codex alimentarius commission

FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS

WORLD HEALTH ORGANIZATION

JOINT OFFICE: Viale delle Terme di Caracalla 00100 ROME Tel: 39 06 57051 www.codexalimentarius.net Email: codex@fao.org Facsimile: 39 06 5705 4593

Agenda Item 9 **CX/MAS 06/27/10**
April 2006

JOINT FAO/WHO FOOD STANDARDS PROGRAMME

CODEX COMMITTEE ON METHODS OF ANALYSIS AND SAMPLING
Twenty-seventh Session
Budapest, Hungary, 15-19 May 2006

UNCERTAINTY OF SAMPLING

BACKGROUND

Sampling has been considered by the Codex Committee on Methods of Analysis and Sampling for many years. One of the primary outputs in this area has been the General Guidelines in Sampling, CAC/GL 50-2004. In addition the CAC has developed Guidelines on [Analytical] Measurement Uncertainty, CAC/GL 54-2004.

However, a number of other initiatives have been progressing, most notably the development of the Eurachem/EUROLAB/CITAC/Nordtest Guide on the "Estimation of Measurement Uncertainty Arising from Sampling". This is currently undergoing development in the Eurachem Sampling Working Group.

It was agreed at the last Session of CCMAS that these activities would be brought to the notice of the Committee. This paper explains some of the background to the issue of uncertainty of sampling. This is given in the Annex to this paper.

RECOMMENDATIONS

It is recommended that the Committee discusses the issue of uncertainty and sampling and decides whether it should develop recommendations in the area in the same way that it already has for [Analytical] Measurement Uncertainty.

The Committee should also decide whether it wishes to progress the topic as a defined New Work Item.

ANNEX: SAMPLING UNCERTAINTY – ESTIMATION AND INTERPRETATION

BACKGROUND

It is widely accepted that repeat analyses of the same sample will almost always produce varying results. These variations may be due to e.g. changes in the operating conditions, and an inhomogeneous sample from which only a small test portion is taken. Persons responsible for producing, appraising and interpreting the results of chemical analyses will be familiar with terms such as reproducibility and repeatability - both are measures of this random variability. They will also be familiar with the use of 'reference materials' and terms such as 'bias' and 'recovery', which are used to check if analytical results are systematically higher or lower than they should be, when compared to a known reference value. The random variability and systematic effects in analytical results are characterised as analytical uncertainty.

Chemical analysis is usually the end part of the measurement process, following the taking of samples (sampling) and grinding, blending and treatment of samples in preparation for chemical analysis (physical preparation). The term 'measurement' (as in measurement uncertainty) encompasses the whole procedure. Each step in the measurement process will introduce variability in the final measurement result, the measurement uncertainty. The International Standards Organisation defines uncertainty of measurement as 'parameter, associated with the result of a measurement that characterises the dispersion of the values that could reasonably be attributed to the measurand' (ISO GUM 1993).

The Codex General Guidelines on Sampling (CAC/GL 50-2004) are based on the principals of acceptance sampling. They are designed to ensure that fair and valid sampling procedures are used when food is being tested for compliance with a particular Codex commodity standard. These Guidelines make the distinction between sampling error and measurement error. For the purpose of the Guidelines measurement error (caused by the measured value of the characteristic failing to accurately represent the true value of the characteristic within the sample) is analogous to analytical uncertainty. Like analytical uncertainty, sampling error (caused by the sample failing to accurately represent the population from which it was collected) has input from both systematic and random effects. The CAC Guidelines advise it is desirable that the sampling errors associated with any sampling plan, as well as measurement errors associated with analysis, should be quantified and minimised. Laboratories are required, as part of 3rd party accreditation, to participate in inter-laboratory trials, data from these and other internal quality control measures allow the estimation of analytical uncertainties. Methods for estimating sampling uncertainty have been published; however neither a consensus method nor routine implementation is forth coming at this time.

The Eurachem/EUROLAB/CITAC/Nordtest Working Group on Uncertainty from Sampling was formed in September 2003. This Working Group includes representatives from a wide range of disciplines, including those from the food sector. The Eurachem Working Group is currently preparing guidance for the evaluation of uncertainties in measurement arising from the process of sampling. This guidance will be applicable to all chemical measurements that require the taking of a sample. It will provide guidance on the assessment of the uncertainty of the measurement that is caused by the process of sampling, and any physical preparation of the sample prior to analysis, and how this can be combined with estimates of uncertainty arising from the analytical process. These guides will be developed in collaboration with relevant international bodies and will be updated as experience is gained in their use.

Once an estimate of uncertainty has been made, we are presented with the question as to whether the estimate is fit for purpose - we are in essence asking 'is it good enough?' It may be the case that a reduction in measurement uncertainty is desirable. A reduction can be reached by simply increasing sample mass, increasing the number of analyses and so on. However, it is unrealistic to expect an unrestricted financial budget or time-scale for completion of analysis. Given these external considerations, a decision needs to be made on how much uncertainty can be tolerated.

