CODEX ALIMENTARIUS COMMISSION

Food and Agriculture Organization of the United Nations

World Health Organization

Viale delle Terme di Caracalla, 00153 Rome, Italy - Tel: (+39) 06 57051 - Fax: (+39) 06 5705 4593 - E-mail: codex@fao.org - www.codexalimentarius.net

Agenda Item 5 **CX/MAS 11/32/5**

JOINT FAO/WHO FOOD STANDARDS PROGRAMME

CODEX COMMITTEE ON METHODS OF ANALYSIS AND SAMPLING
Thirty-second Session
Budapest, Hungary, 7 - 11 March 2011

GUIDANCE ON PROCEDURES FOR CONFORMITY ASSESSMENT AND RESOLUTION OF DISPUTES

REPORT OF THE ELECTRONIC WORKING GROUP ON GUIDANCE ON PROCEDURES FOR CONFORMITY ASSESSMENT AND RESOLUTION OF DISPUTES

Background

1. At its 31st session, the CCMAS agreed to establish an electronic working group led by Brazil with the assistance of New Zealand to prepare a discussion paper that would consider procedures for conformity assessment and resolution of disputes and what further guidance was needed taking into account principles of conformity assessment, principles and guidelines for inspection and certification of food, conformity assessment procedures already developed, measurement and sampling uncertainty (collecting additional information on uncertainty of sampling as necessary), the concept of "fit for purpose", production and process control procedures, means for practical application of the *General Guidelines on Sampling*, risk of wrong decisions on compliance and non-compliance, consequences of non-conformity, resolution of disputes taking into account CAC/GL 70-2009 and the nature and sources of disputes (noting the sources of disputes mentioned in footnote 3 of this document), emerging issues, and documents and papers presented to the session.

Preliminary activities: preparation of the working environment

2. Brazil and New Zealand prepared the environment for facilitating the development of the discussion paper:

- an invitation letter to Codex members and international organizations to nominate participants and guests was circulated on 7 May 2010;
- definition of the desired functions and characteristics of a website to be the forum for discussion;
- elaboration of the proposed outline for the discussion paper (DP);
- elaboration of the proposed timeframe;
- definition of the rules of procedure for the activities of the WG;
- definition of the reference documents and the reference links for helping the elaboration of the DP.

3. New Zealand hosted, developed and managed the website. After the website was ready, a welcome message with information on how to access it was sent on 8 June 2010 to all participants and guests.

4. The names, contact details and statuses of the nominated persons were uploaded in the website. A total of 22 member countries and three international organizations nominated their participants and guests (Annex 1).

5. The timeframe was amended according to a suggestion from Japan.

6. Brazil and New Zealand monitored access to the website and issued messages to encourage participation. The last report showed that 18 of the 22 registered countries had accessed the website.

7. Working through the facilities of a website proved to be appropriate to the development of the activities and to the participation of the nominated persons.

The development of the discussion paper

8. To facilitate the elaboration of the DP, Brazil and NZ agreed to produce a proposed consensus text for each section, to be uploaded at the website. This decision aimed to facilitate the discussions to take place at the electronic forum offering a basis to start the discussions. Some sections of the DP were more complex to prepare and required more time for elaboration, leading to some delays in the progress of the activities, compared to the initial timeframe.

9. The first consensus document was presented for discussion from 14 to 25 June on Section 1 – Introduction and Section 2 - Scope. Contributions to these sections were received from Australia and supported by Netherlands and ICGMA.

On the importance of the DP for Codex and CCMAS activities

10. The DP introduces new concepts and definitions, proposes new procedures and points in the direction of revising some Codex Documents. Sections 6 and 7 on Conclusions and Recommendations and Future Work present the issues that indicate also the value of the DP for the development of Codex and CCMAS activities. The sections of the DP have covered all the points requested by CCMAS as indicated in the first paragraph of the report.

Conclusion

11. CCMAS is therefore referred to the DP (Annex 2), and particularly to the conclusions, recommendations and future work. The amount of proposed future work is substantial. The committee will need to prioritize the work and consider how it should be undertaken. This could be an important point for a future CCMAS agenda.

12. Brazil and New Zealand, based on the above report, recommend that CCMAS considers the contents of this DP as new work and asks CAC for the approval of a project document (Annex 3) where the reasoning is presented to base this decision. The project document relates to recommendation 12 of the DP. This work will make the earliest possible use of the material prepared by the WG; and is a first step in implementing the recommendations. The elaboration procedure will provide opportunity for Codex members to improve the contents of this document as it progresses through the Codex step procedure.

13. Finally, Brazil and New Zealand are grateful for all Codex members and international organizations that followed the discussions and contributed to collaboratively preparing this DP.

Attachments

Annex 1: List of participants
Annex 2: Discussion Paper on Conformity Assessment, Based on Test Results of Foods in Trade, and Implications for Resolution of Disputes
Annex 3: Project document for new work on the development of "Principles and Guidelines for Conformity Assessment Based on Test Results of Foods in Trade and Resolution of Disputes"

Annex 1

List of Participants

Argentina

Verónica Torres Leedham
Director - Laboratory and Technical Control Bureau
National Agrifood Health and Quality Service
Email: vtorres@senasa.gov.ar

Nora Angelini
Laboratory and Technical Control Bureau
National Agrifood Health and Quality Service
Email: nangelin@senasa.gov.ar;

Armenia

Sofik Khachatryan (Mrs)
Leading specialist of the Standardization Department
National Institute of Standards of Armenia
tel.: (37410) 23-47-78, (37477) 09-39-96
Email: sof_Khachatryan@mail.ru

Gayane Margaryan (Mrs)
Advanced Specialist in Innovation, Standardization and Metrology Division
Ministry of Economy of the Republic of Armenia
Tel: 0037410 589324
Email: erkga@mail.ru

Australia

Richard Coghlan
Laboratory Services Manager
National Measurement Institute
Department of Innovation, Industry, Science and Research
Email: richard.coghlan@measurement.gov.au

Codex Contact Point Australia - Codex Australia
Department of Agriculture, Fisheries and Forestry
Email: codex.contact@daff.gov.au

Ms Karina Budd
National Residue Survey,
Department of Agriculture, Fisheries and Forestry
Email: karina.budd@daff.gov.au

Brazil

Hoeck Miranda
Public Health Laboratories General Management
Brazilian National Health Surveillance Agency – ANVISA
Email: hoeck.miranda@anvisa.gov.br

Shirley de Mello Pereira Abrantes;
Public Health Titular Researcher;
National Institute for Quality Control in Health;
Osvaldo Cruz Foundation, Ministry of Health, Brazil
Email: shirley.abrantes@incqs.fiocruz.br

Roberto Gonçalves Junqueira
Professor Doctor (Professor Associado)
Universidade Federal de Minas Gerais (UFMG)
Faculty of Pharmacy, Food Science Department
Belo Horizonte – MG, Brazil
Email: junkeira@netuno.lcc.ufmg.br

Lina Yamachita Oliveras,
Researche Projects Manager;
The Foundation for Science and Technology – CIENTEC
State of Rio Grande do Sul
Email: lina@cientec.rs.gov.br

Marta Severo
Agropecuary Federal Inspector
Ministry of Agriculture, Livestock and Supply
Email: mpfsevero@gmail.com

Dr. Ivone Delazari
Pharmacist, PhD on Microbiology
Brazilian Association of the Food Industries
Email: idelazari@uol.com.br

Canada

Stan Bacler
National Manager, Food Chemistry Laboratory Programs
Canadian Food Inspection Agency
Tower 1 Floor 3 Room 314 1400 Merivale Rd. Ottawa ON K1A 0Y9
(613) 773-5308
stan.bacler@inspection.gc.ca

Chile

Soraya Sandoval Riquelme
Ministry of Health of Chile, Public Health Institute
Marathon 1000 Ñuñoa, Santiago
soraya@ispch.cl

Nuri Gras
Labser
nuri.gras@labser.cl

Jose Chamorro
SAG, Ministry of Agriculture
jose.chamorro@sag.gob.cl

European Commission

Mr Jérôme Lepeintre
Health and Consumers Directorate - General
Tel: --32 – 2 – 299 37 01
Email: jerome.lepeintre@ec.europa.eu

France

M. le Directeur Gérard Philippe GRIMM
Ministère de l'économie, des finances et de l'emploi
Service commun des laboratoires
(Lyon) 10 allée des Saules
69922 OULLINS
tél: +33 4 72 39 51 60
fax: +33 4 72 39 51 81
Email: gerard.grimm@scl.finances.gouv.fr;

Jean-Luc Deborde
Ministère de l'économie, des finances et de l'emploi
jean-luc.deborde@scl.finances.gouv.fr;

Germany

Dr. Gerd FRICKE
Head of Department
Federal Office of Consumer Protection and Food Safety
Mauerstr. 39-42 - 10117 Berlin
Tel.: +49 (0)30 18444 10000
Fax: +49 (0)30 18444 10009
Email: gerd.fricke@bvl.bund.de;

Dr. Carolin STACHEL
Head of Unit
Federal Office of Consumer Protection and Food Safety
Mauerstr. 39-42 - 10117 Berlin
Tel.: +49 (0) 30 18412 2388
Fax: +49 (0) 30 18412 2300
Email: carolin.stachel@bvl.bund.de

Dr. Claus WIEZOREK
Chemisches und Veterinäruntersuchungsamt MEL
Joseph-König Strasse 40
48147 Muenster
Tel.: +49 (0)251 9821.237
Fax: +49 (0)251 9821-7237
Email: claus.wiezorek@cvua-mel.de

Hungary

Prof. Dr. Habil Árpád Ambrus
Deputy Director General
Hungarian Food Safety Office
arpad.ambrus@mebih.gov.hu

Ph.D. Tamás János Szigeti
Business Development Director
Wessling Hungary Ltd
szigeti.tamas@wessling.hu

Ireland

Mr. Dermot Hayes
State Chemist, State Laboratory, Young's Cross, Celbridge, Co. Kildare.
dhayes@statelab.ie.

Japan

Dr. Yukiko YAMADA
Deputy Director-General
Food Safety and Consumer Affairs Bureau
Ministry of Agriculture, Forestry and Fisheries
1-2-1 Kasumigaseki, Chiyoda-ku, Tokyo 100-8950
Email: yukiko_yamada@nm.maff.go.jp

Dr. Takanori UKENA
Associate Director
Food Safety and Consumer Policy Division
Ministry of Agriculture, Forestry and Fisheries
1-2-1 Kasumigaseki, Chiyoda-ku, Tokyo 100-8950
Email: takanori_ukena@nm.maff.go.jp

Mr Taku OHHARA
Assistant Director
Inspection and Safety Division
Department of Food Safety
Ministry of Health, Labour and Welfare
Email: codexj@mhlw.go.jp

Dr Takahiro WATANABE
Section Chief - Division of Foods
National Institute of Health Sciences
Email: tawata@nihs.go.jp

Malaysia

Mrs. Zaiton Ariffin
Section Head- Department of Chemistry
Malaysia Ministry of Science, Technology and Innovation
Jalan Sultan, 46661 Petaling Jaya, Selangor, Malaysia
Tel: + 603 7985 3033
Fax: + 603 7985 3028
Email: zaiton@kimia.gov.my

Ms.Norrani Eksan
Senior Principal Assistant Director
Food Safety and Quality Division
Ministry of Health Malaysia
Tel : +603 8885 0780
Fax: +603 8885 0790
Email :norrani@moh.gov.my
with a copy to : ccp_malaysia@moh.gov.my

Mexico

Jorge Antonio López Zárate
Ministry of Economy.
jalopezz@economia.gob.mx
codexmex1@economia.gob.mx

The Netherlands

Mr Henk van der Schee
Senior Surveillance Officer
Food and Consumer Product Safety Authority
Hoogte Kadijk 401, 1018BK Amsterdam
Tel.:+31 205 344 702
Fax:+31 205 344 700
henk.van.der.schee@vwa.nl

Dr. Saskia van Ruth
RIKILT, Wageningen UR - P.O. Box 230, Wageningen
Tel.:+31-317-480250
Fax:+31-317-417717
saskia.vanRuth@wur.nl

New Zealand

Phillip Fawcet
Principal Adviser (International Standards)
New Zealand Food Safety Authority
PO Box 2835, Wellington
DDI: +64-4-894 2656 MOB: +64-29-894 2656 FAX: +64-4-894 2675
phil.fawcet@nzfsa.govt.nz

Paul Dansted
Principal Adviser (Chemicals), Science, NZFSA,
PO Box 2835, Wellington
Ph: 04 894 2536
Fax: 04 894 2530 - Cell: 029 894 2536
paul.dansted@nzfsa.govt.nz

Michael Clear (Mike)
Senior Programme Manager - Export Standards NZFSA
PO Box 2835, Wellington
Ph: 04 894 2619
Fax: 04 894 2463 - Cell: 029 894 2619

mike.clear@nzfsa.govt.nz

Roger Kissling
Statistician
Fonterra Co-operative Group Ltd
Roger.Kissling@fonterra.com

John Jowett
Statistician
Consultant to NZFSA
JowettJ@xtra.co.nz

Panamá

Mgter. Leticia G. de Núñez
Jefa de la Sección de Alimentos y Bebidas
Instituto Especializado de Análisis
Universidad de Panamá
letician@ancon.up.ac.pa
li_nunez@hotmail.com

The Philippines

Amelia W' Teiada, Ph D
Chair -Sub Committee on Methods of Analysis and Sampling - National Codex Organization, Philippines
Fax:+ 632 838 4692
Tel: + 6328384715
awtejada@yahoo.com;

Ms. Maribel G. Marges
Philippines Codex Contact Point
Office of the Undersecretary for Policy and Planning, Project Development and R&D
3/F Department of Agriculture
Elliptical Road, Diliman, Quezon City, Philippines
(F) 632 928 0590
(T) 632 926 7439
Email: pilipinascodex@gmail.com

