFINITIONS AND TERMINOLOGY USED IN THE GUIDELINES

General analytical terminology is assumed to be accepted when these Guidelines are read, but specific definitions of the terms most pertinent to the Guidelines are given below:

Recovery: Proportion of the amount of analyte, present in or added to the analytical portion of the test material, which is extracted and presented for measurement.

Surrogate: Pure compound or element added to the test material, the chemical and physical behaviour of which is taken to be representative of the native analyte.

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'.)

Native analyte: Analyte incorporated into the test material by natural processes and manufacturing procedures (sometimes called 'incurred analyte'). Native analyte includes 'incurred analyte' and 'incurred residue' as recognised in some sectors of the Analytical Community. It is so defined to distinguish it from analyte added during the analytical procedure.

Empirical method of analysis: A method that determines a value which can be arrived at only in terms of the method *per se* and serves by definition as the only method for establishing the measurand. (Sometimes called 'defining method of analysis'.)

Rational method of analysis: A method that determines an identifiable chemical(s) or analytes(s) for which there may be several equivalent methods of analysis available.

PROCEDURES FOR ASSESSING RECOVERY

<u>Recovery information from matrix reference materials</u>

In principle, recoveries could be estimated by the analysis of matrix reference materials. The recovery is the ratio of the concentration of analyte found to that stated to be present. Results obtained on test materials of the same matrix could, in principle, be corrected for recovery on the basis of the recovery found for the reference material. However, several problems potentially beset this use of the reference materials, namely: (a) the validity of any such recovery estimate depends on the premise that the analytical method is otherwise unbiased; (b) the range of appropriate matrix reference materials available is limited; and (c) there may be a matrix mismatch between the test material and the most appropriate reference material available.

In the last instance the recovery value obtained from the reference material would not be strictly applicable to the test material. The shortfall applies especially in sectors such as foodstuffs analysis where reference materials have to be finely powdered and dried to ensure homogeneity and stability. Such treatment is likely to affect the recovery in comparison with that pertaining to fresh foods of the same kind. However, matrix mismatch is a general problem in the application of recovery information and is treated separately.

Recovery information from surrogates

Where (certified) reference materials are unavailable, the recovery of analyte can be estimated by studying the recovery of an added compound or element that is regarded as a surrogate for the native analyte. The degree to which this surrogate is transferred into the measurement phase is estimated separately and this recovery can, if appropriate, be attributed also to the native analyte. This procedure in principle allows the loss of analyte to be corrected, and an unbiased estimate of the concentration of the native analyte in the original matrix to be made. Such a 'correction-for-recovery' methodology is implicit or explicit in several distinct methods of analysis and must be regarded as a valid procedure if it can be shown to be properly executed.

In order for this procedure to be valid the surrogate must behave quantitatively in the same way as analyte that is native in the matrix, especially in regard to its partition between the various phases. In practice that equivalence is often difficult to demonstrate and certain assumptions have to be made. The nature of these assumptions can be seen by considering the various types of surrogate that are used.

Isotope dilution

The best type of surrogate is an isotopically modified version of the analyte which is used in an isotope dilution approach. The chemical properties of the surrogate are identical with, or very close to, those of the native analyte and, so long as the added analyte and the native analyte come to effective equilibrium, its recovery will be the same as that of the analyte. In isotope dilution methods the recovery of the surrogate can be estimated separately by mass spectrometry or by radiometric measurement if a radioisotope has been used, and validly applied to the native analyte. The achievement of effective equilibrium is not always easy, however.

In some chemical systems, for example in the determination of trace metals in organic matter, the native analyte and the surrogate can be readily converted into the same chemical form by the application of vigorous reagents that destroy the matrix. This treatment converts organically bound metal into simple ions that are in effective equilibrium with the surrogate. Such a simple procedure is usually effective in the determination of trace elements, but might not apply to a pesticide residue. In the latter instance the analyte may be in part chemically bound to the matrix. Vigorous chemical reagents could not be used to release the analyte without the danger of destroying it. The native analyte and surrogate cannot come into effective equilibrium. The recovery of the surrogate is therefore likely to be greater than that of the native analyte. Thus even for this best type of surrogate, a bias in an estimated recovery may arise. Moreover, the application of the isotope dilution approach is limited by the availability and cost of isotopically enriched analytes.

