codex alimentarius commission

FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS

WORLD HEALTH ORGANIZATION

JOINT OFFICE: Viale delle Terme di Caracalla 00153 ROME Tel: 39 06 57051 www.codexalimentarius.net Email: codex@fao.org Facsimile: 39 06 5705 4593

Agenda Item 10 **CX/MAS 09/30/11**

JOINT FAO/WHO FOOD STANDARDS PROGRAMME

CODEX COMMITTEE ON METHODS OF ANALYSIS AND SAMPLING
Thirtieth Session
Balatonalmádi, Hungary, 9 - 13 March 2009

DISCUSSION PAPER ON SAMPLING PLANS FOR MILK PRODUCTS IN THE PRESENCE OF SIGNIFICANT MEASUREMENT ERROR

(prepared by New Zealand)

Summary

1. This paper presents a review of the CCMMP project on sampling plans[1] in the presence of *significant* measurement error. In this context *significant* means that the measurement error standard deviation is greater than 30% of the standard deviation describing the total variation of test results, due to product variation and measurement error, as set down in the Codex General Guidelines on Sampling.

2. The topic was initiated after a review by an IDF/ISO working group found that the Codex General Guidelines on Sampling could not be applied to many standards for milk products due to the presence of significant measurement error. Subsequently CCMMP asked whether it was possible to identify sampling plans which would overcome this problem.

3. A preliminary review was not able to readily identify any general methodology for suitable sampling plans when the between-laboratory component of measurement error was significant. This conclusion was reached in light of the principle that sampling plans should be developed according to the theory of acceptance sampling plans. This approach enables users to select sampling plans which allow no more risk than is required so that they can provide producers and consumers with prescribed levels of assurance relating to the quality of the product overall. This is the basis underlying the sampling plans in General Guidelines and used in general when measurement error is negligible.

4. This paper proposes that a working group should investigate the issue of sampling plans in the presence of significant measurement error more fully to determine whether recommendations can be developed, to enhance the existing Codex General Guidelines on Sampling,

[1] A sampling plan, or more correctly an Acceptance Sampling Plan (referred to as a sampling protocol in the Codex Procedural Manual) is a procedure which stipulates:

- The size and number of individual items forming the sample taken from the lot or consignment being assessed
- The way in which those samples are taken
- The analytical method used to analyse those samples
- The statistical criteria used for acceptance or rejection of the lot on the basis of those samples, using the results obtained from the testing of those samples.

Background

5. The Sixth Session of CCMMP (2004) agreed[2] to request the IDF/ISO/AOAC Working Group on Methods of Analysis and Sampling to prepare recommendations for sampling plans for milk products on the basis of the General Guidelines on Sampling, CAC/GL 50, which had recently been finalised by CCMAS.

6. The IDF/ISO Working Group's findings were documented in CX/MMP 06/7/13 Part III, and were discussed at the Seventh Session of CCMMP (2006).

7. The Working Group came to the conclusion that there were several reasons why the General Guidelines on Sampling could not immediately be applied to the assessment of compliance against CCMMP compositional standards, in particular:

   i) The presence of significant measurement error associated with the testing of many provisions in milk products means the sampling plans in the General Guidelines cannot be used. These sampling plans apply only in the situation where it is not necessary to make allowance for measurement error, i.e. when the measurement error standard deviation is less than 30% of the overall standard deviation, including both measurement error and sampling variation. Some examples are presented in Appendix 1.

      It should be borne in mind that although more precise methods might be available, this conclusion is based on comparisons against the errors of the defining or reference methods, as these are the methods referred to in the Codex standards and would be used for testing in dispute situations.

   ii) CCMMP has not defined the level of stringency required, so it is not possible to select sampling plans from the General Guidelines.

8. The IDF/ISO working group also reviewed some options for sampling plans, to explore strategies that could satisfactorily take account of significant measurement error.

9. This review followed the principles of statistical acceptance sampling theory which allows the construction of sampling plans that control consumer's and producer's risks within defined levels of protection. This is the same approach followed by the Codex General Guidelines. This basis for this principle is discussed in more detail in the section *Objectives of the Sampling Method* below.

10. The review identified procedures which could be used to develop sampling plans when bias and/or repeatability type measurement error variation were present but that a reasonable solution to cope with significant between-laboratory measurement error seemed possible only when reference samples were tested in parallel with samples from the lot under assessment, and an adjustment made for the "local" effect of this component of measurement error.

11. The extent to which this issue affects assessments in foods other than dairy products is not known, but it suggested that for example, protein testing by the Kjeldahl method could be problematic regardless of the product matrix tested, given that the reproducibility of the Kjeldahl method is known to be of the order of 1% of the concentration.