Spain

Dr. Pedro A. Burdaspal Pérez
Subdirección General de Coordinación Científica -Jefe de Área
E-mail: pburdaspal@msps.es
Fax: 915097913

Agustin Pons Carlos-Roca
Director Técnico del CICC
Centro de Investigación y Control de la Calidad (CICC).
agustin.pons@consumo-inc.es

Switzerland

Mr. Gérard Gremaud
Scientific adviser
Compliance with foodstuffs legislation section
Swiss Federal Office of public health
gerard.gremaud@bag.admin.ch

Erik Konings Ph.D.
Method Management Group - Quality and Safety Department - Nestlé Research Center
PO Box 44, Vers-chez-les-Blanc, CH-1000 Lausanne 26
Phone: +41 21 785 8232 Cell: +41 79 279 6884
erik.konings@rdls.nestle.com

United States of America

Dr. Gregory Noonan
Research Chemist
Division of Analytical Chemistry
Center for Food Safety and Applied Nutrition, FDA
5100 Paint Branch Parkway
College Park, Maryland 20740
Gregory.noonan@fda.hhs.gov

Larry Freese
Mathematical Statistician
USDA/GIPSA/TSD
10383 N. Ambassador Drive
Kansas City, Missouri 64153
(816) 891-0453
Larry.D.Freese@usda.gov

Zambia

Andrew Chipongo
Senior Laboratory Analyst
Zambia Bureau of Standards,
P. O. Box 50259,
LUSAKA
Tel: +260 211 231385
Fax: +260 211 238483
Cell: +260 979 180 368
anchipongo@ymail.com
Andrewc@zabs.org.zm

ICGMA

Shannon Cole, MS, PMP
Director of Science Operations, Chemistry/Biochemistry
Grocery Manufacturers Association (GMA)
1350 I Street NW, Suite 300 - Washington, DC 20005
(202) 639-5979 phone
(202) 639-5991 fax
(202) 294-6124 mobile
scole@gmaonline.org

ICUMSA

Dr. Roger Wood
Fir Tree Lodge 65 Colney Lane
Cringleford, Norwich NR4 7RG, United Kingdom
Tel: +44/16-03255000
Fax: +44/16-03507723
roger.shirley@btinternet.com

IDF

Ms. Aurélie Dubois
Standards Officer - International Dairy Federation
Diamant Building, Boulevard Auguste Reyers, 80
1030 Brussels, Belgium
Tel.: +32 2 706 86 45
Fax: +32 2 733 04 13
Email: ADubois@fil-idf.org

Dr. Jaap Evers
Senior Regulatory Strategist
FIL-IDF New Zealand c/o Fonterra Co-operative Group
Private Bag 11 029 - Palmerston North , New Zealand
Tel.: +64 6 350 46 13
Fax: +64 6 350 4676
Email: jaap.evers@fonterra.com;

DISCUSSION PAPER ON CONFORMITY ASSESSMENT, BASED ON TEST RESULTS OF FOODS IN TRADE, AND IMPLICATIONS FOR RESOLUTION OF DISPUTES

PREPARED BY BRAZIL AND NEW ZEALAND WITH THE ASSISTANCE OF A WORKING GROUP[1]

1 Introduction

The purpose of this discussion paper is to consider what guidance is needed to ensure that the procedures for determining conformity to specifications and for resolving disputes are fair.

Food products moving in trade are often accepted or rejected on the basis of their conformity to a specification established by a competent authority. There are many types of such specifications, including specifications that define the identity of a food (such as composition or colour), specifications that indicate good practice (such as maximum levels of pesticide residues or veterinary drugs), microbiological specifications, and specifications relating to label claims. Codex standards also include many such specifications.

Disputes might arise from several causes: differences due to sampling; differences in the analytical procedures; and differences in the interpretation of test results concerning the compliance of the product with specifications.

Particular problems may arise in certain areas, and these have become more prominent in recent times. A more complex situation exists when a product is not stable during transportation and storage and its quality can change between the port of origin and the port of destination. This situation can also lead to a non-homogeneous product after the transportation period. In the area of contaminants, including certain pesticides, disputes may be more likely to occur at the import stage in face of the stability of the food product

[1] Working Group: Argentina, Armenia, Australia, Canada, Chile, European Commission, France, Germany, Hungary, Ireland, Japan, Malaysia, México, Panamá, Spain, Switzerland, The Netherlands, The Philippines, United States of America, Zambia, ICGMA, ICUMSA, IDF

during the transportation period. These situations deserve an appropriate approach towards sampling and analysis to determine the conformity of the food product.

Due to the specific characteristics of microbiological contamination, microbiological analysis has been excluded from the scope of the guidelines on settling disputes, but this and other areas could be addressed in future in additional annexes that could cover specific areas of food analysis.

Document CAC/GL 70-2009, *Guidelines for Settling Disputes over Analytical (Test) Results*, concentrates only on the settlement of differences due to analytical procedures. One important aspect of this document is that its application depends on the agreement in advance between the parties to utilize the procedure for settling a dispute.

2 Scope

This discussion paper considers procedures for conformity assessment and resolution of disputes and what further guidance is needed besides those given by the document CAC/GL 70-2009. It therefore considers firstly the determination, by sampling and testing, of the conformity of consignments of food in trade to official specifications.

The discussion paper takes into consideration, but is not limited to, existing Codex documents on conformity assessment and related subjects such as sampling and analysis and their associated uncertainties, and also documents prepared by international organizations that deal with these subjects.

The document goes on to consider guidance for governments on procedures to resolve disputes which arise between food control authorities about the implications of product test results on the status of a food consignment. This part of the document provides general information on the approaches and perspectives for collaboratively resolving disputes. It discusses the relationship to conformity assessment, it is complementary to document CAC/GL-70-2009, and it aims to avoid time-consuming and costly litigation.

3 Definitions

Conformity Assessment Procedure

Any procedure used, directly or indirectly, to determine that relevant requirements in technical regulations or standards are fulfilled.

Ref: WTO-TBT-Agreement.

> Note: This definition could embrace a wide variety of procedures, regulations and standards. For the purposes of this discussion paper, conformity assessment procedures are limited to sampling and testing activities leading to test results used to determine whether relevant provisions in official standards for food are fulfilled.

Dispute

A situation of disagreement between countries concerning a conformity assessment by an importing country of a consignment or lot of product in international trade. The situation is initiated when the aggrieved country presents a formal appeal to the importing country, explaining its grounds for disagreement and requesting the establishment of a dispute resolution process, and ends when a consensus is reached about the outcomes of the dispute resolution process.

Perishable food

Perishable foods are those foods that lose normal characteristics and are prone to deterioration if they are not properly stored, refrigerated and handled within a short time. Examples of perishable foods include fish, fruit, milk and all seafoods.

Ref: http://www.dhs.vic.gov.au/health/foodsafety/downloads/unrefrig_jan05.pdf

4 Part A: Conformity Assessment (CA)

4.1 PROCEDURES FOR CONFORMITY ASSESSMENT

4.1.1 Overview

The objective of conformity assessment is to determine whether food meets specified requirements. Ideally, conforming product would always be cleared and non-conforming product would always be captured, but this is rarely possible in the real world. In practice there will always be some uncertainty when deciding on whether a product conforms with a defined standard.

The objective could be restated as controlling both the risk of an incorrect conclusion that conforming product is determined as non-conforming (producers' risk) and the risk of an incorrect conclusion that non-conforming product is actually found to conform with the requirements (consumers' risk).

The producers' and consumers' risks can only be controlled through the use of statistical methods based on the theory of probability. There are several alternative approaches but none are satisfactory. 100% testing is costly, it may be impractical (especially when destructive analysis is necessary), and is likely to result in poor quality for the consumer. Another alternative is *ad hoc* sampling, for example making occasional random checks, or following no set sampling plan. These approaches offer no means by which the appropriate probabilities of acceptance can be calculated. A policy of testing a fixed percentage of product in a lot allows the capacity for discrimination of the sampling plan to vary considerably, according to the size of the lot. In either case there is no logical basis for accepting or rejecting product. Single result assessments allow either the producers' risk or the consumers' risk to be controlled, but not both.

A consequence of the statistical approach is that a risk manager needs to make a decision on the levels to which risks are to be controlled, in order to devise an appropriate sampling plan.

There are various factors that contribute to the risk of reaching an incorrect conclusion, such as sampling error and analytical measurement error. There are also other sources of error outside the scope of this discussion, such as human error and systems failure.

Where a need has been identified to determine fulfilment of specified product requirements (i.e. conformity assessment), ISO/IEC 17000, *Conformity assessment – Vocabulary and general principles*, sets out a functional approach involving the basic steps of:

- Selection,
- Determination,
- Review and Attestation, and
- Surveillance, where required.

The following sections consider each of these steps in turn.

Selection

When selecting samples a number of factors should be considered.

a) Specification of sampling requirements

The standard may specify how and where samples should be taken, or appropriate guidance may be available in published standards or published literature. Preference should be given to standards and guidelines produced by international organizations occupying themselves with the elaboration of these standards and guidelines. If necessary, appropriate advice should be sought on whether the product specification is amenable to determination by sampling and testing, and if so, what procedures to follow.

b) Selection of the samples to be assessed

This is how and where to sample. This would be dependant on the distribution of the characteristic of interest and the nature of the product. There are recommended procedures for sampling bulk materials of specified types or qualities (e.g. rice, apples, milk, fruit juice, etc.) according to different levels of granularity. For example:

> Codex examples: sampling procedures for aflatoxin in peanuts and tree nuts in the General Standard for Contaminants and Toxins in Food and Feed, Codex Stan 193 (http://www.codexalimentarius.net/download/standards/17/CXS_193e.pdf); sampling procedures for pesticide residues in horticultural and animal products; and ISO 707|IDF 50:2008, Milk and milk products – Guidance on sampling.
>
> National examples: GIPSA requires 4.5kg corn, sampled in a certain way for aflatoxins - http://www.gipsa.usda.gov/GIPSA/documents/GIPSA_Documents/sampling.pdf) and
>
> International examples: 'Grain storage techniques - Evolution and trends in developing countries' FAO grain sampling guideline. (http://www.fao.org/docrep/t1838e/T1838E0l.htm#Sampling,%20equipment%20and%20methods)

c) Specification of statistical sampling techniques

Statistical sampling is always necessary for food, as it is possible and practical to select only a sample of the total population for analysis. Therefore, to assess conformity of the total population (e.g. a lot, consignment, or source), competent authorities need to make inferences, using criteria specified as part of the sampling plan, that go beyond the available data (e.g. test results) that only relate to the tested samples.

"Statistical inference involves drawing conclusions that go beyond the data and having empirical evidence for those conclusions. These conclusions have a degree of uncertainty, whether or not quantified, accounting for the variability that is unavoidable when generalizing beyond the immediate data to the population or a process" (Bakker, *et al*).

In plainer language, statistical sampling involves the drawing of conclusions about the (status of a) lot as a whole from data obtained from sampling the lot under examination and the testing of those samples. In drawing these conclusions, note that it is not possible to determine, without possibility of error, whether a lot is compliant.

Further to the above, it is important to note that the aim of the statistical method is to make an assessment of the product *as a whole*, not the individual samples, as certainly the issue is whether the product complies as a whole.

Note also that the *General Guidelines on Sampling* apply to assessments of individual lots or to series of lots originating from the same supplier, being quantities of product that are homogeneous[2]. The assessment of a consignment, being a collection of lots, should be undertaken in a stratified manner – with the possibility that rejection of an individual lot might not lead to rejection of the entire consignment, depending on the severity of the defect: to do otherwise this would commit the same sort of error as using a single sample to assess a lot.

Determination

ISO suggests several means of obtaining data on the sample, such as testing, inspection, audit and evaluation: the scope of the present document is focused on sampling and testing (i.e. applying a method of analysis to produce a result[3]). The report produced at the completion of the determination stage of conformity assessment summarises the necessary information, including laboratory performance information.

The performance of the method of measurement (the test method) may have considerable impact on the suitability of a sampling plan for a particular purpose, and may need to be considered here. Further comments on performance characteristics of test methods are in Annex 1.

Note that there are different types of assessment. A common type is a general assessment of the concentration of some product characteristic, such as fat in milk powder, throughout the lot. On the other hand when assessing for banned substances, positive identification of the substance is used to judge the conformity of the lot from which that sample was drawn.

Review and Attestation

The results of the determination (the test results) should be reviewed, resulting in a recommendation for an attestation. Persons responsible for review and attestation make a conclusion on conformity of the product by applying a decision rule with suitable risk characteristics, or assess the test results themselves, if necessary. The decision rule, or acceptance criterion, is normally part of the sampling plan (which is set before the assessment of conformity is undertaken). As part of a review anomalous results may be identified and reported back to the analyst.

---

[2] Homogeneity is defined in the *General Guidelines on Sampling*, section 2.2.10, as follows:
A lot is **homogenous** *relative to a given characteristic* if the characteristic is uniformly distributed according to a given probability law throughout the lot[8].
NOTE: A lot being homogeneous for a given characteristic does not mean that the value of the characteristic is the same throughout the lot.
A lot is **heterogeneous** relative to a given characteristic if the characteristic is **not** uniformly distributed throughout the lot. Items in a lot may be homogenous on one characteristic whilst heterogeneous on another characteristic.

[8] After checking, if necessary by an appropriate statistical test for comparison of 2 samples, i.e. a parametric test of a mean/variance of the characteristic (e.g. Aspin-Welch test) or a non parametric test of the characteristic for the proportions (e.g. Chi-square test or Kolmogorof-Smirnof test)

[3] **Result**: Set of values attributed to a measurand together with any other available relevant information. (CAC/GL 72)

It is recognised that, in some cases, it may not be possible to assess conformity in a cost-effective manner based on product testing alone, because of lack of suitable analytical methods or the practicalities, especially cost, involved in testing a suitable number of samples to achieve an acceptable level of confidence that the lot is of acceptable quality.