This document looks firstly at the methods of estimating uncertainty and uses real case studies to exemplify each. The role of measurement uncertainty in the decision making process is also addressed, as is the assessment of fitness for purpose. The second part of this document examines whether it is a good idea to set global fitness for purpose criteria for sampling uncertainty. This document is focussed on measurement processes that result in quantitative data. Qualitative data (e.g. yes / no responses) are not addressed

METHODS FOR ESTIMATING SAMPLING UNCERTAINTY

Sampling theorists consider that sampling error can be minimised by using a correct sampling protocol. Theory of sampling relies on a great deal of prior knowledge of both the sampling target and characteristics are required, e.g. particle-size distribution factors, density, shape factors (Gy, 1979). This methodology was initially applied to bulk particulate materials, and wider application requires a correct interpretation of the theory. Without a good grasp of the principles, adaptation of the theory across the range of commodities covered by Codex will be difficult.

The method described herein has been found to be broadly applicable across the food sector. The examples cover a range of sampling exercises that span from grower level to retail sampling.

The duplicate method – general principles

A sampling protocol (detailing, how many samples, how to sample, sample mass etc.) is a prerequisite for all food surveys, assessments etc. The duplicate method requires a second (duplicate) sample to be taken for 10% (or a minimum of 8) of the total number of sampling targets. This second 'duplicate' sample should be taken to represent the ambiguity in interpreting the protocol, what this means is perhaps better explained using the examples.
The duplicate samples are then each subject to independent physical preparation (i.e. they are not combined). Two analytical test portions are drawn from each of the duplicate 'prepared' samples – see Figure 1.

All test portions are anonymised (so it is unclear which are duplicates) and subsequently analysed in a randomised order.

A statistical technique called analysis of variance (ANOVA) is applied to the resultant data to separate out between-target variances, sampling (or within-target) variances and analytical variances.

The inclusion of certified reference materials (CRM) and /or spike samples within the analytical run will allow the systematic effects of analysis to be quantified. This is generally routine in most laboratories. As described, the duplicate method does not permit the estimation of systematic effects from the sampling process. When the duplicate method of uncertainty estimation is utilised, the costs will increase by 10% for sampling and 30% for analysis.

EXAMPLE 1 – NITRATE CONCENTRATION IN GLASSHOUSE LETTUCE

This example is included in the draft Eurachem Uncertainty from Sampling Working Group document entitled: Estimation of measurement uncertainty arising from sampling

Aim: To estimate the average concentration of nitrate (in mg kg^{-1}) in a bay of lettuce.
For this study each 'bay' was considered equivalent to a batch of lettuce, and a bay of lettuce was the sampling target.

The routine sample: The sampling was planned for the winter growing season (October – April). The concentration of nitrate in glasshouse grown lettuce is regulated by EC Regulation 563/2002.
The routine sampling protocol applied for this analyte-commodity combination required 10 heads of lettuce to be cut from each bay of lettuce. The protocol instructs samplers to cut the samples whilst walking either a 'W' or '5-point die' through the bay under investigation. The first sample (S1 - usually be the only sample taken from the bay) was collected by the samplers using their routine interpretation of the protocol.

The duplicate sample: The protocol did not give any specific information on how to orient either design. In this respect either a W or 5-point die could be applied, and orientated in any direction (examples are given in Figure 2). All are equally valid under the protocol.

For the purpose of estimating sampling uncertainty the duplicate sample (S2) was taken by the samplers using a different interpretation of the protocol (as instructed by the researchers).

Both samples (S1 and S2) were transported to the analytical laboratory in identical ice-packed cool boxes.

Sample preparation and analysis: On receipt at the laboratory each 10-head sample was reduced in size (i.e. each head was cut into four and opposite quarters selected) and macerated. Two 30g test samples were drawn from the homogenate, i.e. A1 and A2, for each of the duplicate sample. Extraction was by routine accredited procedures (water extraction with quantification by HLPC). Spike samples were run concurrently with the samples to provide an estimate of recovery.
No significant analytical bias could be detected and so bias correction was considered unnecessary for the resultant data.

This measurement process was repeated for eight sampling targets (bays). In practice the eight duplicate samples were achieved during two sampling exercises.

CALCULATING UNCERTAINTY

Analysis of variance provided estimates of sampling and analytical uncertainty as standard deviations – see Table 1.