Objectives of the Sampling Method

12. This paper considers sampling methods in the context of assessing foods in international trade for compliance to specifications in Codex standards.

13. Codex procedures include several provisions relating to sampling methods for compliance assessment of foods[3]. Codex Methods of Sampling 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. The sampling protocol may include the statistical criteria to be used for acceptance or rejection of the lot on the basis of the sample. A Codex Committee should, whenever possible, provide information for each sampling plan relating to the scope or field of application, the type of sampling (e.g. bulk or unit), sample sizes, decision rules, details of plans (e.g. "Operating characteristic" curves), inferences to be made to lots or processes, levels of risk to be accepted and pertinent supportive data. The General Guidelines are intended to assist commodity committees to follow these procedures.

---

[2] ALINORM 04/27/11, para 135.

[3] *Codex Procedural Manual*, 17th edition, pages 84, 85 and 108 (English version).

14. ISO 2859[4] provides the following guidance on sampling plans:

> *The main aim of any acceptance inspection must be to see that the customer gets the quality required, while remembering that financial resources are not unlimited and that the cost of the article must reflect the cost of inspection as well as the cost of production.*
>
> *There are three possible ways of selecting items for inspection:*
>
> *a) 100% inspection, in which every item produced is examined*
> *b) Sampling based on the mathematical theory of probability*
> *c) Ad hoc sampling based on sampling theory, for example the inspection of a fixed percentage, or occasional random check.*

The standard dismisses Method a), 100% inspection, as an expensive and often impractical option, particularly when testing is destructive, and then goes on to say:

> *Method b) has the disadvantage, compared with 100% inspection, that some of the items produced will not be inspected. But the risks involved can be precisely calculated, and a plan chosen to allow no more risk that is required.*
>
> *Method c) is not recommended since it leads to uncalculated risk, and often to unjustifiably high risks; further, there is no logical basis for either the acceptance or rejection of the product.*

15. These considerations imply that sampling plans should be developed using statistical considerations (i.e. method b), so that risks of making incorrect decisions about the status of product can be managed to required levels by both producers and consumers.

16. Where significant measurement error is present, arbitrarily prescribed limits to risk may not always be achievable. However even in these cases, tabulation of schemes together with their associated risks (or maximum risks) would be of great value to users in managing and balancing risks associated with compliance testing.

Measurement Uncertainty Approach

17. CCMAS is also considering a different procedure for compliance assessment as part of the explanatory notes on measurement uncertainty[5]. This approach envisages determining compliance to specifications by taking account of Measurement Uncertainty as follows:

> *A dispute will arise when considering a Codex specification, which is a maximum value, if:*
> *• the export certificate states that the analytical result to which its associated measurement uncertainty is then added is less than the Codex specification (i.e. "x + U" < L, where x is the reported analytical result, U is the expanded uncertainty and L is the Codex specification, which is a maximum limit) and so the sample meets the Codex specification, and*
> *• the import certificate states that the analytical result after deduction of its associated measurement uncertainty is greater than the Codex specification (i.e. "x – U" > L, where x is the reported analytical result, U is the expanded uncertainty and L is the Codex specification, which is a maximum limit) and so the sample does not meet the Codex specification.*
>
> *This assumes that the laboratory at importation will deduct the measurement uncertainty, as implied in Section 5, above, of this guidance. If the value after deduction is still greater than the specification, then it may be stated, **beyond reasonable doubt**, that the sample is not compliant with the specification.*

18. This approach would appear to have several shortcomings from a statistical point of view:
- The tolerance prescribed to make reasonable allowance for measurement error appears quite inadequate. A coverage factor of k=2 would be appropriate to achieve 95% coverage if the measurement uncertainty was estimated with a large number of "degrees of freedom" but typically estimates have very few degrees of freedom. For example, estimates of reproducibility obtained

---

4 ISO 2859-0:1995, Sampling procedures for inspection by attributes -- Part 0: Introduction to the ISO 2859 attribute sampling system

5 CX/MAS 08/29/9

following the standard template with L=8 laboratories probably have only about 8 degrees of freedom. In addition, the commonly advocated correction using a coverage factor based on the Student's t-distribution is not correct either, as this provides only 95% coverage *on average*, and then only approximately, which does not seem to constitute a sufficiently high standard of proof to claim non-compliance. Recourse must be made to the theory of statistical tolerance intervals to determine the correct coverage factors (based on the non-central t-distribution for normally distributed data).