As a first step in considering costs, it could be proposed as a general principle that compliant product should not be subject to high risk of failure, i.e. there should not be a high producers' risk. If this is accepted, there will remain a trade-off between consumers' risk and cost of testing that needs to be resolved.

The distinction between compliant and not compliant, simply at a single analyte level, can in some cases lead to unnecessarily high costs. It may reasonably be questioned whether this approach is always appropriate in the testing of food. For instance, the analyte levels resulting from good manufacturing practice may be considerably below those at which product poses a significant risk to the consumer. Sampling schemes that require large sample sizes or accurate measurement methods, in order to penalize poor manufacturing practice by a high risk of failure, at analyte levels that do not present any threat to the consumer, may be out of place in such contexts. A solution might be to control producers' risk at one analyte level and consumers' risks at another. However this falls outside standard procedures for acceptance sampling, and the appropriate sampling methods would need to be addressed if this approach is considered.

Looking at this problem in another way, it may be that simply controlling the percentage of non-conforming product, without consideration of how far out of specification it is, may not always be appropriate.

It must also be recognized that, although product testing may expose a producer to increasing risks of rejection as product quality deteriorates, product testing will not reliably divide a mixture of conforming and non-conforming lots into the conforming and the non-conforming. One cannot "inspect quality into" product.

While it is not always possible to make statistically powerful assessments of product quality, nevertheless the need to control incorrect decisions about the conformity of a food remains, in order to ensure an acceptable level of risk and an acceptable level of quality. There may be a need to consider complementary means of determining conformity, and confidence in conformity assessments from analytical testing can sometimes be augmented by those alternative means (e.g. audit of the system producing the food). See further comments in section 4.2.1 below.

The attestation may be in the form of a statement of conformity, or an official certificate attesting compliance to a standard or a clearance to market or export.

There may be a separate decision on the measures to be taken as a consequence of the review assessment. Appropriate actions on non-conforming product are discussed in section 4.3 below. Such actions may include resolution of a dispute, which is discussed in section 5 below.

Surveillance

Conformance activities are used to allow or prevent product entering the market. Surveillance, on the other hand, is an activity which monitors a characteristic of the product (such as the mean level) over a longer time frame. Surveillance could be, for instance, post-market monitoring, including both imported and domestic product, and traceable back to origin. The information might be used to provide confirmation of the continuing low levels of a contaminant known from risk assessment not to endanger health. Since surveillance is not a conformity assessment, it is not appropriate to take action based only on surveillance activities to withdraw or recall product not in compliance (unless there are health risks).

4.1.2 Existing guidance on conformity assessment

This section provides a survey of guidance on conformity assessment that is available from Codex and other sources, as a prelude to considering what further guidance is needed.

Procedures for conformity assessment based on product testing are available from a number of sources:

- The Codex Procedural Manual.
- Documents already adopted. In particular documents developed by CCFICS contain procedures applying to conformity assessment based on product testing, and a number of committees have developed procedures for specific purposes.
- Conclusions from CCMAS 31.
- The WTO Agreement on Technical Barriers to Trade (TBT Agreement). Conformity assessment forms a major part of the Agreement, and includes points that could be considered as procedures. The conformity

assessment provisions of the Agreement also apply to sanitary provisions covered by the Agreement on the Application of Sanitary and Phytosanitary Measures (SPS Agreement). Not all Codex members are signatories to these agreements, but nevertheless, because Codex documents are "international standards" for the purposes of these Agreements, it would be helpful for Codex documents to align where possible.

The following procedures are derived from these sources.

General procedures

The *Principles for Food Import and Export Inspection and Certification* (GL 20) includes many procedures for inspection. The definition of "inspection" includes conformity assessment based on product testing. The definition reads: "the examination of food or systems for control of food, raw materials, processing, and distribution including in-process and finished product testing, in order to verify that they conform to requirements."

The Principles set out a range of concepts applicable to the inspection and certification of food: food inspection and certification systems and their effectiveness, risk assessment, non-discrimination, efficiency, harmonization, equivalence, transparency, special and differential treatment, control and inspection procedures, and certification validity. All these principles therefore apply to conformity assessment based on product testing.

Several other documents develop the concepts in the Principles, and in doing so provide guidance on aspects of conformity assessment based on product testing. Relevant documents include (among others) *Guidelines for Food Import Control Systems* (GL 47), *Guidelines for the Design, Operation, Assessment and Accreditation of Food Import and Export Inspection and Certification Systems* (GL 26), *Guidelines on the Judgement of Equivalence of Sanitary Measures Associated with Food Inspection and Certification Systems* (GL 53) and *Guidelines for Design, Production, Issuance and Use of Generic Official Certificates* (GL 38).

The Procedural Manual makes a number of points relating to conformity assessment procedures:

- 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. (Page 60, 19th edition)
- Commodity committees should report to CCMAS on matters connected with provisions in commodity standards which require analytical or statistical procedure and provisions for which elaboration of specific methods of analysis or sampling are required. (Page 37, 19th edition)
- Supporting information for each sampling plan for a commodity should include: 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. (Pages 37 and 64, 19th edition)
- Before it elaborates any sampling plan, or before any sampling plan is endorsed by CCMAS, the commodity committee should indicate the basis on which the provision has been drawn up, and the relative importance of the provision and the basis for judgement of conformity. Instructions on the procedure for taking samples should be provided. The sampling protocol may include the statistical criteria for acceptance or rejection; and the procedures to be adopted in cases of dispute. (Pages 62-63, 19th edition)
- Commodity committees should state information in commodity standards on allowing for measurement uncertainty (except for direct health hazards), and on recovery correction, the units and the number of significant figures. (Pages 64-65, 19th edition)
- Risk assessment should take into account methods of analysis, sampling and inspection; and in achieving agreed outcomes, risk management should take into account relevant production, storage and handling practices used throughout the food chain including methods of analysis, sampling and inspection, and feasibility of enforcement and compliance. (Pages 88 and 90, 19th edition)

CCMAS 31 noted a number of points as follows. (This does not include points made by individual delegations.)

- Conformity assessment procedures should account for the specificities of particular types of products (Alinorm 10/33/23, para. 37). An example might be the presence of non-negligible measurement error associated with some parameters or provisions.

- Sampling plans should be based on valid statistical principles (Alinorm 10/33/23, para. 37). This seems to mean that statistically valid procedures should be used to make an assessment of the conformity of the lot as a whole, with a known or derivable relationship between the probability of acceptance of the lot and the quality level of that lot.
- When a sample is taken according to the relevant guideline its measurand content should meet the specification and consequently sampling uncertainty should not be taken into account (Alinorm 10/33/23, para. 91). This point could be questioned: one would always have to take sampling error into account, unless that component is negligible compared to measurement error. It has nothing to do with whether the sampling has been done according to a specified procedure.

The TBT Agreement has substantial sections on conformity assessment, notably Article 5, Procedures for Assessment of Conformity by Central Government Bodies, and Article 6, Recognition of Conformity Assessment by Central Government Bodies. Procedures particularly relevant to conformity assessment based on product testing include the following:

- Importing country conformity assessment procedures should not be more strict or be applied more strictly than is necessary to give adequate confidence that products conform with the applicable official standards. (TBT Agreement, 5.1.2)
- The selection of samples for conformity assessment should not be such as to cause unnecessary inconvenience [to applicants or their agents]. (TBT Agreement, 5.2.6)
- Acceptance of alternative conformity assessment procedures is conditioned on demonstration that the alternative procedures offer an assurance of conformity with the relevant provisions equivalent to those adopted (TBT Agreement, 6.1). The term "equivalent" may need interpretation. It may for instance mean a similar level of strictness, in accordance with the first bullet of this paragraph.

In addition, note that standardization organisations offer considerable information on conformity assessment. An example is ISO CASCO, which produces a range of standards and guides.

Conformity assessment procedures already adopted by Codex

Codex has adopted conformity assessment procedures, or elements of them, for various types of contaminants and for types of food, as well as general procedures. The following sections provide an outline.

General Guidelines on Sampling (GL 50)

The *General Guidelines on Sampling* (GGS) provides guidance on sampling and sets out sampling plans intended for use by Codex commodity committees or, if applicable, by governments in case of international trade disputes.

The GGS is designed to be used when food is being controlled at reception by statistical inspections for compliance with a particular Codex commodity standard. In this context, "reception" seems to mean the point of import (port of entry). They are intended to guarantee an acceptable quality level.

The GGS states on the one hand that producers and/or traders can use the guidelines for self-inspection; but also states they may not be applicable for control of end-products and for process control during production.

The GGS covers the following sampling situations:

- the control of only homogeneous goods (with exceptions noted in the next set of bullets);
- control of percentage of defective items using inspection by attributes or inspection by variables, for goods in bulk or in individual items;
- control of a mean content.

The GGS does not cover:

- the control of non-homogeneous goods (see further comments on this point in Annex 2);
- the control of homogeneous goods in cases where measurement error is not negligible compared to sampling error; and
- double, multiple and sequential sampling plans, deemed too complex in the frame of the GGS.

Pesticide residues

The *Recommended Methods of Sampling for the Determination of Pesticide Residues for Compliance with MRLs* (GL 33) describe sampling procedures to enable a representative sample to be obtained from a lot, for analysis to determine compliance with Codex MRLs for pesticides. MRLs for meat and poultry apply to a bulk sample derived from a single primary sample, whereas MRLs for plant products, eggs and dairy products apply to a composite bulk sample derived from 1-10 primary samples.

GL 33 recommends the number of primary samples to be taken from a lot, and the size and nature of the samples, and sets out criteria for determining compliance.

Residues of veterinary drugs

The *Guidelines for the Design and Implementation of National Regulatory Food Safety Assurance Programme Associated with the Use of Veterinary Drugs in Food Producing Animals* (GL 71) provide guidance on the design and implementation of food safety assurance programmes for residues of veterinary drugs. They recommend verification using a combined system of inspection/audits and sampling/laboratory analysis. The frequency, point and type of activity should be based on assessment of the food safety risk. The verification programme may be chosen as appropriate from system verification, risk-targeted verification or surveys. In designing a sampling protocol it is essential to define both the purpose of the programme and the population of interest. It is also important to define the criteria to be applied when analysing the results with respect to the need/desirability for any further action, and especially how such criteria and actions directly relate to the protection of human health.

The guidelines recommend port of entry testing programmes only as a secondary system verification tool. The guidelines recommend the type of programme that is appropriate for this purpose, and actions following the detection of non-compliant results.

Contaminants

The *General Standard for Contaminants and Toxins in Food and Feed* (GSCTF, Codex Stan 193) lists the maximum levels and associated sampling plans of contaminants and natural toxicants in food and feed. Among the criteria to be considered when developing MLs are validated qualitative and quantitative data on representative samples and appropriate sampling procedures.

The GSCTF includes sampling plans for aflatoxins in peanuts and tree nuts.

Microbiological criteria

The *Principles for the Establishment and Application of Microbiological Criteria for Foods* (GL 21) include a section on sampling plans, methods and handling. The section describes the information that should be included in the sampling plan, and notes the issues to be taken into account in designing the plan. The issues include the risks to public health associated with the hazard, the heterogeneity of distribution of microorganisms where variables sampling plans are employed, the AQL, the desired statistical probability of accepting a non-conforming lot, and the administrative and economic feasibility.

Commodity standards

The Procedural Manual sets out a number of procedures, noted above, for the development of conformity assessment provisions in commodity standards.

A brief review of several commodity standards revealed a wide range of provisions for conformity assessment. Some standards do not mention sampling plans (e.g. Standard for Edible Fats and Oils not Covered by Individual Standards; Standard for Sugars), some refer to general guidance on sampling plans (e.g. the Guideline for Visual Inspection of Cans calls for statistically-based sampling plans chosen according to the nature of the inspection; standards for milk products refer to standards for inspection by attributes and inspection by variables), some refer to specific sampling plans (e.g. many standards for fish and fishery products and for processed fruits and vegetables refer to a sampling plan with an AQL of 6.5 percent and a sampling plan for determination of net weight), and some specify a sampling plan and acceptance procedure in the standard itself (e.g. the procedure for determining sodium chloride in food grade salt; procedure for defects and composition of dates).

4.1.3 Areas where conformity assessment procedures are needed

The following topics have been suggested as areas where further Codex guidance would be helpful:

- Means for practical application of the *General Guidelines on Sampling*. It has been noted that the sampling plans recommended in these guidelines are too complex for routine use, and the costs of sampling, testing and administration are too high.

  Consideration could be given to simplifying the way in which sampling plans are presented.

- On the other hand, it has been noted that alternative, simpler, sampling plans should take account of the consumers' and producers' risks, and the costs of wrong decisions.

- Consideration could be given to managing risks in the medium to long term, rather than lot by lot, and to using lot-by-lot acceptance in conjunction with other tools as part of a wider risk management strategy. (See further comments in Annex 3.)

- Sampling plans in the presence of significant measurement uncertainty. As noted above, the *General Guidelines on Sampling* does not cover the control of homogeneous goods in cases where measurement error is not negligible compared to sampling error. ISO is undertaking work that may lead to guidance in this area; New Zealand is also doing some work.

- Specificities of particular commodities. CCMMP raised this as a point that should be taken into account, and in the past the same committee referred the problem of sampling plans in the presence of significant measurement uncertainty to CCMAS, because the problem applies to many foods. There may be other commodity-specific issues that need to be considered.

- Guidance might be needed on establishing levels of acceptable risk. The discussion may need to develop further before we are in a position to know how to pose this question.

4.2 ISSUES RELATING TO CONFORMITY ASSESSMENT

4.2.1 Recommended actions for producers and manufacturers in the light of regulatory standards

CCMAS requested that the discussion paper should take into account production and process control procedures to achieve compliance with specifications in a more effective manner than end-product testing[4].