Classical Results:
Mean = 4345.5625
Sums of Squares = 12577113, 4471511, 351320
Sigma values (between bay, sampling, analysis) = 556.2804, 518.16089, 148.18063
Percent variance (between bay, sampling, analysis) = 51.583582, 44.756204, 3.6602174
Sigma (total) = 774.5296

Robust Results:
Mean = 4408.3237
Sigma values (between bay, sampling, analysis) = 565.39868, 319.04834, 167.94308
Percent variance (between bay, sampling, analysis) = 71.090791, 22.636889, 6.2723172
Sigma (total) = 670.57617

Table 1: Results of the analysis of variance of data from the duplicate sampling and analysis of lettuce for nitrate. Estimates of sigma and mean are quoted in concentration units of mg kg^{-1}

For uncertainty estimation, the robust estimates are used preferentially to the classical estimates. Robust statistics allow the effects of outliers to be down-weighted and incorporated into the calculations. Classical estimates can be adversely affected outlying values. For a more in-depth discussion on robust statistics please see 'Robust statistics–how not to reject outliers. Part 1. Basic concepts, Analytical Methods Committee, Analyst, 1989, (12), 1693-1697'

22.6% of the overall variability arises due to sampling and only 6.3% is due to the chemical analysis. The most variation was observed between bays of lettuce (71.1% of total).

UNCERTAINTY ESTIMATES

Mean = 4408 mg kg^{-1}	Sampling (within-batch)	Analytical	Measurement
As 's' in mg kg^{-1}	319.04	167.94	360.55
As U% at 95% confidence (k = 2)	14.475	7.619	16.358

EXAMPLE 2 – INFANT WET MEALS (RETAIL SURVEY)

Aim: To estimate concentrations of cadmium in infant wet meals (in mg kg^{-1}), as part of a survey. Survey data may be used in risk/exposure assessment. Wet meals can be considered as food given to an infant at mealtimes, which does not require the addition of water/fluid. For this study a batch of a particular wet meal (identified by unique batch code) was considered to be the sampling target.

The routine sample: Each sampling target (in terms of provenance, brand name, product type, size) was identified prior to the sampling event. Samplers were instructed to purchase three of each target, all from the same batch, i.e. 3 glass jars, metal cans etc. Two of the pots were analysed independently to produce two discrete concentration estimates. This could allow a rudimentary estimate of within-batch variability. For the purpose of uncertainty estimation one of the pots was randomly selected as S1. The third pot was retained by the laboratory as a reference sample.

The duplicate sample: The protocol did not specify specific batch codes from which to sample. At each retail establishment, there was more than one batch available for purchase. The likelihood of the sampler or a member of the public selecting from either batch was considered equivalent. Therefore the 'duplicate sample' was taken from a second batch (S2) which allowed the preservation of the original experimental design. Duplicate samples were taken for 8 wet meals (sampling targets). The estimate of sampling uncertainty represented between-batch variability.

Sample preparation and analysis: The Cd concentration was determined for each sample using a UKAS accredited methods, quantification was by ICP-MS (ELAN 6000). Both S1 and S2 samples were analysed in duplicate. All other samples were analysed singularly.

CALCULATING UNCERTAINTY

Analysis of variance provided estimates of sampling and analytical uncertainty as standard deviations – see Table 2.

CLASSICAL RESULTS
Mean = 7.5748205
Sigma values (between-wet meal, sampling, analysis) = 1.9620616, 0, 2.2179029
Percent variance (between-wet meal, sampling, analysis) = 43.90219, 0, 56.097815
Sigma (total) = 2.9612124

ROBUST RESULTS
Mean = 7.574821
Sigma values (between-wet meal, sampling, analysis) = 2.2573242, 1.234913, 1.0998778
Percent variance (between-wet meal, sampling, analysis) = 65.074678, 19.47587, 15.449451
Sigma (total) = 2.798259

Table 2: Results of the analysis of variance of data from the duplicate sampling and analysis of infant wet meals for Cd. Estimates of sigma and mean are quoted in concentration units of µg kg^{-1}

The benefit of robust statistics is emphasised here. A number of samples returned outlying values for one of the duplicated analyses, e.g. S1A1 = 10.903, S1A2 = 3.901 µg kg^{-1}. Such outliers mask the influence of the sampling when classical statistics are used.

For this food-analyte combination the processes of sampling and analysis contribute similar amounts to the overall variability, 19.5% and 15.4% (of the total variance) respectively. This is unsurprising due to the amount of processing that precedes packing and the low concentrations of Cd.

UNCERTAINTY ESTIMATES

Mean = 7.575 µg kg^{-1}	Sampling (between-batch)	Analytical	Measurement

As 's' in µg kg^{-1}	1.235	1.100	1.654
As U% at 95% confidence (k = 2)	32.61	29.04	43.66

EXAMPLE 3 – MOISTURE IN WHOLESALE BUTTER (OFFERED FOR EU SUBSIDY)

Aim: To estimate the moisture content of a batch of butter put forward for subsidy payment (EC 2571/97). To achieve the minimum quality standards required, the batch must contain a maximum of 16% moisture (m/m). It should be noted that other quality requirements should be satisfied before a subsidy is paid.
For this study a c. 20 tonne batch of unsalted butter (typically a days production) was considered to be the sampling target. Each batch was comprised of 25 kg (individually cased) blocks of butter, i.e. 40 * 25 kg per 20 tonne batch.