Spiking

A less costly expedient, and one very commonly applied, is to estimate in a separate experiment the recovery of the analyte added as a spike. If a matrix blank (a specimen of the matrix containing effectively none of the analyte) is available the analyte can be spiked into that and its recovery determined after application of the normal analytical procedure. If no matrix blank is available, the spike can be added to an ordinary test portion that is analysed alongside an unspiked test portion. The difference between these two results is the recovered part of the added analyte, which can be compared with the known amount added. This type of recovery estimate is called here the 'surrogate recovery' (the added analyte acts as a surrogate for the native analyte). It is analogous to the method of standard additions. It suffers from the same problem as that encountered with isotopically modified analyte, namely that added analyte may not come to effective equilibrium with the native analyte. If the added analyte is not so firmly bound to the matrix as the native analyte, the surrogate recovery will tend to be high in relation to that of the native analyte. That circumstance would lead to a negative bias in a corrected analytical result.

Internal standards

A third type of surrogate used for recovery estimation is the internal standard. When internal standardisation is used in recovery experiments the surrogate is an entity chemically distinct from the analytes, and therefore will not have identical chemical properties. However, it will normally be selected so as to be closely related chemically to the analytes, thus representing their chemical behaviour to the highest degree practicable. The internal standard would be used, for example, in recovery estimation where numerous analytes are to be determined in the same matrix and marginal recovery experiments would be impracticable for each of them individually. The question of practicability goes beyond the costs of handling numerous analytes: some analytes (for example, new veterinary residues, or metabolites) may not be available as pure substances. While it may be the most cost-effective expedient in some circumstances, the internal standard at best is technically less satisfactory than the spike as a surrogate, because its chemical properties are not identical with those of the analytes. Biases in both directions could result from the use of a recovery estimate based on an internal standard. Internal standards may also be used for other purposes.

Matrix mismatch

Matrix mismatch occurs when a recovery value is estimated for one matrix and applied to another. The effect of matrix mismatch would be manifested as a bias in the recovery in addition to those considered above. The effect is likely to be most serious when the two matrices differ considerably in their chemical nature. However, even when the matrices are reasonably well matched (say two different species of vegetable) or nominally identical (for example, two different specimens of bovine liver), the analytical chemist may be forced to make the unsubstantiated assumption that the recovery is still appropriate. This would clearly increase the uncertainty in the recovery and in a recovery-corrected result. Matrix mismatch can be avoided in principle by a recovery experiment (for example, by spiking) for each separate test material analysed. However, such an approach will often be impracticable on a cost-benefit basis so a representative test material in each analytical run is used to determine the recovery.

Concentration of analyte

The recovery of the surrogate or the native analyte has up to this point been treated as if it were independent of its concentration. This is unlikely to be strictly true at low concentrations. For instance a proportion of the analyte may be unrecoverable by virtue of irreversible adsorption on surfaces. However, once the adsorption sites are all occupied, which would occur at a particular concentration of analyte, no further loss is likely at higher concentrations. Hence the recovery would not be proportional to concentration. Circumstances like this should be investigated during the validation of an analytical method, but a complete study may be too time-consuming for *ad hoc* use.

ESTIMATION OF RECOVERY

There is no generally applicable procedure for estimating recovery that is free from shortcomings. However, it is possible to conduct a 'thought experiment' in which an ideal procedure is used. This provides a reference point for real procedures. In this ideal procedure a definitive analytical method is available: the analyte can be determined by a method that is completely unbiased with no recovery losses. The method is too resource-intensive for use in routine analysis, but there is an alternative routine method with imperfect recovery. The recovery obtained in the routine method is estimated by using both methods to analyse a large set of typical test materials, a set that covers the required range of matrices and analyte concentrations. This gives the recovery (and its uncertainty) for the routine method for any conceivable situation.

In practice there may be no such definitive method available for reference, so reference materials or surrogate studies have to be used for the estimation of recovery. However, reference materials are few, and lack of resources restricts the range of test materials that can be used to estimate recovery by using surrogates. Additionally, the use of surrogates in itself adds an uncertainty to a recovery estimate because it may not be possible to determine whether some proportion of the native analyte is covalently or otherwise strongly bound to the matrix and hence not recoverable.