There are a number of possible reasons why end-product testing at the border may not be an effective means of determining conformity with specified requirements. In many situations cost is likely to prevent adequate control of consumer's risk if reliance is placed solely on the lot-by-lot assessment of incoming product, since all, or almost all, product would have to be tested, the number of samples required per lot is often substantial, and some analyses are expensive. If the analysis is time consuming and the product is non-stable or perishable, there may be high costs for delays and storage at the point of entry. An importing country, or an exporting country providing certificates to an importing country, may therefore find such testing is an undue burden and a barrier to trade. A risk analysis may indicate that the expense is not warranted, especially if the characteristic of interest is not a hazard of great public health significance. Lot-by-lot assessment of end products therefore needs to be considered along with other risk management tools as a part of an overall risk management strategy.

Such a strategy involves the effective control of food production and manufacture. For example, if a process is well controlled and documented, acceptance of product may be decided on summaries of first-party conformity assessment, and the purpose of testing may be primarily for verification of those summaries. The type of control depends on the nature of the parameters:

(a) Some parameters are best controlled using "good practices", for instance control of pesticide residues using good agricultural practice.

(b) Some parameters can be maintained in a state of statistical process control (SPC). This is a situation where the variation of some parameter about a target level is essentially random with constant mean zero and a standard deviation equal to a specified value that is characteristic of the process. If a process is in a state of SPC, that may contribute to acceptance of product; however if the process is not in control it does not necessarily mean there is a risk of product being out of specification. Furthermore, if a process is in control it does not necessarily mean that the processor is following good manufacturing practice. CCMAS should consider providing further guidance on how SPC used by an processor could be taken into account in assessing risk to an importing country.

[4] ALINORM 10/33/23, para. 98

Where compliance with a large number of requirements is required a manufacturer may well be unable to provide evidence for statistical control of some of them, particularly those where there is no reason to suppose that any violation of limits is likely or even possible.

(c) Some parameters are not maintained in a state of statistical process control but are nevertheless under statistical control. This is likely to be the case for the compositional parameters of many foods based on naturally-occurring raw materials. The criteria for such control would need to be established on a case-by-case basis using appropriate expertise, and the implementation would need to be documented and audited.

The strategy may be guided by existing Codex documents. The Procedural Manual provides working principles for risk analysis for application in the framework of Codex. (Note that in this case the risk being managed refers to the probability of adverse health effects, as distinct from consumer's risk or producer's risk.) These principles advise that when conducting risk assessment or risk management, activities throughout the food chain should be taken into account (Pages 88 and 90, 19th edition):

> ***RISK ASSESSMENT***
>
> 21. Risk assessment should take into account relevant production, storage and handling practices used throughout the food chain including traditional practices, methods of analysis, sampling and inspection and the prevalence of specific adverse health effects.
>
> ***RISK MANAGEMENT***
>
> 30. In achieving agreed outcomes, risk management should take into account relevant production, storage and handling practices used throughout the food chain including traditional practices, methods of analysis, sampling and inspection, feasibility of enforcement and compliance, and the prevalence of specific adverse health effects.

The *Principles and Guidelines for Imported Food Inspection Based on Risk* (Appendix to the *Guidelines for Food Import Control Systems* (GL 47)) makes the following recommendations for designing an imported food control programme based on risk:

> **SECTION 4 - DESIGNING AN IMPORTED FOOD INSPECTION PROGRAMME BASED ON RISK**
>
> 6. The competent authority should use relevant information to assess the level of risk associated with the imported food. This information could include, *inter alia*: …
>
> - The adequacy of processing controls in place in the exporting country as evidenced by its laws, regulations, and other policies; its infrastructure; and its ability to effectively enforce food safety requirements, as may be verified by audits and on-site visits by the competent authority of the importing country[18]…
> - The compliance history of the food with respect to the source of the food including, where available, the compliance history of … the producer and manufacturer.
>
> [18] Laboratory sampling programmes and results may provide this type of information. Audits are another way of gaining information.

This provides a sound basis for proceeding to consider production and process control procedures to achieve compliance with specifications. The question is, how should such procedures be implemented in a balanced way, maintaining fair trade, and without unnecessary burden and cost to the parties involved? The answer is to adopt cooperative efforts to solve difficulties related to the necessary improvements when implementing control and conformity assessment procedures to guarantee high quality food based on risk analysis.

Somewhat greater detail is provided in this and other Codex guidelines. Firstly, GL 47 includes detail on the possibility of recognizing the food control system in the exporting country, including the controls applied during production and manufacture:

PROVISION OF THE IMPORTING COUNTRY FOR RECOGNITION OF THE FOOD CONTROL SYSTEM APPLIED BY AN EXPORTING COUNTRY'S COMPETENT AUTHORITY

13. Food import control systems should include provisions for recognition as appropriate of the food control system applied by an exporting country's competent authority. Importing countries can recognize the food safety controls of an exporting country in a number of ways that facilitate the entry of goods, including the use of memoranda of understanding, mutual recognition agreements and equivalence agreements and unilateral recognition. Such recognition should, as appropriate, include controls applied during the production, manufacture, importation, processing, storage, and transportation of the food products, and verification of the export food control system applied.

POINT OF CONTROL

17. Control of imported food by the importing country can be conducted at one or more points including the points of :

• origin, where agreed upon with the exporting country; …

18. The importing country can recognize controls implemented by the exporting country. The application of controls by the exporting country, during production, manufacture and subsequent transit should be encouraged, with the aim of identifying and correcting problems when and where they occur, and preferably before costly recalls of food already in distribution are required.

The exchange of information leading to this recognition has to be done in a cooperative way to be effective and to lead to an agreement to both parties in trade.

There are international organizations dealing with development around the world and these organizations can be of help to take advantages of the opportunities for improvements in the conformity assessment and certification systems. There are also agreements between countries to foster developments and these agreements can be of help.

The *Guidelines for the Design, Operation, Assessment and Accreditation of Food Import and Export Inspection and Certification Systems* (GL 26) recommend that voluntary use of quality assurance by food businesses should be encouraged. Section 4, Quality Assurance (paragraphs 6-8), reads:

6. The voluntary utilization of quality assurance by food businesses should also be encouraged in order to achieve greater confidence in the quality of products obtained. If safety and/or quality assurance tools are used by food businesses, the official inspection and certification systems should take them into account in particular through the adaptation of their control methodologies.

7. Governments do, however, retain the fundamental responsibility to ensure by official inspection and certification[7] the conformity of foodstuffs to requirements.

8. The degree to which industry effectively utilizes quality assurance procedures can influence the methods and procedures by which government services verify that requirements have been met, where official authorities consider such procedures to be relevant to their requirements.

[7] For the purpose of these guidelines, "inspection and certification" means "inspection and/or certification".

The *Guidelines for the Design and Implementation of National Regulatory Food Safety Assurance Programmes Associated with the Use of Veterinary Drugs in Food Producing Animals* (GL 71) provide guidance on the design and implementation of food safety assurance programmes for residues of veterinary drugs. They recommend verification using a combined system of inspection/audits and sampling/laboratory analysis.

The *Principles for the Establishment and Application of Microbiological Criteria for Foods* (GL 21) recommend that food business operators can use microbiological criteria, which may be stricter than those used for regulatory purposes, to verify and/or validate the efficacy of their HACCP plan:

3.1.2 Application by a food business operator

In addition to checking compliance with regulatory provisions (see § 3.1.1) microbiological criteria may be applied by food business operators to formulate design requirements and to examine end-products as one of the measures to verify and/or validate the efficacy of the HACCP plan.

Such criteria will be specific for the product and the stage in the food chain at which they will apply. They may be stricter than the criteria used for regulatory purposes and should, as such, not be used for legal action.

Guidance is also available from other international organizations. For instance the ISO/UNIDO handbook, *Building Trust: The Conformity Assessment Toolbox*[5], describes systems that combine these features (e.g. testing by producer and third-party audit, with testing of outputs as a verification of the system) – see chapter 3, page 46.

The committee may wish to consider further work in order to provide practical guidance that will assist producers and manufacturers in the area of sampling and testing. Possible items of work include:

- Acceptable standards for effective control of food production and manufacture to ensure conformity upon receipt. The work could include consideration of processes in a state of statistical process control, processes under statistical control, and good practices.
- Guidance for verification of food control procedures on the basis of sampling and testing (as distinct from lot-by-lot conformity assessment); and action on non-conformances. This discussion paper covers only lot-by-lot conformity assessment; a different approach is needed for verification purposes.
- Consideration of how to improve cooperation when implementing control and conformity assessment procedures to guarantee high quality food based on risk analysis, while maintaining fair trade, and without unnecessary burden and cost to the parties involved. This should include taking advantage of procedures that are already well established, such as evaluating competence of laboratories, and transparency of procedures for conformity assessment and procedures for verification.
- Consideration of examples of programmes that are implemented nationally. A real-life example that goes some way towards addressing the problems involved could provide valuable guidance. FAO could be invited to produce such a document, covering products of animal and also vegetable origin. The main concern is focused on non-stable or perishable products like unprocessed and minimally processed foods (fruit and vegetables; meat and meat products; milk and milk products; fish and fish products; eggs and eggs products; and honey).

4.2.2 Considerations on sampling plans for non-stable or perishable food

During the 31st CCMAS meeting the electronic working group was asked by the Committee *"to prepare a discussion paper that would consider procedures for conformity assessment and resolution of disputes and what further guidance was needed taking into account emerging issues in relation to conformity assessment and resolution of disputes based on product testing"* (ALINORM 10/33/23, §98).

International trade in non-stable or perishable food is increasing in volume and adequate control based on risk analysis[6] is gaining importance as non-stable or perishable food needs special procedures from production, through transportation, storage, processing, packaging and distribution until retail sale to preserve the quality of the food.

Responsibilities throughout the entire chain from production to retail sale should be well defined, and also the system for quality evaluation and law enforcement, taking into consideration risk analysis, the appropriate level of protection to be reached, cost effectiveness and possible delays.

When food is sampled at the destination, there is a possibility that the inspected samples are not of the same quality or from the same lot as the samples inspected at the origin if analytes are unstable or if they are rapidly metabolised by an endogenous enzymatic system. If the results from the importer and exporter are statistically divergent, the cause may need to be investigated. Before proceeding to take and test further samples, the data obtained through monitoring of the entire chain from production to retail sale and those from traceability of production should be retrieved and evaluated, in order to identify the steps where the possible problems related to storage or handling have occurred and to take the appropriate decisions.

---

5 http://www.iso.org/iso/casco_building-trust.pdf

6 Codex Procedural Manual, 19th Ed. Section IV, p.85

Cooperation between the competent authorities from the exporting and the importing countries is of the utmost importance. GL 47, *Guidelines for Food Import Control Systems*, recommends the competent authority should use the information given by the exporting country related to the adequacy of processing controls in place in the exporting country as evidenced by its laws, regulations, and other policies; its infrastructure; and its ability to effectively enforce food safety requirements, as may be verified by audits and on-site visits by the competent authority of the importing country and the compliance history of the source of the food, the food, the producer, the manufacturer, the exporter, the shipper and the importer.

An appropriate analysis of these records, to identify the steps where possible problems related to storage or handling have occurred, can be cost effective and useful to avoid costly and time consuming sampling and analysis at the point of destination. This information also helps in choosing the appropriate sampling plan to confirm the occurrences mentioned.

The non-stability or perishability of these foods may create a situation where the food cannot be considered homogeneous for the purpose of sampling at the destination.

To make the selection of the appropriate sampling plan for non-stable or perishable foods, it is necessary to take into consideration the principles given in the Procedural Manual for selection of methods of sampling, as mentioned in section 4.1 above.

The GGS gives information on how to proceed for sampling of heterogeneous foods. In p.17 it is stated:

> If the lot is heterogeneous, a random sample may not be representative of the lot. In such cases, stratified sampling may be a solution. Stratified sampling consists of dividing the lot into different strata or zones, each stratum being more homogenous than the original lot. Then a random sample is drawn from each of these strata, following specified instructions which may be drafted by the Codex product committees. Each stratum can then be inspected by random sampling which usually includes from 2 to 20 items or increments per sample (see the sampling plans of ISO 2859-1 of letter-codes A to F at the inspection level II). But before sampling, it is necessary, where appropriate, to refer to the specific instructions of the Codex product committees.

Stratified sampling is a method of obtaining a more precise estimate of a population characteristic (e.g. mean analyte level, percentage defects) where it is possible that the characteristic is distributed differently in different, and identifiable, parts of the population. The population may be, for instance, a lot or a consignment. Note that stratified sampling does not in itself imply that the various strata into which the lot is divided are then to be accepted or rejected individually, depending on the samples taken within each stratum. If in fact it is desired to formulate a decision rule that involves accepting/rejecting various combinations of sub-lots (corresponding to strata) in various circumstances, the statistical characteristics of the decision rules should be evaluated, and considered in combination.

The GGS also advises that, except for economic reasons, preparing composite samples is not to be recommended given the loss of information on sample-to-sample variation due to the combination of primary samples. There are other implications of composite sampling that also need to be considered: it leads to less precise estimates of sampling variation, and increases the impact of measurement error of the repeatability type, compared to averaging an equal number of samples analysed separately.

On the other hand, if there are economic reasons for reducing the number of analyses, it has been suggested that it may be possible to use composites but compensate for the loss of information by other means. This would require further work.

Further, composite sampling can be an appropriate means of control of *average* analyte level, as distinct from control of percentage defective or percentage above a certain limit.

There are situations where stratified sampling is not possible or the accepted control or enforcement procedure considers only one primary sample to be taken from the lot or consignment. The question of whether, in such a situation, a composite sample together with information on uncertainty of sampling, is appropriate for sampling food, including non-stable or perishable food, is considered further in section 4.2.4.