- 95% compliance does not seem consistent with the principle of "beyond reasonable doubt" which implies more like 99% compliance.
- No demonstration of the effectiveness of this proposal has been provided to allow potential users to make an informed choice about its suitability, or to reassure producers that they will not be unfairly penalised by unwarranted rejections.
- Unlike the statistical theory of acceptance sampling, where an assessment is made about the quality of the lot overall, there is an assumption that compliance of the sample means compliance of the product. This conflicts with other, commonly used attribute sampling plans in which non-zero acceptance numbers are permitted for acceptance of a lot.
- No guidance is given about the number of samples required to be taken, but it is implied that n=1 will suffice in all situations. However there seems to be no theory that would allow the number of samples to be determined on a statistical basis under the measurement uncertainty proposal to reflect the principle that there should be some pay-back for using more samples since more information about the lot is available.
- Even if exact compensation could be made for measurement error, any sampling plan based on a single result has a substantial chance of accepting a grossly non-conforming lot. For instance there would be a 50% chance of accepting a lot which contains 50% of the product outside specification. Further, it is apparent that this sort of sampling plan can control only the consumer's risk or the producer's risk (but not both), so consumers must resist the temptation to introduce offsets that might unduly penalise producers.
- Finally, the procedure always takes measurement uncertainty into account, even when it is negligible in the scheme of things. This conflicts with the criteria appearing in the Codex Guidelines on Sampling which require measurement error to be taken into account only when it exceeds of 30% of the total standard deviation. This conflict has not been satisfactorily resolved, although it is unclear whether measurement uncertainty includes or does not include sampling uncertainty.

19. The two proposals being considered by CCMAS (being the extension of the CCMMP topic and the measurement uncertainty approach) are fundamentally different. It is suggested that CCMAS should consider the appropriate way forward, taking account of statistical advice along with the usual consideration of the practicalities of compliance assessment in exporting and importing countries and cost implications.

Conclusions

20. A review found that sampling plans are needed for CCMMP at least, to cover situations when measurement error, and particular the between-laboratory component of measurement error, is significant.

21. Sampling plans need to be developed in accordance with the accepted theory of acceptance sampling plans, to enable users to select plans that allow no more than prescribed levels of risk, i.e. of controlling the chance of accepting a lot containing a certain level non-conforming product or of rejecting a lot of satisfactory quality. Where this not possible (as seems likely in the case of significant measurement error), a statistical evaluation of a sampling plan should be undertaken to determine the risks inherent in the use of that plan before it is used.

22. The choice of sampling plan appears to depend on the relative magnitude of measurement error to the total variation. There are three situations:

    i) Measurement error standard deviation is less than 30% of the overall standard deviation. In these cases sampling plans in the Codex General Guidelines are appropriate.

    ii) The ratio of the product variation/sampling error standard deviation to the total variation [standard deviation] is less than 30%. In these cases sampling plans need allow only for measurement error.

    iii) When both product variation and measurement error are significant. This case is the motivation for the project proposed by CCMMP.

23. Guidelines need to be developed for situations 2 and 3, to enhance the existing Codex Guidelines.

Recommendations

24. It is proposed that CCMAS establish a working group to investigate sampling plans for foods in the presence of significant measurement uncertainty, with a view to recommending amendments to the General Guidelines on Sampling. The working group should:

    i) Review the plans proposed in CCMMP and CCMAS.

    ii) Recommend plans based on valid statistical principles.

    iii) Evaluate, or provide tools for evaluation of, risks inherent in the plans so that users (both consumers and producers whose product might be assessed using these plans) can be aware of the risks.

    iv) Further to points (ii) and (iii), consider whether it is necessary to specify a maximum producer's risk (i.e. the maximum chance of rejection of conforming product), and make recommendations to CCMAS about the maximum chance of rejection allowable for conforming product, for example a 5% chance of rejection.

Appendix 1: Examples where Measurement Error is Significant

1. Milkfat in Dry Matter for Cheddar Cheese.

Codex Standard 263-1966 specifies that unless labelled specifically, cheddar cheese should contain a minimum of 48% milkfat in dry matter, with listed reference methods given as IDF5/ISO1735:2004, Milkfat tested by the Schmid-Bondzynski-Ratzlaff method and IDF4A:1982/ISO5534:1985, Moisture tested by gravimetry, drying at 102°C.

Data from 131 lots of manufactured cheddar of a similar specification were analysed to determine the within lot standard deviation pertaining to milkfat in dry matter. The results from the analysis carried out in Minitab ® are as follows – one outlying observation identified in a preliminary analysis was omitted:

Analysis of Variance for FDM, using Adjusted SS for Tests

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Batch	130	1669.794	1669.794	12.845	153.38	0.000
Error	1990	166.654	166.654	0.084		
Total	2120	1836.449				

Variance Components, using Adjusted SS

Source	Estimated Value
Batch	0.78860
Error	0.08375

From this analysis the estimate of the within lot standard deviation is $s_w = 0.2894$ on 1990 degrees of freedom, including repeatability type measurement error variation.