The routine sample: Prior to the physical taking of the sample, an appropriate number of 25 kg blocks were selected from the batch under inspection. The number of blocks is dependent on the mass of the batch, e.g. for a 20 tonne batch, 6 blocks were selected. The six blocks were left to temper for 48 hours. On the day of sampling a 500 g increment was cut from the edge of each of the blocks. The six increments were later processed and combined to produce two 3-fold composite samples.

The duplicate sample: For this case study an estimate of within-batch sampling uncertainty was required. Decisions on whether to award the subsidy are made on a batch-by-batch basis. As two independent results are routinely presented, the two 3-fold composite samples can be considered as the duplicate samples. Although this is not an implementation of the duplicate method in the purest sense, it is time, cost and space efficient for routine surveys.

Sample preparation and analysis: The six 500 g increments were transported to the laboratory. For each increment, 200 g was removed for further analysis and the remainder was used for sensory analysis. Two 3-fold composites were produced using the 6 increments. Each composite sample was analysed for moisture, as determined by drying of a known mass of butter at 102°C ± 2°C and weighing to determine loss. Moisture is expressed in units of % m/m.

CALCULATING UNCERTAINTY

Analysis of variance provided estimates of sampling and analytical uncertainty as standard deviations – see Table 3.

Classical Results:
Mean = 15.746875
Sums of Squares = 2.0025887, 1.2109982, 0.024100207
Sigma values (between-batch, sampling, analysis) = 0.18351385, 0.27374122, 0.038810603
Percent variance (between-batch, sampling, analysis) = 30.582991, 68.049141, 1.3678644
Sigma (total) = 0.3318401

Robust Results:
Mean = 15.754773
Sigma values (between-batch, sampling, analysis) = 0.25005788, 0.19465271, 0.042098258
Percent variance (between-batch, sampling, analysis) = 61.188376, 37.077355, 1.7342675
Sigma (total) = 0.31967309

Table 3: Results of the analysis of variance of data from the duplicate sampling and analysis of moisture in wholesale butter. Estimates of sigma and mean are quoted in concentration units of % (m/m)

Butter is produced within specification limits and compared to other food-types is fairly homogenous. As expected the variation in moisture concentration dominates between batch (61.2% of total). Compared to the

analytical uncertainty (0.042 % m/m), the sampling uncertainty (0.195%) dominates the measurement procedure.

UNCERTAINTY ESTIMATES

Mean = 15.754 % (m/m)	Sampling (between-batch)	Analytical	Measurement
As 's' in % (m/m)	0.1947	0.0421	0.1992
As U% at 95% confidence (k = 2)	2.471	0.5344	2.5282

OTHER METHODS OF UNCERTAINTY ESTIMATION

It is possible to extend this relatively simple method to a more comprehensive approach that will allow the estimation of sampling bias. The concepts of analytical collaborative trials and proficiency tests can also be applied to primary sampling, i.e. a collaborative trial in sampling (CTS) or a sampling proficiency test (SPT). Because multiple samplers (at least 8 in each case) are utilised, any within-sampler bias becomes randomly distributed between the trial participants, and is consequently included in the estimate of sampling uncertainty. CTS can be used to assess the performance of a particular protocol, with each participant applying the same sampling protocol to the same sampling target. In an SPT the participants apply a protocol that they have selected as being most appropriate, to the same target. The SPT allows the proficiently of the samplers to be assessed. These inter-organisational trials can be implemented, not only for uncertainty estimation, but also for identifying the need for further training of sampling personnel and/or a more appropriate protocol.

One perceived difficulty in applying these concepts to the food sector is finding a suitable sampling target where both the analyte concentration, and commodity remain relatively invariant in time (e.g. pesticides loss / degradation and dairy products become rancid). As the commodities and analyte concentrations change with time, the characteristics of the sampling target (e.g. heterogeneity) will also change, ultimately affecting the uncertainty estimates. Provisional work applying a collaborative trial to foodstuffs has been undertaken with positive results. The UK Food Standards Agency is funding a 2-year research project to look at the general feasibility of sampling proficiency test within the food sector. The final report is due in October 2007.

UNCERTAINTY AND THE DECISION MAKING PROCESS

The acceptability of food is often assessed using concentration estimates that are compared against a statutory limit. If there is no value for uncertainty quoted, the concentrations are simply classed as above the threshold (e.g. batch failed) or below the threshold (e.g. batch passed). When a value of measurement uncertainty is available we have more evidence to support these decisions. This has been recognised within the legal framework of enforcement. In order to bring a case of non-compliance the measured value and the analytical uncertainty must exceed the statutory limit – see Figure 3.