From the above statements we can reach the conclusion that non-stable or perishable food can be sampled at the destination by a stratified sampling procedure to obtain more reliable results about the variability in the quality of food. Random sampling followed by compositing the samples might not give the necessary information.

Some Codex commodity standards include specific sampling plans. These sampling plans may be applicable to non-stable or perishable foods if stratification is applied. On the other hand if random sampling with

composite sample is done, the variability of the non-stable or perishable food is not adequately evaluated and can lead to wrong decisions on conformity assessment.

The GGS also gives a series of advice on the criteria for choosing an appropriate sampling plan, according to considerations such as the type of characteristic; whether the lot is considered in isolation or is part of a continuous series; whether the measurement is qualitative or quantitative; and whether determining percentage non-conforming or average content.

4.2.3 End product testing for any non-stable or perishable food after storage and transportation between the exporting and importing countries: issues relating to CA

The topic of this section is an issue that was raised during discussion at CCMAS 31. It could be regarded as an "emerging issue", a topic that is included in the terms of reference of the working group.

Many foods in international trade are non-stable or perishable. The condition of a food may change during storage and transportation between the exporting and importing countries. A common situation, already raised in 4.2.2, is that the exporting country, on the basis of its testing, may have found the food conformed to requirements, but after storage and transport the importing country may find the food does not conform.

The importing and exporting countries should consider whether the difference in CA may be attributable to contamination, deterioration or other change in the food during storage, transport or distribution. An investigation could reveal the cause and avoid a more costly investigation of sampling and testing conditions.

Codes of practice highlight the storage and transportation conditions that are needed. Possible failures can be identified using records of traceability/product tracing and records of the storage and transportation conditions, including records of prior cargoes and prior cleaning. Potential hazards during storage and transportation may be identified and controlled using the HACCP approach.

The investigation should be undertaken with care, to ensure a material difference in analytical testing is not mistakenly interpreted as being due to product deterioration, particularly if product is near its best before date or use-by date, or if there has been similar previous experience with the same product.

If there is no evidence that storage and transportation conditions are at fault, then it may be useful to investigate sampling and testing conditions as possible causes of the difference in CA. This is discussed in section 5.

It may be necessary to use a reference standard when testing is carried out, to establish that the difference in CA is caused by product deterioration and not analytical differences.

Another situation is the case where non-stable or perishable products have apparently been exposed to a source of deterioration (e.g. refrigeration failure). CA may be used to determine whether deterioration has occurred or not and hence clear or reject the product on the basis of more than one source of information.

4.2.4 How to deal with uncertainty of sampling in the CA of foods sold in bulk

The 31st Session of CCMAS discussed uncertainty of sampling in the light of a discussion paper[7]. The committee decided to discontinue further development of the paper as a separate issue, and uncertainty of sampling was included as an issue to be considered by the working group on conformity assessment and resolution of disputes.

Context of the discussion on uncertainty of sampling

CCMAS discussed uncertainty of sampling at the 30th and 31st Sessions, in the context of several relevant considerations:

- The question of whether the *Guidelines on Measurement Uncertainty* (CAC/GL 54) should be extended to cover uncertainty of sampling, and whether it is necessary to estimate it and report it;
- The recent publication of procedures for estimating uncertainty of sampling;
- The awareness that in many cases uncertainty from sampling is as large as or larger than analytical uncertainty; and
- The question of how sampling uncertainty should be taken into account when assessing conformity.

The latter point was discussed in the knowledge that, while the *General Guidelines on Sampling* (CAC/GL 50) provide sampling plans that take account of uncertainty of sampling, there are certain limitations to their

[7] ALINORM 10/33/23, paragraphs 83-98

scope and difficulties in applying them, for instance (a) in some cases commodity committees do not select specific appropriate sampling plans from the Guidelines for the provisions they establish, as required by the Procedural Manual; (b) the sampling plans are sometimes considered too complex and costly for importing country authorities to apply, so single-sample assessments may be used instead; and (c) the Guidelines generally only provide sampling plans for situations where uncertainty from sampling is the dominant uncertainty.

The committee considered that it was important to address this complex issue on a scientific basis.

The discussion that follows deals with the issues that have been raised at CCMAS and in written comments.

Foods sold in bulk

A description is needed that covers a wide variety of foods such fruits, grains, powders, oily, viscous and liquid products. Existing definitions are not fully satisfactory. "Bulk material" could be described as any product within which component units are not aggregated and for which any division of the material into units for the purposes of sampling is necessarily arbitrary.

The term "bulk" does not limit the size of the lot. For instance, certain cereals, such as wheat, barley and maize, are usually traded in large bulk, while fish is usually traded in much smaller bulk with large variability among units.

The *General Guidelines on Sampling* (GGS) make a distinction between foods sold in bulk and other foods, because the sampling plans presented there are indexed partly by lot size, expressed as number of items in the lot. This of course is impossible with bulk materials. However the distinction seems artificial because apart from cases where a substantial proportion of a lot (at least 10%) is taken in samples, the operating characteristic of a sampling plan does not depend on the number of items in a lot. Furthermore the example given and the indexing of the plans themselves make it reasonably clear that the producer's risk specification should be treated as fixed when selecting a sampling plan: within this constraint, the consumers may choose a risk specification, and thence a sampling plan according to their assessment of the importance of the lot in question.

There may also be practical differences such as those concerned with the taking of random samples, but once a certain number of samples are taken the statistical procedures and operating characteristics of the sampling plan would be the same, once "percentage of items" out of specification is re-interpreted as "percentage of product" out of specification.

Recognizing that sampling plans may be selected to fit a consumer's own assessment of risk, rendering lot size irrelevant, and that the only statistical feature of bulk materials that distinguishes them from other types of lot is the difficulty of defining the number of items in a lot, it is not obvious that there is any reason why the sampling plans recommended in the GGS for lots consisting of discrete items should not also be used for bulk materials.

Perhaps partly because of the difficulties in defining a lot size for bulk materials, the procedures suggested in GGS for bulk materials focus on control of the lot mean analyte level, rather than control of the percentage defective. The GGS discusses sampling for control of the lot mean in general in Section 4.4 and for the special case of bulk materials in Section 5. The discussion of bulk materials in Section 5 is unclear on some points. Attempts to clarify the quantity "D" (the discrimination distance) used there lead back to Section 2 where it is difficult to interpret, in view of some confusion in Section 2 about the use of the terms "producers' risk" and "consumers' risk". Clarification and fixing would be helpful.

Current Codex practice for dealing with uncertainty of sampling in the CA of foods sold in bulk

If the term "uncertainty of sampling" is understood in a general sense, as expressing the random variation in some test criterion or criteria due to the random selection of samples within a lot, it is covered in the GGS, where sampling plans are proposed to achieve stated producer's and consumer's risks taking this variation into account. However, the term "uncertainty of sampling" has been proposed for a more specific sense akin to that used in the theory of measurement uncertainty. This is discussed later in this document.

The *General Guidelines* do not cover the situation where measurement uncertainty is significant relative to sampling uncertainty – but see comments in the next section.

So-called convenience (or pragmatic, empirical) sampling plans generally ignore uncertainty of sampling and are not covered in the *General Guidelines*. A single sample or a group of samples is selected because it is readily available or convenient. Such a sample cannot be used scientifically to make generalizations about

the lot as a whole because there is no mathematical model which would enable the probabilities of the various possible outcomes of the sampling to be related to the true lot composition.

Issues relating to uncertainty of sampling

A number of issues have been raised during discussions in CCMAS on uncertainty of sampling.

- *Definition of uncertainty of sampling*

There is no clear definition of uncertainty of sampling. The *General Guidelines on Sampling* (GGS) describes the term **sampling error** as follows:

> Part of the total estimation error due to one or several of the following parameters:
>
> • the heterogeneity of the inspected characteristics,
>
> • the random nature of a sampling,
>
> • the known and acceptable characteristics of the sampling plans.

IUPAC defines **uncertainty from sampling** as follows:

> The part of the total measurement uncertainty attributable to sampling.
>
> Note. Also called sampling uncertainty
>
> > IUPAC (2005). Terminology in Soil Sampling (IUPAC Recommendations 2005), prepared for publication by De Zorzi P, Barbizzi S, Belli M, Ciceri G, Fajgelj A, Moore D, Sansone U, and Van der Perk M. *Pure and Applied Chemistry*, 77 (5), 827–841.

Neither definition is particularly clear: however the distinction seems similar to that between measurement error (a random variable with a probability distribution) and measurement uncertainty (an assessment of a plausible range for a true value). In the measurement uncertainty definition it is not clear what this "true value" actually is, but presumably it is the average analyte level in the lot.

Concern has been expressed at this interpretation. It is far from clear that questions such as controlling probabilities of wrong decisions can be addressed satisfactorily within this framework. By summarizing the various quantities and statistical calculations involved in evaluating the operating characteristics of sampling plans (as required for an acceptance sampling approach) under the heading of "sampling uncertainty" they may be incorporated into a concept of and approach to measurement uncertainty that is of questionable relevance in risk control, these same concepts and approaches then being applied to situations in which measurement uncertainty is in fact negligible and which were already well dealt with by statistical methods orientated to control of risks.

A different view has been put forward[8], that sampling and analytical uncertainties are independent and should be dealt with separately.

Rather than attempting to construct ranges within which the true values of a particular characteristic of the lot can be said to lie, in order to formulate a procedure which allows the risks of incorrect decisions to be controlled we need a measure of test error about the true values.

In this case a definition of uncertainty of sampling is not needed; but if it were decided to develop one, it should be appropriate to food sampling and analysis. A model has been proposed by Eurachem describing sampling and analytical uncertainty:

> "…a single measurement of an analyte concentration (x) on one sample (composite or single), from one particular sampling target [which may be a lot]:
>
> $$x = X_{true} + \varepsilon_{sampling} + \varepsilon_{analysis}$$
>
> where X_{true} is the true value of the analyte concentration in the sampling target, i.e. equivalent to the value of the measurand. The bias of the analytical test method is represented by b, the total error due to sampling is $\varepsilon_{sampling}$ and the total analytical error is $\varepsilon_{analysis}$..."

(It should be noted that the analytical measurement bias b does not appear in the formula, thus $\varepsilon_{analysis}$ will in general have a non-zero expectation.)

[8] Wunderli, S. Uncertainty and sampling. *Accred. Qual. Assur.,* (2003) 8:90.

The measurand being considered in this model appears to be the mean analyte level in the "sampling target." This may restrict its scope of application significantly. In some situations the mean analyte level may indeed be the most appropriate quantity to control, particularly in the presence of significant measurement error. The model does not however deal with other situations where another approach may be more appropriate, for example control of the percentage of product that does not conform to specified limits on analyte level.

If this definition is developed further a number of issues would need to be made explicit. For instance, the difference between using a single sample versus using a composite to estimate X_{true}; bias in the testing procedure, or the between- and within-run structure of measurement errors; the uncertainty of the estimates of uncertainty of sampling; and changes in uncertainty of sampling over time and between manufacturers.

- *Estimation of uncertainty of sampling*

An estimate of uncertainty of sampling is needed for conformity assessment purposes only if a single sample or single result assessment is being used in which it is not possible to obtain an estimate of the variation within the lot under examination, either implicitly (i.e. as part of the conformity assessment procedure) or explicitly. This seems to be a common situation.

An estimate of sampling uncertainty is not generally required if the sampling uncertainty is small compared to analytical measurement uncertainty. However one does not know this before any data is obtained and analysed, but the situation is very common for the analysis of substances that occur at a very low concentration, such as ppb or ppt (parts per trillion).

The *General Guidelines on Sampling* (GL 50) advise it is desirable that the sampling errors associated with any sampling plan should be quantified and minimised.

A guide to the estimation of uncertainty of sampling has been published, developed by a Eurachem/EUROLAB/CITAC/Nordtest Working Group[9]. The guide is stated to be applicable to the estimation of uncertainty from the full range of materials that are subject to analytical measurement, and to sampling by any protocol. It offers guidance on the estimation of the uncertainty of the measurement that is caused by the process of sampling, and any physical preparation of the sample prior to analysis.

The Guide presents the methods of estimating uncertainty for quantitative data and uses real case studies to exemplify each. It recommends the duplicate method, in which duplicate samples are each subject to independent physical preparation. Statistical procedures are applied to the resultant data to separate out the components of uncertainty.

In addition the Nordtest handbook *Uncertainty from sampling*[10], based on the Eurachem *Guide*, is intended to provide practical guidance for sampling planners on sampling quality assurance and uncertainty estimation.

Before this method can be recommended it needs to be made clear:

(a) under what conditions a lot, or the various parts into which it may be divided, are accepted or rejected;

(b) whether the whole lot is at risk, or only its various parts;

(c) whether estimates of variation obtained from one lot can be applied to another, and under what circumstances.

As for any conformity assessment method, some statistical evaluation of its performance characteristics would be desirable.

Depending on the answers to the above questions, it will need to be considered whether, in applying an imprecise estimate of sampling uncertainty to a potentially large number of lots, a conservative estimate (e.g. an upper confidence limit) needs to be made of the sampling uncertainty, to prevent an underestimate leading to a large number of lots being exposed to an excessive producer's risk.

It should be noted that the duplicates method deals only with (a) sampling variation and (b) measurement error of the repeatability type. While it is stated that "studies of environmental systems have shown that between-operator and between protocol-effects (which are not allowed for) are often much smaller than those caused by heterogeneity" it is not clear why the duplicate method would be expected to yield better results than random or stratified sampling in such a situation. The same tolerance for bias and between

9 http://www.eurachem.org/guides/pdf/UfS_2007.pdf

10 http://www.nordicinnovation.net/nordtestfiler/tr604.pdf

laboratory variation would be needed in either case: but neither method deals with this aspect of the problem. CCMAS could address this issue by including guidance in the *General Guidelines* on the best way to proceed.