IDF5:2004 Milkfat by Schmid-Bondzynski-Ratzlaff Method

The repeatability and reproducibility values reported in the method are 0.35 pp and 0.55pp respectively.

IDF4A:1982 Moisture by Gravimetric Method after Solvent Extraction

The repeatability and reproducibility values reported in the method are 0.30 pp and 0.40pp respectively.

Assuming an average Milkfat = 36.50% and an average Moisture = 35.3% and using the approximate formula for the variance of a ratio:

$$\mathrm{var}\left(\frac{X}{Y}\right) = \left(\frac{\overline{X}}{\overline{Y}}\right)^2 \left(\frac{\mathrm{var}(X)}{\overline{X}^2} + \frac{\mathrm{var}(Y)}{\overline{Y}^2}\right)$$

the apparent repeatability of the FDM test method is:

$$\left(r(FDM)\right)^2 = \left(\frac{0.365}{1-0.353}\right)^2 \left(\left(\frac{0.35/100}{0.365}\right)^2 + \left(\frac{0.30/100}{1-0.353}\right)^2\right)$$

or r(FDM)=0.60 pp so that the repeatability standard deviation is 0.22pp.

Similarly, the apparent reproducibility of the FDM test method is:

$$(R(FDM))^2 = \left(\frac{0.365}{1-0.353}\right)^2\left(\left(\frac{0.55/100}{0.365}\right)^2 + \left(\frac{0.40/100}{1-0.353}\right)^2\right)$$

or R(FDM)=0.90 pp so that the reproducibility standard deviation is 0.33pp.

Correcting the within lot standard deviation for repeatability type measurement error produces an estimate of the net within lot process standard deviation of $s_{adj} = \sqrt{0.2894^2 - 0.22^2} = 0.188\,pp$
Clearly this is of the same order as either of the measurement error standard deviations.

2. Milkfat in Dry Matter for Cheddar Cheese II

Data from 213 lots of cheese produced by a different manufacturing plant were analysed in the same way as the example above.

Analysis of Variance for FDM, using Adjusted SS for Tests

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Batch	212	483.5880	483.5880	2.2811	9.70	0.000
Error	425	99.9400	99.9400	0.2352		
Total	637	583.5279				

Variance Components, using Adjusted SS

Source	Estimated Value
Batch	0.6830
Error	0.2352

Correcting the within lot standard deviation s=0.485 for repeatability type measurement error produces an estimate of the net within lot process standard deviation of $s_{adj} = \sqrt{0.485^2 - 0.22^2} = 0.43\,pp$
Clearly this is of the same order as either of the measurement error standard deviations.

3. Moisture in Butter

The Codex Standard A-1-1971 for butter specifies a maximum moisture content of 16% m/m for butter, tested using the IDF80/ISO37271:2001 reference method.

Data obtained from testing from 104 lots of manufactured butter were analysed using Minitab to determine the within batch standard deviation. This standard deviation was estimated as sd = 0.074pp, which will include of repeatability type error.

Analysis of Variance for Moisture

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Batch	103	4.853238	4.853238	0.047119	8.53	0.000
Error	2760	15.240278	15.240278	0.005522		
Total	2863	20.093516				

Variance Components, using Adjusted SS

Source	Estimated Value
Batch	0.00151
Error	0.00552

ISO8851-1/IDF191-1:2004 quotes the following precision statements:

Repeatability
The absolute difference between two independent single test results, obtained using the same method on identical test material in the same laboratory by the same operator using the same equipment within a short interval of time, will in not more than 5% of cases be greater than 0.31%.

Reproducibility
The absolute difference between two independent single test results, obtained using the same method on identical test material in different laboratories with different operators using different equipment, will in not more than 5% of cases be greater than 0.42%.

On this basis $\sigma_r = 0.112$ and $\sigma_R = 0.152$ so that $\sigma_L = 0.10$

The estimate of the overall within batch standard deviation $\sigma = 0.074$ is comparable to the between laboratory measurement error standard deviation. We note that the overall within batch standard deviation includes a component due to repeatability type measurement error but the overall estimate is actually smaller than the repeatability error standard deviation taken from ISO8851/IDF191.

However the laboratory performing the testing that generated the estimates of the within and between batch process variation was part of a continuing inter-laboratory proficiency study for butter testing. Analysis of this data led to an estimate of repeatability of $r=0.17$, or equivalently $\sigma_r=0.06$. However this does not alter the conclusion that the measurement error and process standard deviations are of similar magnitudes.