At present no consideration is given to the sampling uncertainty, often it is assumed that the use of standard protocols will ensure a 'correct' and representative sample. This, however is rarely the case. If we re-examine the example in Figure 3, and include the sampling uncertainty also, it is clear that erroneous decisions can still be made – see Figure 4. It should be borne in mind that in most cases the sampling uncertainty is the dominant contributor to the measurement uncertainty as a whole.

EXAMPLE – NITRATE IN LETTUCE

Scenario 1: A routine sample has been taken from a batch of lettuce grown under glass in October. The concentration of nitrate estimated from the 10-head sample was 5028 mg kg^{-1}. On first impressions, the measured nitrate concentration appears to exceed the regulatory limit, i.e. 5028 mg kg^{-1} > 4500 mg kg^{-1}.

The relative analytical uncertainty (AU) has been estimated to be 7.619% (of the measured value – Example 1). The batch will be considered as non-compliant if the measured concentration, less the AU is greater than the threshold value:

What is the concentration of c_m – AU?

c_m - AU = 5028 – 5028/100*7.619
c_m - AU = 5028 – 383.08
c_m – AU = 4644 mg kg^{-1}

A nitrate concentration of 4644 mg kg^{-1} is still greater than the threshold value of 4500 mg kg^{-1}. The batch can therefore be classed as non-compliant with a high degree of certainty.

However, if the total measurement uncertainty (MU, including sampling uncertainty) is considered in this assessment will the same classification apply? We must now consider if the measured concentration less the total measurement uncertainty is greater than the threshold?

c_m – MU = 5028 – 5028/100*16.358
c_m – MU = 5028 – 822.48
c_m – MU = 4205 mg kg^{-1}

When sampling uncertainty is estimated and considered in the assessment of results the non-compliance cannot be proven with a high level of certainty. If a sample returns a concentration greater than the threshold, there is a real possibility that the true assessment should be compliance.

Scenario 2: On another occasion, a second routine sample has been taken from a batch of lettuce grown under glass in October. The concentration of nitrate estimated from the 10-head sample was 4061 mg kg^{-1}. A brief assessment of the result would indicate that this batch of lettuce complies with EC legislation.

Under usual circumstances this result would be considered compliant without consideration of the analytical uncertainty. However, for the purpose of this document we will check the validity of this conclusion in view of the measurement uncertainty estimate.

In this example, it is necessary to ask if the measured concentration (c_m) plus the MU (16.3 % - see Example 1) is less than the threshold value?

c_m – MU = 4021 + 4021/100*16.358
c_m – MU = 4021 + 657.76
c_m – MU = 4679 mg kg^{-1}

This basic calculation shows that although the batch has been deemed compliant, there is a real probability that the nitrate levels are greater than the threshold, and the correct assessment would have been that of non-compliance. This is what is known as a false-compliance scenario.

Clearly each misclassification scenario (false compliance and false non-compliance) will result in different consequences. In terms of enforcing legislation, a false non-compliance scenario (if subsequently proved) would result in a case being lost/rejected and associated financial liabilities. Unlike a false non-compliance situation, false compliance could result in serious risks to public health, e.g. the release of food that is unfit for human consumption, to the market place.

HOW MUCH MEASUREMENT UNCERTAINTY IS ACCEPTABLE?

Ideally there would be zero uncertainty and the results of measurements would always be reported as the true value. In reality, the answer to this question does in part depend on the aims of the end user of the data. Differentiating between high and low levels of a particular contaminant can be successfully achieved with relatively high levels of measurement uncertainty. Conversely, proving guilt in a disputed non-compliance enforcement case would require low uncertainty. Budgetary constraints will influence the number and size of

samples, as well as the analytical methods employed, and so will have some influence on the uncertainty. It often follows that having a high budget for sampling and analysis will result in a low level of measurement uncertainty and a reduced risk of erroneous decisions – this is not always the case! Larger samples can be taken and more precise methods of analysis employed. However, the analyte and food combination for which the survey is designed may not warrant these high costs, i.e. data that is adequate for the end purpose could be achieved by more uncertain methods at a lower cost. Where there is a direct risk to consumer health and/or for a contentious analyte (e.g. GMO) the consequence costs of making wrong decisions due to high levels of measurement uncertainty would be very large. Such costs would not only be high in terms of monetary loss, but also in terms of damaging the reputation of the regulatory body.