- *Conformity assessment and uncertainty of sampling*

Alternative conformity assessment procedures have been suggested, to overcome some of the perceived disadvantages of the sampling plans in the GGS. It is sometimes thought that in many cases these require unrealistically large samples, such as those recommended in Tables 10 and 14 in the GGS for the larger lot sizes. This problem may not actually be as severe as it appears. It is partly due to the indexing of product plans to "number of items" in the lot as a measure of the lot's importance. This will not always be appropriate. Another contributing factor is the apparent emphasis on control of the percentage defective rather than control of the mean analyte level in the lot (see Annex 6 "Conformity assessment in the presence of significant measurement uncertainty"). As discussed earlier, the indexing to number of items can be superseded by a consumer's own assessment of the risks involved, provided that the producer's risk and associated quality level are maintained.

Before considering possible alternatives, we set out criteria for management of risks that the procedures should meet. This management has to take into consideration the Codex principles of protecting consumers' health and ensuring fair practices in the food trade. In regard to managing producers' and consumers' risks, in principle it is undesirable to expose compliant product to unknown and uncontrolled risks of failure – in other words, importing countries should manage their consumer's risk in a way that does not unduly penalize exporting countries. This implies a maximum producer's risk, for instance lots consisting entirely of compliant product being subjected to a maximum rejection rate of 5%.

Given that such a criterion on producer's risk is agreed, it would then be up to the control authorities to define the adequate measures to control both the producer's and consumer's risk, in a balanced way. This may include a definition of what kind of conformity assessment is deemed as appropriate, using appropriate tools like sample sizes, analytical methods or bias correction techniques. This is an area where guidance would need to be developed. Guidance would also be needed for exporting countries to ensure that exported foods will meet requirements when subject to import inspection. Codex Recommended Codes of Practice (CAC/RCP) and the Standards (Codex Stan) are helpful in this respect. Controls on food production and manufacture (as noted in section 4.2.1) are an example of where such guidance could be developed.

Turning now to a consideration of alternatives to GGS plans, we consider single sample tests. As an alternative to the procedures suggested in the GGS, conformity assessment procedures have been proposed for determining the conformity of a lot on the basis of the test result from one sample together with uncertainty information[11]. These include the Nordtest procedure discussed under bullet point "Estimation of uncertainty of sampling", and possible alternative methods. Such procedures have been proposed as a way of reducing sampling and testing costs, and may reflect the reality of import control procedures. The impact of such procedures on producer's and consumer's risks should therefore be considered, including:

- Both analytical measurement uncertainty and uncertainty of sampling.
- The effect of using a sample that may be a single item or increment, or may be a composite of several randomly selected items or increments.
- The effects of the method of estimation (including the effect of underestimation) and the uncertainty of the estimate.
- Managing producers' and consumers' risks appropriately, taking into account all possibilities of rejection to which the lot is exposed.
- The purpose of the conformity assessment.

- *Cases where the uncertainty of sampling changes during storage and transport*

[11] Note: It may be worthwhile to note that the tests of Section 5 of GGS (which is however put forward only for bulk materials), or even Section 4.4.2, can be applied for single sample assessments (which may be composite), provided that the relevant standard deviations are known and stable. Substantial investigations are however prescribed in Section 5 to assure this. To some extent, the single sample proposals seem orientated towards methods of conducting, or replacing, these investigations. However, as usual, the GGS methods are only applicable in the absence of significant measurement error.

Questions were raised, during the seminar given by Brazil at the last session of CCMAS on issues relating to the *Guidelines for Settling Disputes Over Analytical (Test) Results*, about uncertainty of sampling related to sampling of bulk materials, whose composition is not constant after harvest and during storage and transportation. The concerns of the EU are related to nuts and grains that have a microbiological load that can pose problems of quality, when they reach the importing country. This has been attributed to the non-Gaussian distribution of natural contaminations and the conditions during storage and transportation that could be unable to impede the spread of contamination on natural products.

This does not seem to be a compliance assessment issue. A lot was "compliant" when it left the exporter and "non-compliant" when it arrived at its destination. This is certainly a problem, but is one of product management rather than compliance testing, provided it is accepted that the lots concerned did it fact change.

These situations are very common around the world and indicate that uncertainty of sampling at destination is greater than at the origin. So the total uncertainty at destination is greater than the total uncertainty at origin for all non-stable or perishable products. Sampling plans that allow for this increased uncertainty should be chosen. In some situations the sampling uncertainty may have to be considered as unknown. The GGS currently lacks guidelines on how to control producer's and consumer's risk in tests for the mean analyte level when sampling variation is unknown, and CCMAS could prepare guidance to address this problem.

- *Can uncertainty of sampling be predicted?*

The variation of foods is diverse due to the origins of raw materials, varied farming and manufacturing processes, and differences between manufacturers. It therefore seems impossible to establish, either theoretically or empirically, reliable estimates of uncertainty of sampling that could be used in conformity assessments to provide fair assessments of foods from any given supplier.

One cannot take a small sample and use the estimate of uncertainty universally across all products from one manufacturer, or indeed several manufacturers. The representativeness of the sample would be open to question, and the uncertainty of such an estimate would need to be allowed for, say by the use of an upper confidence limit.

Similarly, it does not seem appropriate to apply fitness for purpose considerations to any foods by comparisons against a theoretically or empirically derived formula, as any such formula would not necessarily represent the particular conformity assessment under consideration.

The only exception seems to be the use of standard deviations calculated over the longer term in cases where the product is sufficiently stable.

4.2.5 How does the uncertainty of analytical measurement affect CA procedures?

The terms of reference for this discussion paper include taking into account measurement uncertainty. It is here referred to as analytical measurement uncertainty to clearly distinguish it from uncertainty of sampling or from combined uncertainty, both sampling and analytical.

Context of the discussion on analytical measurement uncertainty

CCMAS 31 discussed the explanatory notes of the *Guidelines on Measurement Uncertainty* (GL 54), including the section dealing with consideration of measurement uncertainty when deciding whether or not a sample meets a specification. The notes illustrated situations that can occur when a test result with associated measurement uncertainty is compared against a maximum level, demonstrating the effect measurement uncertainty could have on the apparent compliance of product and compliance decisions in general, and therefore the need to take it into account appropriately.

Current Codex practice for dealing with analytical measurement uncertainty and conformity assessment

The only adopted procedures for conformity assessment that take account of analytical measurement uncertainty are those for determining compliance of residues with MRLs.

Recommended Methods of Sampling for the Determination of Pesticide Residues for Compliance with MRLs (GL 33) sets out criteria for determining compliance. These criteria include that results must be supported by acceptable quality control data (e.g. for instrument calibration and pesticide recovery).

Where results for the bulk sample exceed the MRL, a decision that the lot is non-compliant must take into account: (i) the results obtained from one or more laboratory samples, as applicable; and (ii) the accuracy and precision of analysis, as indicated by the supporting quality control data.

The *Guidelines on Estimation of Uncertainty of Results* (GL 59) sets out procedures for estimating measurement uncertainty for pesticide residues, and provides guidance on the use of this information for determining compliance of residues with MRLs. The Committee on Pesticide Residues is currently developing an appendix to GL 59 containing further detail.

However, these procedures are reliant on the estimates of uncertainty intervals provided by laboratories. There may be some fundamental issues concerning whether these intervals are in fact the appropriate ones for use in product testing procedures. This is discussed further below.

Issues relating to conformity assessment

A number of issues have been raised during discussions in CCMAS on analytical measurement uncertainty in relation to conformity assessment.

- *Special considerations relating to producer's risk when measurement error is present*

Where measurement error is negligible, sampling plans are geared towards a 5% rejection rate at a certain low rate of non-conformity (e.g. 2.5%). In the absence of measurement error a producer could reduce his risk to zero by supplying no non-conforming product at all. When measurement error is present, as it usually is, the possibility arises that product containing nothing but conforming items will be rejected. When an analyte level of zero is required, this presents to the producer a risk which is inescapable. It is clearly necessary that this risk should be limited and there seems a good case for considering that the value decided on should be considered as a "guarantee" to the producer.

- *Purpose*

Codex has adopted the *Guidelines on Measurement Uncertainty* (GL 54), and is currently developing Explanatory Notes to the *Guidelines.* However differing views of measurement uncertainty are still being discussed in the scientific literature[12]. The issue relates to whether uncertainty is seen as an attribute of the experimental method, which is only properly characterized by experimental data, or is seen as an attribute of our 'state of knowledge' about the measurand after the measurement has been made[13].

For the purpose of conformity assessment, the first point of view seems necessary. In CA, a producer needs know that, if he produces product at a certain analyte level, then at least 95% of the time that product will pass. The consumer needs know that if product is presented at a certain other analyte level, at least 90% of the time that product will fail. (The 95% and 90%, used for the purposes of illustration, are the conventional values used in setting producer's and consumer's risks.)

These probabilities of failure at various analyte levels are in principle at least, objectively defined and open to experimental determination and verification, the only obstacle standing in the way being the provision of enough suitable test material.

GL 54 adopts the second, less relevant, point of view. So, CCMAS could develop further guidance on this issue, aiming to define the best approach for CA and resolution of disputes. (See also a similar recommendation in Annex 1.)

- *Definition of analytical measurement uncertainty*

Measurement uncertainty has been suggested as a way to deal with measurement error in conformity assessment. For this purpose, it should be considered alongside uncertainty of sampling. However the concept of "confidence" underlying the measurement uncertainty approach differs from the concept of "confidence" used in the traditional approach to product sampling which is focussed on control of producer's

[12] Hall B D (2008) Evaluating methods of calculating measurement uncertainty, Metrologia **45** L5–8
Lira I (2009) *On the meaning of coverage probabilities*, Metrologia **46** 661–618
Willink R (2010) *On the validity of methods of uncertainty evaluation* Metrologia **47** 80–89
Willink R (2010) *Probability, belief and success rate: comments on "On the meaning of coverage probabilities",* Metrologia (in press)

[13] Lira I and Wöger W (2006) *Comparison between the conventional and Bayesian approaches to evaluate measurement data*, Metrologia **43** S249–S259

and consumer's risks[14]. The two concepts are incompatible without a reappraisal of what "measurement uncertainty" is actually to mean in practical terms.

As defined, it seems that the proposal's concept of numerical confidence is a 'degree of belief', an expression of personal assurance of the people developing and applying the test procedure, and the values of u and U are a quantification of this degree of belief. It is not clear how, if at all, this degree of belief, and the values of u and U, correspond to observable phenomena (such as success rates) by which they could, even in principle, be given objective meaning and verified. (Some guidelines on the subject of measurement uncertainty are based on this philosophy[15].) Because such an evaluation would be essentially subjective, the testing and validation of a corresponding statement of uncertainty would be problematic. How exactly is the statement of uncertainty intended to perform? What does a 95% level of confidence of this type actually mean in practical terms? The *Guidelines* do not deal with these important questions.

The definition should also acknowledge the imprecision of experimentally determined estimates of measurement uncertainty. Information on imprecision is needed if the uncertainty intervals are to be used for conformity assessment purposes, to avoid substantially overestimating the coverage provided by the intervals. As a first approximation, an upper confidence limit for the relevant precision parameter could be used in place of its estimated value. Even when estimates of uncertainty are not experimentally derived, the same approach should be used. It is not a question of "what width of interval do we *estimate* would give a 95% success rate?" but "what width of interval is required *to be confident* of a success rate of 95% or better?"

The accompanying paper, provided in the Annex 4, discusses in the context of attribute sampling and classification error the effect of the uncertainty of estimates of test method performance on the control of risks in product acceptance. The conclusion there is that disregarding the uncertainty of such estimates can have a marked effect on the conformity assessments carried out using a particular sampling plan. Similar effects would be expected in general, and can indeed be shown to occur.

- *Components of analytical measurement uncertainty*

Measurement uncertainty includes various components, such as precision of the method, bias, matrix effect, and competence of the laboratory, combined into a single figure. However for conformity assessment purposes information on the separate components is needed, because the random and systematic components (such as bias) contribute differently to the risks. For instance, the size of the 'within-laboratory' component of uncertainty is needed[16] in order to characterize the 'success rate' of measurement intervals quoted by any individual laboratory and to take into account the effect of product sampling. The bias particularly needs to be taken into account in assessing probabilities of failure.

A widely used model for measurement error expresses it as the sum of a bias, a 'between-laboratories' (or between-runs, when it is more relevant to consider a single laboratory) component, and a repeatability component.

Annex 5 gives an explanation of the effects of bias, between-laboratory variation and repeatability variation.

- *Estimation of analytical measurement uncertainty*

If measurement uncertainty were to be used for conformity assessment purposes, an agreed method of estimation would be needed. The method of estimation should be transparent and technically sound, and appropriate for the conformity assessment undertaken.

Procedures referenced in GL 54 might not be suitable. Estimates are likely to be based on numbers of samples that are considerably too small to achieve the expected nominal coverage rates; the procedures do not adequately account for bias; and there is no verification procedure for the estimates.

---

[14] The difference is similar to the difference between the Bayesian (strength of the evidence) and classical (relative frequency) concepts of probability. The measurement uncertainty approach bases confidence on the Bayesian "probability" that the true value of a parameter falls within a given range; the traditional approach bases confidence on the relative-frequency "probability" with which the ranges obtained will include the true value.

[15] Joint Committee for Guides in Metrology 2008 *Evaluation of measurement data – Supplement 1 to the "Guide to the expression of uncertainty in measurement" – Propagation of distributions using a Monte Carlo method,* JCGM 101

[16] An exception might be when only a single sample is tested, including when that single sample is a composite taken from throughout the lot under examination.

The formula $U=2u$ would also need further consideration. This prescription only yields a coverage rate of 95% for the resulting intervals when u is the true standard deviation of the probability distribution of measurement errors. If some guarantee (e.g. 95% confidence) is to be given that the 95% coverage rate is met or exceeded, for this formula to apply u would have to be in effect an upper confidence limit for this true standard deviation. This would not be consistent with the normal use of the letter u which is used to denote, indiscriminately, either the true value of the standard measurement uncertainty or an estimate of this true value, rather than a confidence limit for this true value.