A strategy commonly employed to handle this problem is to estimate recovery during the process of method validation. Recoveries are determined over as wide a range of pertinent matrices and analyte concentrations as resources allow. These values are then held to apply during subsequent use of the analytical method. To justify that assumption, all routine runs of the method must contain a reference material (or spiked samples) to act as internal quality control. This helps to ensure that the analytical system does not change in any significant way that would invalidate the original estimates of the recovery. The following points are therefore suggested as requiring consideration, even if lack of resources prevents their complete execution in practice.

<u>**Representative recovery studies**</u>

The entire range of matrix types for which the method will be applied should be available for the method validation. Moreover, several examples of each type should be used to estimate normal range of recoveries (the uncertainty) for that matrix type. If it is likely that the history of the material will affect the recovery of the analyte (for example, the technical processing or cooking of foodstuffs), then examples at different stages of the processing should be procured. If this range cannot be encompassed in the validation, there will be an extra uncertainty associated with the matrix mismatch in the use of the recovery. That uncertainty may have to be estimated from experience.

An appropriate range of analyte concentrations should be investigated where that is technically and financially possible, because the recovery of the analyte may be concentration-dependent. Consider adding an analyte to a matrix at several different levels. At very low levels the analyte may be largely chemisorbed at a limited number of sites on the matrix or irreversible adsorbed onto surfaces of the analytical vessels. Recovery at this concentration level might be close to zero. At a somewhat higher level, where the analyte is in excess of that so adsorbed, the recovery will be partial. At considerably higher concentrations, where the adsorbed analyte is only a small fraction of the total analyte, the recovery may be effectively complete. The analytical chemist may need to have information about recovery over all of these concentration ranges. In default of complete coverage, it may be suitable to estimate recovery at some critical level of analyte concentration, for example at a regulatory limit. Values at other levels would have to be estimated by experience, again with an additional uncertainty.

When spiking is applied to a matrix blank then the whole range of concentrations can be conveniently considered. When the concentration of the native analyte is appreciable the spike added should be at least as great, to avoid incurring a relatively large uncertainty in the surrogate recovery.

Internal quality control

The principles and application of internal quality control (IQC) are described. The purpose of IQC is to ensure that the performance of the analytical system remains effectively unchanged during its use. The concept of statistical control is crucial in IQC applied to routine analysis (as opposed to *ad hoc* analysis). When applied to recovery, IQC has some special features that have to be taken into account. This IQC of recovery can be addressed in two distinct ways, depending on the type of control material that is used.

(a) A matrix-matched reference material can be used as a control material. The recovery for this material and an initial estimate of its between-run variability are determined at the time of method validation. In subsequent routine runs the material is analysed exactly as if it were a normal test material, and its value plotted on a control chart (or the mathematical equivalent). If the result for a run is in control, then the validation-time estimate of the recovery is taken as valid for the run. If the result is out of control, further investigation is required, which may entail the rejection of the results of the run or possibly a re-investigation of the recovery. It may be necessary to use several control materials, depending on the length of the run, the analyte concentration range, etc.

(b) Spiked materials can also be used for quality control. As usual, initial estimates of the average recovery and its between-run variability are made during method validation, and are used to set up a control chart. Either of two variant approaches can be used in routine analysis, depending on the stability of the material: (a) a single long-term control material (or several such materials) can be prepared for use in each routine run, or (b) all, or a random selection, of the test materials for the run can be spiked. In either instance the surrogate recovery is plotted on a control chart. While the recovery remains in control it can be deemed to apply to the test materials generally. Of the two alternative methods, the latter (involving the actual test materials) is probably the more representative, but also the more demanding.

There is a tendency for the role of IQC to be confused with the simple estimation of recovery (where deemed appropriate). It is better to regard IQC results solely as a means of checking that the analytical process remains in control. The recovery estimated at method validation time is usually more accurate for application to subsequent in-control runs, because more time can be spent on studying their typical levels and variability. If real-time spiking is used to correct for recovery, this is more like a species of calibration by standard additions.

--