The Optimised Uncertainty (OU) method provides a way in which the level of uncertainty can be assessed against the potential financial costs that may be incurred as a result of misclassification. This methodology utilises an economic loss function to calculate the expectation of financial loss for a given level of uncertainty. Such a loss includes contributions due to misclassification, and also the real cost off making the measurements. It is called an 'expectation of loss' because the loss is not always realised. Estimates of measurement uncertainty (as standard deviations) are used to compute the probability of misclassification at a chosen concentration of analyte. Using estimates of consequence costs of misclassification (e.g. recall, litigation, compensation etc.) and real measurement costs an overall value for the expectation of loss can be computed. Once a range of values for measurement uncertainty is used in the loss function, a graph can be plotted which shows how the expected loss changes as a function of the uncertainty – see Figure 5. It is the uncertainty that gives us the lowest financial loss that we should aim to achieve.

Once the optimal level of MU has been determined it can be divided to give the optimal proportions between SU and AU. Comparing the original estimates (from the duplicate designs) with the calculated optimal values it is possible to see what improvements may be needed. Furthermore optimal levels of expenditure per sample and per analysis are computed. Practical examples of this methodology can be found in the literature.

An alternative approach to managing levels of sampling uncertainty is to set a standard fitness for purpose criteria. The advantages and disadvantages of this proposed method are explored further below.

IS IT A GOOD IDEA TO SET GLOBAL FITNESS FOR PURPOSE CRITERIA FOR SAMPLING UNCERTAINTY?

Requirements for performance characteristics of analytical methods are increasingly defined in regulations. Such characteristics include target values of repeatability, reproducibility and recovery across a range of concentration intervals. The estimation and acceptability of analytical uncertainty estimates are expressly referred to within these official documents and will soon be widely used within the decision making process. Currently, the quality of sampling is not addressed, beyond the broad requirement that, for example 'the sampling method applied shall ensure that the aggregate sample is representative for the lot that is to be controlled' (2003/78/EC: Annex I, Section 4). The question therefore arises as to whether equivalent limits can be placed on the performance characteristics of sampling methods and should sampling quality be legislated and, if so, in what way?

One potential method of legislating sampling quality would be to define 'global fitness for purpose criteria' for the sampling process. What this means is that an acceptable level of sampling uncertainty will be agreed and implemented for all food types. This level may be for example ± 20% (of measured concentration), e.g. for a measured nitrate concentration of 4061 mg kg^{-1}, the sampling uncertainty will be considered fit for purpose as long as it was <20% of 4061 mg kg^{-1}, 812 mg kg^{-1}. Previous work within the geochemical sector has used a similar fitness for purpose criteria, in that the sampling variance should be <20% of the total variance (not the measured value). This criterion has been applied for stream sediments and soil sampling in order to make a reliable interpretation of the measurements.

In principle the idea of setting such fitness for purpose criteria is desirable. Effective implementation of the system will first require a clear definition of who is responsible for the sampling quality. At present there is no explicit guidance on this e.g. is it the regulators, local authority, contracted laboratory, survey planners or survey operators? It may be that the responsibility will vary on a case-by-case basis.

Routine estimation of sampling uncertainty will be a prerequisite of the proposed method and a consensus on the appropriate estimation method (or harmonised method) would also be necessary. Section 1 of this document describes the duplicate method for uncertainty estimation. This has been applied successfully across a broad range of analyte-foodstuff combinations. The increased financial expenditure associated with this method is justified by the supplementary information gained and subsequently used in the decision making process. Knowledge of the measurement uncertainty would allow the assessor to distinguish between samples which are contaminated from those which are 'uncontaminated', 'possibly contaminated' and 'probably contaminated', at a defined level of statistical confidence. Probabilistic calculations of maximum risk could be undertaken based on the uncertainty values defined by the fitness for purpose criterion. This could potentially result in improved reliability of decisions and assessment of risks. Theoretically this could in turn lead to increased public confidence (e.g. reduction in food scares) and a reduction in trade disputes, given that all parties will be required to conform to the same legislative uncertainty requirements.

Over a longer time frame the uncertainty data can be used to improve the overall assessment process by indicating whether sampling or analysis is the dominant source of uncertainty. Subsequently the uncertainty estimates can be used to justify a more appropriate budgetary allocation between sampling and analysis. Where a particular analyte-foodstuff combination persistently reports uncertainty estimates outside of the criterion a reduction in uncertainty would be necessary. Consequently, the introduction of such sampling legislation would lead to a tightening of the controls of sampling quality. However the required changes in uncertainty may not be realised in practice due to limiting factors, such as uneven sample heterogeneity that cannot be predicted by current sampling theory.

Before any global fitness for purpose criteria can be laid down, a greater knowledge of the levels of sampling uncertainty across the analyte-foodstuffs covered by the Codex remit is needed. Setting any limits without obtaining such knowledge would be both difficult and scientifically unsound.