Uncertainty intervals would also need to be appropriate for situations where several samples are tested from the same lot, or where a single sample is analysed for more than one analyte. In these situations the probability of rejection can accumulate; to avoid this, the uncertainty intervals may have to be expanded by a significant factor. For example, for independent measurement errors with a Gaussian distribution, there is a probability of 0.05 that a single measurement error will lie outside the range $\mu \pm 1.96\sigma$, where μ and σ are the **true** mean and standard deviation of the Gaussian distribution. However, if ten measurements are made there is a probability of $1-(1-0.05)^{10} = 0.40$ of at least one of the measurement errors lying outside this range. To reduce this probability to 0.05 we need a multiplier of 2.80 rather than 1.96.

- *Conformity assessment*

In situations where measurement uncertainty is negligible relative to sampling uncertainty the sampling plans in the *General Guidelines on Sampling* are applicable for conformity assessment lot by lot, and no allowance for measurement uncertainty needs to be made.

When measurement uncertainty is significant, the exact calibration of producer's and consumer's risks for all but the simplest cases poses considerable mathematical difficulty. A possible solution is presented in Annex 6. This should be considered if CCMAS undertakes work on the development of methods for dealing with situations where both analytical and sampling uncertainty play a significant role (recommendation 4, section 7).

Sampling plans for inspection by variables in the presence of measurement error have been devised. These involve increased sample sizes to allow for the measurement uncertainty, and may be practical only when measurement uncertainty, though not negligible compared to sampling variation, can be assumed to contribute a relatively minor part to the combined uncertainty from sampling and measurement. The considerations in Annex 6 may be relevant to extending the scope of such methods. Again, this needs to be considered if the work in recommendation 4, section 7, is undertaken.

Conclusion

If measurement (or analytical) uncertainty information is to be used as a component of a procedure for conformity assessment, it is necessary to develop procedures that simultaneously address both sampling uncertainty and a suitable concept of measurement uncertainty. Uncertainty of sampling needs to be taken into account because compliance of a lot cannot be judged accurately from compliance of a single sample, when only analytical uncertainty is taken into account. This procedure will not provide consumers with any degree of protection for product in which uncertainty of sampling is material, and it could unduly penalize producers. The primary consideration is a kind of variability that can be discussed in performance-oriented and testable terms. It seems appropriate to build procedures around product sampling with potential for the addition of factors relating to measurement uncertainty. A possible approach has been mentioned in section 4.2.4.

In considering measurement uncertainty in relation to conformity assessment, clearer guidance will be needed on where measurement error does and does not need to be taken into account in the conformity assessment criterion. There is material in the General Guidelines, although that may have to be revisited as it is overly simplistic in that it does not distinguish between the various components of measurement error, or consider potential increases in risk to either producers or consumers. CCMAS could consider the preparation of further guidance on these issues.

4.2.6 How does the concept of "fitness for purpose" affect CA procedures?

The terms of reference for this discussion paper include taking into account the concept of "fit for purpose".

At its 28th session CCMAS noted that it was premature to apply fitness for purpose for regulatory purposes and agreed to postpone the consideration of the matter for the time being and to monitor the international activities currently on-going in this area[17].

Since then Codex has adopted[18], as an analytical term, a definition of fitness for purpose which reads:

> ***Fitness for purpose:*** Degree to which data produced by a measurement process enables a user to make technically and administratively correct decisions for a stated purpose.

When the concept thus defined is considered in relation to conformity assessment procedures, it could be understood as follows. The 'stated purpose' refers to the assessment of conformity of a consignment or lot of product, or a system, to a specification. The 'data produced by a measurement process' refers to the sampling report and the laboratory report (test results and measurement uncertainty information). 'Technically and administratively correct decisions' are those that satisfy technical and administrative requirements and permit the purpose to be achieved. The 'technically correct' requirements could refer to following the correct principles, statistical, scientific and other, to control acceptance of lots to prescribed levels. The 'degree' of enabling correct decisions can be seen as the level of control of wrong decisions, or the degree of avoiding misclassifying product, or the levels of consumers' risk and producers' risk.

In this context, then, fitness for purpose of an analytical method could be expressed as the effect on consumers' and producers' risks (defined in terms of the relative frequencies of the possible outcomes for lots of given quality) from the specified conformity assessment system (sampling plan and the use of that analytical method). Thus a test method and a sampling plan for a parameter specified in a commodity standard is an implied statement of the required fitness for purpose of that parameter, since the test method and sampling plan in combination imply the levels of consumers' risk and producers' risk that are deemed appropriate.

In the light of this understanding, a concept of fitness for purpose in the context of CA and resolution of disputes might be described as follows:

> An analytical method is fit for use in a given conformity assessment procedure if the method and CA procedure used in conjunction meet specified criteria for consumer's or producer's risk.

This new concept specifically related to conformity assessment, highlights the types of consideration involved and their implications and could be considered by CCMAS.

This understanding of fitness for purpose leads to several consequences. First, when considering alternative analytical test methods, the effect on fitness for purpose of the conformity assessment should be considered. Use of alternative test methods may necessitate modifications to the sampling plan, specifically the number of samples taken and the product acceptance criterion, to achieve a similar level of fitness for purpose. Codex procedures for the adoption of test methods and criteria for test methods should be considered in this light.

Another way of looking it may be to recognize that the limits being tested against, and at which the consumers' and producers' risks apply, will generally be, if uncertainty due to measurement error is present, different from their specified values. The difference will include both systematic and random variation. The range of likely variation of these "de facto" limits, and its impact on consumer and producer, need to be assessed. If this range of variation is too wide, it might not be possible to provide adequate safety to both producer and consumer at the same time, and the method would then be considered unfit for purpose.

The example attached in Annex 7 illustrates the sort of consideration that may be required.

It should be noted that one needs the components of measurement error to be able to carry out an assessment of fitness for purpose, rather than using just the reproducibility or the measurement uncertainty (as currently defined). Note also that one needs to know the uncertainties of the estimates of the components to do the job properly, as is highlighted in section 4.2.5 and its Annex.

Further, comparison of performance characteristics such as reproducibility against the Horwitz criterion or any other similar rule does not constitute a fitness for purpose assessment, because there is no consideration of the "purpose", i.e. the way in which the results are ultimately used – in a product assessment for Codex purposes. CCMAS could consider developing further guidance on the matters raised in this section.

---

17 ALINORM 05/28/23

18 *Guidelines on Analytical Terminology*, GL 72.

4.3 CONSEQUENCES OF NON-CONFORMITY

This topic is one that CCMAS requested should be included in the discussion paper[19]. The consequences of non-conformity include the immediate actions resulting from the decision of non-conformity, and the corrective action that is taken to reduce the likelihood of future non-conformances.

Appropriate action on non-conformity is important in order to protect consumers, and to treat exporting countries fairly.

Codex has developed guidance on actions that might be taken on non-conformity. The *Guidelines for Food Import Control Systems* (GL 47) suggest a range of possible decisions following sampling and analysis, as follows:

> ***DECISIONS***
>
> 27. Decision criteria (without prejudice to the application of customs procedures) should be developed that determine whether consignments are given:
>
> - acceptance;
> - entry if cleared upon inspection or verification of conformance;
> - release of non-conforming product after re-conditioning and/or corrective measures have been taken;
> - rejection notice, with redirecting product for uses other than human consumption;
> - rejection notice, with re-exportation option or return to country of export option at exporter expense;
> - rejection notice with destruction order.
>
> 28. Results of inspection and, if required, laboratory analysis, should be carefully interpreted in making decisions relating to acceptance or rejection of a consignment. The inspection system should include decision-making rules for situations where results are borderline, or sampling indicates that only some lots within the consignment comply with requirements. Procedures may include further testing and examination of previous compliance history.
>
> 29. The system should include formal means to communicate decisions regarding clearance and status of consignments.[6] There should be an appeal mechanism and/or opportunity for review of official decisions on consignments.[7] When food is rejected because it fails to meet national standards of the importing country but conforms to international standards, the option of withdrawing the rejected consignment should be considered.
>
> ***DEALING WITH EMERGENCY SITUATIONS***
>
> 30. The responsible authority should have procedures that can respond appropriately to emergency situations. This will include holding suspect product upon arrival and recall procedures for suspect product already cleared and, if relevant, rapid notification of the problem to international bodies and possible measures to take.
>
> 31. If the food control authorities in importing countries detect problems during import control of foodstuffs which they consider to be so serious as to indicate a food control emergency situation, they should inform the exporting country promptly by telecommunication.[8]
>
> ---
>
> [6] Paragraph 4 of the *Guidelines for the Exchange of Information Between Countries on Rejections of Imported Food* (CAC/GL 25-1997) should be consulted in this regard.
>
> [7] Paragraph 6 of the *Guidelines for the Exchange of Information Between Countries on Rejections of Imported Food* (CAC/GL 25-1997) should be consulted in this regard.
>
> [8] *Guidelines for the Exchange of Information in Food Control Emergency Situations (*CAC/GL 19-1995).

In cases where the importing country recognizes the food control system applied by an exporting country's competent authority, the Guidelines, paragraph 32, recommend that

---

[19] ALINORM 10/33/23, para. 98

> the importing country should establish mechanisms to accept control systems in an exporting country where these systems achieve the same level of protection required by the importing country. In this regard, the importing country should … conduct verification procedures for example, occasional random sampling and analysis of products upon arrival. (Section 5 and Annex of CAC/GL 26-1997 deal with the provision and verification of systems that provide certification for food in trade) …

In these cases the conformity assessment is part of the risk management procedures for verifying the controls applied by the exporting country, and the consequences of non-conformity should be appropriate to the relationship between importing and exporting countries.

The *Guidelines for the Design and Implementation of National Regulatory Food Safety Assurance Programme Associated with the Use of Veterinary Drugs in Food Producing Animals* (GL 71) provide detailed guidance on regulatory action in the case of non-compliant results[20], recalling that these guidelines recommend sampling and laboratory analysis as only one part of a combined system of a verification programme. A range of actions is recommended, including investigation of the non-compliances, measures that may be taken in regard to the conduct of responsible parties, to the product and to the corrective action to prevent reoccurrence.

An exporting country may disagree with the conclusion reached by the importing country and the action it has taken, and this may lead to a dispute. The causes of disputes and procedures for resolving disputes are considered in section 5.

These texts lead to questions that need further clarification, and open up areas that may need further work on practical application in the context of sampling and testing. Possible points include:

- What is meant by "Results of … laboratory analysis should be carefully interpreted in making decisions relating to acceptance or rejection of a consignment", quoted in paragraph 28 of the Guideline mentioned above?
- What decision rules are appropriate for "situations where results are borderline", quoted in the same paragraph?

  Comment on these two points:

  Paragraph 28 refers to the situation where an importing country (or an exporting country acting on its behalf) is making inferences on the conformity of consignments using test results and decision-making rules. If the decision-making rules have been laid down, it is not clear where there is room for "careful interpretation" in deciding whether to accept or reject a lot. Giving scope for "careful interpretation" could easily lead to a situation where the risks to which product was exposed were unquantifiable.

  The authority will have certain data at its disposal. It will have results from one or more samples from a lot (taken in accordance with a sampling plan) and in many cases these results will be subject to significant measurement uncertainty. Criteria for acceptance and rejection should have been set down, by Codex commodity committees, or by the authorities of the importing country, such that the risk of compliant product failing is suitably small: nonetheless, over time, as more and more compliant lots are tested, a few will fail. If non-compliant lots are also presented, a higher fraction of them will fail. Depending on the distribution of incoming product quality, failing lots could be all compliant, all non-compliant, or anywhere in between. It might even be that the only lots to come to the importing authority's attention are those that fail.

  In such a situation the only reasonable course open to an authority, on a lot-by-lot basis, is to abide by the decision rules. Otherwise there will be an unquantifiable impact on the operating characteristic of the sampling plan.

- When is it appropriate to conduct "further testing", suggested in paragraph 28? Procedures to further investigate a case of failure, e.g. by taking additional samples, should either be built into the decision rules so that their impact on producers' and consumers' risk can be ascertained, or be treated as "for information only" and not allowed to affect the decision regarding acceptance or rejection of the lot in question.

[20] See paragraphs 110 to 129 of GL 71.

- How should the information arising from non-compliances be used as a basis for further action? At this point the existence of a record of results is of the utmost importance to guide the decision process about what further action is necessary.

  The characteristics of the record of results give the basis to guide the frequency and the strategy of sampling. The importing authority may have a range of information, for instance: How many lots from a particular importer have been tested, and how many have failed? Have the failures occurred over a fairly short time frame? Have there been other failures from unrelated suppliers within this time frame? Such information should not be used to over-ride the decision rules in respect of particular lots but to suggest the need for increased or reduced inspection levels, re-auditing of the supplier or analysing laboratory, and so on, in accordance with a risk management strategy as discussed in Sections 4.1 and 4.2.

  The exchange of information with the exporter and the competent authorities in the exporting country will be of value to for improvements along the production, transportation, storage and distribution chain. It is recommended that the exchange of information be done according to the *Guidelines for the Exchange of Information Between Countries on Rejections of Imported Food* (CAC/GL 25-1997), especially the contents mentioned in the paragraphs 14 to 17 related to 'reasons for rejection'.

- How should test results be used as part of a wider verification programme, as envisaged in paragraph 32 quoted above?

  The purpose in this case is to verify by testing and other means that the exporting country is fulfilling agreed requirements. An importing country may wish to assess the overall performance of a sector or a product class e.g. meat, dairy, fish or honey, or a particular exporter.