Sampling uncertainty arises, largely, although not solely, due to the characteristics of the analyte-foodstuff under investigation. Heterogeneous contamination produces high levels of sampling uncertainty (e.g. for total aflatoxin in pistachio nuts, 45%), whereas more homogenous anayte distributions produce minimal levels of sampling uncertainty (e.g. moisture in butter, 2.47%). It should be noted that high levels of relative uncertainty arise inevitably for foodstuffs with low analyte concentrations due to how percentage values are calculated. A low (absolute) level of sampling uncertainty estimated at low mean analyte concentration can give a high value of percentage uncertainty giving an apparently pessimistic viewpoint of the sampling quality. A real example of this can be found in Example 3 in Section 1 of this document. The mean concentration of Cd in infant wet meals was estimated to be 7.57 μg kg^{-1} and the standard deviation for sampling was 1.23 μg kg^{-1}. Both estimates when viewed as absolute concentration are relatively low, however, the Usamp% is calculated as 32.6%. If a legislative limit where to be set, then due consideration should be paid to both the analyte distribution and the expected concentrations. The fact that uncertainty changes as a function of concentration should also be considered. In view of this, a model would need to be constructed for each analyte-foodstuff. The models would allow establishment of fitness for purpose criterion that are appropriate over the full range of concentrations routinely considered.

A blanket limit on the level of uncertainty would not take into account the variable potential consequences of the resultant misclassification of food both to human health and also to financial budgets. A product under investigation that is suspected of containing a toxic substance would require tighter controls on sampling uncertainty compared to an investigation of non-toxic constituents, due to the different potential risks to human health e.g. Cd in infant food, compared to moisture in butter. The potential consequences would require much consideration prior to the setting of fitness for purpose criterion. The Optimised Uncertainty method (described above) is another method for assessing fitness for purpose that acknowledges the financial consequences of measurement and decisions made as a result.

Finally, the ease with which the criterion can be monitored or indeed enforced deserves comment. The current approach to ensuring sampling quality involves following a sampling design prescribed for a specified analyte-commodity combination. Setting of global FFP criterion would also require sampling protocols designed to result in a sampling uncertainty that achieves compliance with the criterion. However even if the samplers follow these specially designed protocols, the estimate of sampling uncertainty may fall above the criterion value. This could occur thorough no fault of the sampler, but as a result of the degree of heterogeneity of the contamination, e.g. outliers or sporadic contamination (e.g. as observed for mycotoxins).

This does not necessarily mean that the results are not of acceptable quality. However if a limit were set on the acceptable level of sampling uncertainty allowed, these results would not be used. If the sampling process had been contracted to a company on the grounds that the sampling quality would be met, a dispute between the two parties may ensue and in such cases it would be difficult to apportion blame and financial culpability.

SUMMARY

- Estimate measurement uncertainty arising from both sampling and analysis to be made as a routine procedure.
- Consider acceptable levels of uncertainty with consideration to a cost – benefit analysis, e.g. the analyte under investigation may not have any associated health risks and so high levels of sampling uncertainty would not pose a serious risk.
- Report both a combined estimated of uncertainty (measurement uncertainty) and individual estimates of individual components of sampling and analytical uncertainty.
- Consider the combined estimate of uncertainty with individual cases of compliance assessment
- It may be that for specific analyte-foodstuff combination a specific FFP criterion can be set, for measurement uncertainty including sampling, as a function of concentration, based upon this approach.
- Setting global fitness for purpose criteria for sampling uncertainty should only be undertaken with care.

FIGURES

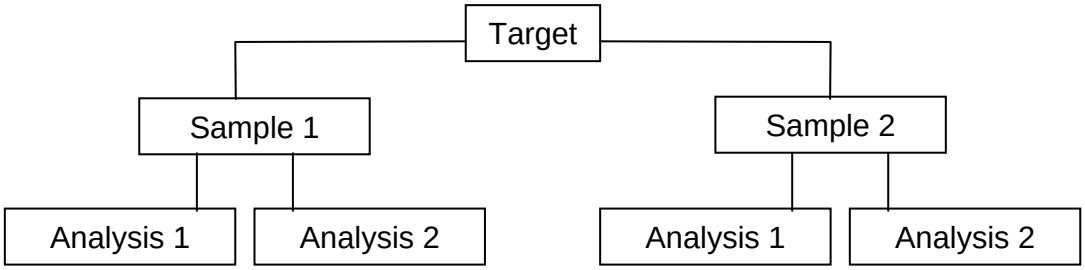

Figure 1: The duplicate method of estimating uncertainty from sampling (and chemical analysis).
S1 is taken using routine methods; S2 is taken to reflect the ambiguity in the protocol.

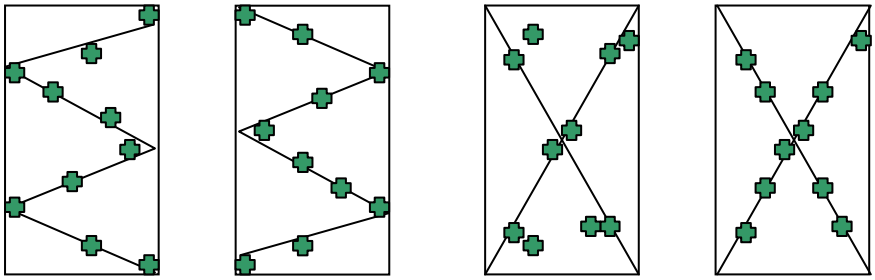

Figure 2: Three possible interpretations of the sampling protocol routinely used for measuring nitrate in lettuce. Some interpretations varied significantly from the protocol, due to practical constraints, these are not shown, but are valid within the duplicate method.

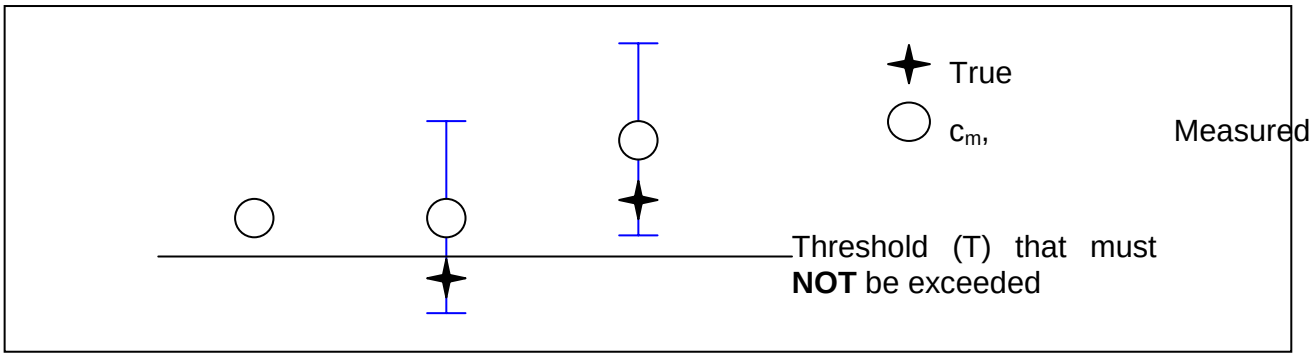

Figure 3: A batch is observed as non-compliant (c_m > T) when the analytical uncertainty (AU) is not quoted. However, if AU is known then there is doubt (i.e. [c_m – AU] < T) and a risk of a false positive result (i.e. the true value < T). However if [c_m – AU] >T then non-compliance can be proven.

Total MU (including both AU and SU). In this example [c_m – MU] is not > T.
It cannot be proven that c_m > T.
If only AU is considered an erroneous decision of non

Threshold

Figure 4: Measurement uncertainty (MU) including contributions from both analytical uncertainty (AU) and sampling uncertainty (SU), and decision making.

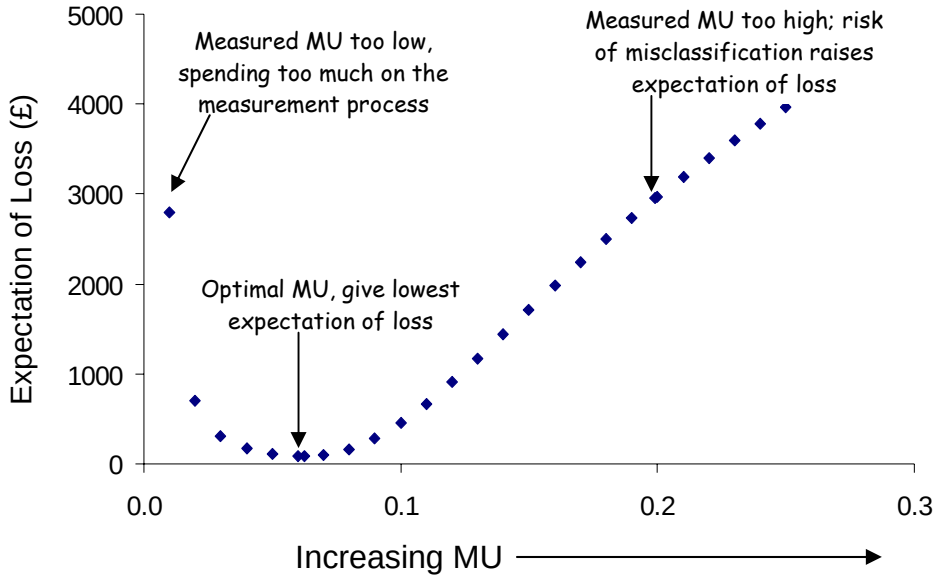

Figure 5: The results of the expectation of loss function, using a range of input MU values are shown graphically. The optimal uncertainty is identified as 0.06 units – it is this MU value that returns the lowest expectation of loss.