  What documentation and data are required, and how are they assembled to build a picture over time of the adherence to the desired quality? Is "occasional random sampling and analysis of products upon arrival" appropriate? What action is appropriate for non-conformity, bearing in mind the history of records? These points are discussed further in section 5.2, regarding the dispute resolution process, and section 5.4 regarding the risks of wrong decisions.

- What action is appropriate on a "non-conformity" that has been found when conformity assessment is conducted outside established rules, for instance by testing a single sample from a lot? This point is discussed further in section 5.2, regarding the dispute resolution process, and section 5.4 regarding the risks of wrong decisions.

- When is it appropriate to assess conformity on a lot-by-lot or consignment-by-consignment basis? Are there specific hazards for which this should apply? This point is discussed further in section 5.2, regarding the dispute resolution process, and section 5.4 regarding the risks of wrong decisions.

5 Part B: Dispute Resolution Related to CA

Introduction

In the ideal situation trading partners would have all the necessary data in hand, obtained through the application of the best control procedures and methods of analysis to base the decisions to be made using valid CA procedures.

The limitations we have in the real world give rise to the possibility of differences in the results of CA from the exporting and importing countries' CA according to the adopted procedures.

The next Sections of the DP aim to offer guidance on how to resolve the disputes which take place related to CA, taking into consideration the procedures from the approved document CAC/GL 70-2009, the recommendations arising from the 31st Session of CCMAS and the content of the previous Sections of the DP.

The CCMAS context concerning dispute resolution

During the 31st Session of CCMAS, the Committee agreed to establish a working group to prepare a discussion paper on procedures for conformity assessment and resolution of disputes, taking into account measurement uncertainty, sampling uncertainty and other relevant issues (Alinorm 10/33/23, para. 98). The Committee considered also the terms of reference for the work to be done by the working group which included the following items: a) nature and sources of disputes noting the sources of disputes mentioned in footnote 3 of GL 70 and b) resolution of disputes taking into account GL 70.

5.1 CAUSES OF DISPUTES

A dispute may take place when the analytical results of the importing country and the analytical results of the exporting country are divergent concerning the conformity assessment of a lot or consignment related to the specifications.

A dispute would normally be expected to be based on the inherent unlikeliness of failure in view of the control measures adopted by the exporting country. The product was assessed by the exporter as unlikely to fail through the application of an appropriate method of testing according to a prescribed procedure by related Codex standard and guidelines or other reference method recognized by an international organization, but did in fact fail. There are many possibilities that may need to be investigated. For example it could be that the analytical methods were not appropriately selected and run, the differences in the analytical performances of the methods were greater than expected for the two laboratories, the analytical results were wrong, that the sampling plan used was unspecified, or that the method of taking samples, or the transportation or storage of samples taken was unsuitable to the product.

In the 3rd paragraph of Section 4.1.1 of the DP we noted that the producers' and consumers' risks can only be controlled satisfactorily through the use of statistical methods based on the theory of probability. While several other approaches have been proposed, none is satisfactory because these approaches offer no means by which the appropriate probabilities of acceptance can be calculated.

On the other hand there are limitations in the application of procedures for inspection, sampling and analysis due to complexity of operations involved, the cost, the shelf-life and the time needed to obtain the results.

Concerning this point, the penultimate paragraph of section 4.2.2 informs us in connection with the discussion on non-stable or perishable food that if random sampling in combination with compositing is done, the variability of the food is not adequately evaluated due to heterogeneity and can lead to wrong decisions on conformity assessment and this may cause a dispute to take place.

Section 4.2.4 of the DP presents a set of explanations on uncertainty of sampling and issues involved in developing a satisfactory approach to the control of wrong decisions.

In the face of practical limitations, decisions on conformity assessment have to be made based on limited data and this offers a certain degree of uncertainty that must be taken into account to control the producer's and the consumer's risks of wrong decisions.

The aforementioned facts contribute to the generation of disputes on conformity assessment of lots and consignments in international trade and they have to be resolved in a case-by-case basis.

Considering the necessary measures to be taken for obtaining a fair dispute resolution, the risk managers from the importing and exporting countries have to consider the data from the entire chain from production to retail sale to verify how possible causes of dispute occurred and how to improve the control procedures to prevent the occurrence of mentioned causes.

The mentioned causes of disputes can occur in different combinations, affecting the probability of acceptance or rejection of a lot or consignment.

Sections 5.1.1 to 5.1.3 describe situations that may occur in isolation and may give rise to a dispute. If we consider that these situations may also occur in different combinations we can realize that there is always a chance for a situation of dispute to take place during the international trade of food. This emphasizes the need to keep these situations under strict control, from production to retail sale, to guarantee the success of international trade. To reach this goal a cooperative effort has to be made by the parties in trade to prevent, identify and to solve the problems that may occur during the entire food chain.

5.1.1 Disputes Covered by the *Guidelines for Settling Disputes Over Analytical (Test) Results* (CAC/GL 70-2009)

The scope of the document presents, through the two paragraphs indicated below, a picture of situations in which the provisions of this document apply.

The scope states:

> These guidelines provide guidance to governments on the procedures to resolve disputes which arise between food control authorities about the status of a food consignment, when the assessment based on test results made in the importing country disagrees with the assessment made by the exporting country on the same lot.

> These guidelines only address disputes related to methods of analysis or laboratory performance and do not address questions of sampling. The procedure examines only the validity of the importing country's results on which non-compliance is alleged.

These statements from the scope of the document indicate that it covers a limited set of causes of disputes and aims at the validation of the results obtained by the laboratory of the importing country.

While this is a very limited scope, it may well cover most of the situations in which a dispute may not be resolved in advance. For instance, whether a sampling protocol is appropriate to the product being tested, whether the sampling plan used gives an unduly high producer's risk, whether the method used to estimate "uncertainty intervals" is statistically valid and so on, could reasonably be expected to be resolved in advance before product is tested, either by agreement or by reference to the standards and guidelines produced by Codex commodity committees.

The end of the first paragraph of the Scope limits the resolution of dispute to the assessments made on the same lot, and the procedure requires comparison of results from split samples. This is to allow for the fact that the importing and exporting countries' samples may in fact be significantly different in analyte level, particularly where a sample classified as non-compliant has been found in a lot which an exporting country has reason to believe compliant. For some products, there may even be a possibility of a change in true analyte levels between the two assessments.

It should also be noted that GL 70 does not deal with the conditions under which a dispute may be raised. In general an exporting country may have strong evidence of a history of compliance which may need to be taken into account in this context. Considerations such as those expressed in GL 47 and discussed below may need to be taken into account in determining any limitation on the rights of an exporting country to appeal against the measurements obtained by the importing country using GL 70. These considerations have to be taken into account in order to ensure fair trade practices.

The present DP aims to deal with dispute situations that were not considered by the document CAC/GL 70-2009, as noted in the terms of reference, and will be considered during the next Session of the CCMAS.

The document CAC/GL 47-2003(revision 2006) presents in its para. 13 provisions for the recognition of food control systems applied by an exporting country's competent authority. It recommends that "Food import control systems should include provisions for recognition as appropriate of the food control system applied by an exporting country's competent authority". The same paragraph indicates that this recognition can be done in "a number of ways that facilitate the entry of goods, including the use of memoranda of understanding, mutual recognition agreements and equivalence agreements and unilateral recognition".

The Appendix to the document CAC/GL 47-2003 (revision 2006) states in its Section 4 the relevant information the competent authorities should use to assess the level of risk associated with the imported food. This information could include, *inter alia*:

> - The adequacy of processing controls in place in the exporting country as evidenced by its laws, regulations, and other policies; its infrastructure; and its ability to effectively enforce food safety requirements, as may be verified by audits and on-site visits by the competent authority of the importing country.
>
> - The compliance history of the food generally, irrespective of the source of the food.
>
> - The compliance history of the food with respect to the source of the food including, where available, the compliance history of:
>
> - the exporting country or region/area within an exporting country;
> - the producer and manufacturer;
> - the exporter;
> - the shipper; and
> - the importer.
>
> - reports from officially recognized inspection and/or certification bodies.

The mentioned Appendix also states in its § 11 *"Exporting countries can provide information on the control systems in place in their country and, as appropriate, may provide assurance to the importing country that a particular food complies with the food safety requirements of the importing country"*

It can be seen, from the statements above, the document "**Principles and Guidelines for Imported Food Inspection Based on Risk**" recommends procedures that take into consideration the data from the exporting country.

However Section 32 (bullet 5) of the GL 47 document also provides for *importing countries to conduct verification, for example the occasional random sampling and analysis of products on arrival*. This provision guarantees an independent assessment may be done by importing countries as appropriate.

The above statements from CAC/GL 47 clearly recommend procedures to promote agreement between the parties to lead them to fair trade and, consequently, to avoid unnecessary disputes to take place.

Taking these points into consideration, the statement in the 2nd paragraph of the Scope of CAC/GL 70-2009, "**examines only the validity of the importing country's results on which non-compliance is alleged,**" creates an inconsistency with other Codex approved texts, like CAC/GL 47-2003 (revision 2006), unless GL 47 has been taken into account in specifying the form of CA to be used.

The present DP aims to help Codex member countries to deal with dispute situations covering important aspects that were not covered by CAC/GL 70-2009 and also to ensure that the procedures in that document are balanced with other relevant guidance.

5.1.2 Codex Documents on Dispute Situations

The following are the only Codex documents that present some kind of provision related to dispute situations. As can be seen, none of them presents a definition of dispute which may be considered satisfactory for Codex purposes or a specific procedure to resolve it. These documents only mention the situation of a dispute where some specific provisions can be applied. The documents are listed and their provisions are quoted for ease of reference.

1 - CODEX STAN 193-1995: "Codex General Standard for Contaminants and Toxins in Food and Feed" states in its § 2, p.3: "National measures regarding food and feed contamination should avoid the creation of unnecessary barriers to international trade in food and feed commodities. The purpose of the GSCTFF is to provide guidance about possible approaches to eliminate or reduce the contamination problem and to promote international harmonization through recommendations which in turn may prevent trade barriers and ***disputes***".

2 – CAC/GL 34 –1999: "Guidelines for the development of equivalence agreements regarding food import and export inspection and certification systems", presents only a definition for dispute settlement under the WTO agreements, in its p.6 as follow:

> "***dispute settlement***: A description of the consultative procedures, joint committee, and/or other mechanisms that should be employed by the participants ***to resolve disputes*** under the agreement. Such procedures and mechanisms should not limit the rights or obligations of the parties under the World Trade Organization (WTO) Agreements".

3 – CAC/GL 50-2004: "General Guidelines on Sampling": this document presents two provisions for dispute situations. The first provision described in its item 1.2, p.5, states: "These Guidelines are above all aimed at Codex Commodity Committees which select from the plans recommended in sections 3, 4, and 5 those which at the time of the drafting of a commodity standard appear to them best suited for the inspection to be made. These Guidelines can also be used, if applicable, by governments ***in case of international trade disputes****"*.

The second provision for dispute situations is mentioned in its item 2.1.2, p. 10 and states:

> "The precise definition of an acceptance sampling procedure will require the setting or selection of:
>
> - The characteristic to be measured
> - Lot size
> - An attribute or variables plan
> - The Limiting Quality (LQ) level, for isolated lots; or the AQL (Acceptable Quality Level), for a continuous series of lots
> - The level of inspection
> - The size of the sample
> - The criteria for acceptance or rejection of the lot

- ***The procedures to be adopted in cases of dispute****"*

4 - CAC/RCP 58-2005: "Code of Hygienic Practice for Meat" states in its item 4.3 - **Laboratory analysis,** §15, p.51**:** "Methods for detection and enumeration should be practical, accurate, reproducible, sensitive and selective. Only methods for which the reliability and reproducibility have been validated should be used. Inter-laboratory testing should be a feature of a microbiological verification programme. ***In cases of dispute, recognised reference methods should be used.****"*

5 - CAC/GL 71-2009: "Guidelines for the design and implementation of national regulatory food safety assurance programme associated with the use of veterinary drugs in food producing animals" presents only one provision for dispute situations and it states in its § 53, p.8 that "verification programmes should define sampling and identification procedures that allow tracing each sample back to its origin and independent confirmation of the finding ***in case of dispute***."

There are other Codex documents that mention the word dispute although they do not characterize a dispute situation. As an example the document **CODEX STAN 190-1999 - Codex General Standard for Quick Frozen Fish Fillets,** in its Annex A, Items 4 and 5, can be mentioned where the wording "disputed material" appears, with the sense that it refers to the portion of the product on which the sensory evaluation has to be re-evaluated after cooking procedures.

This guidance, scattered in different documents, is necessary but not sufficient to allow the building of a comprehensive framework for an efficient and fair approach aiming at the resolution of disputes over the status of a lot or consignment in international trade of foods.

In this situation, the present Discussion Paper is pioneering in its intention to offer a broad set of consistent information to help member countries to take the necessary measures for dispute resolution in accordance to the Codex principles of protection of consumer's health and fostering fair trade practices.

5.1.3 Other Causes of Disputes

The footnote nº 3 of the document CAC/GL 70-2001, describes the following causes of disputes:

- the existence, appropriateness and statistical validity of the sampling plan used to assess the product;
- the allowances made for normal measurement error and within-lot product variation;
- differences in physical sampling procedures;
- differences in composition of the samples tested due to product inhomogeneity; or
- changes occurring during storage and/or transport of the product.

Other causes of dispute may arise from:

- sample differences between lots;
- sample preparation for analysis;
- withdrawal of aliquot for analysis;
- selection of analytical procedure;
- differences in laboratory performance greater than those expected between the two laboratories;
- adequacy of the interpretation of the results; or
- correct application of the decision rules on conformance to the results obtained.

Regarding the first cause in the list, there are some specific dispute situations that should be mentioned:

The CA procedure is based on Codex standards or guidelines

This is a situation where a CA procedure has been established, taking into consideration Codex Guidelines and Standards, and a claim is made that the prescribed procedure has not been properly carried out. The party that raises the claim must demonstrate to the other party, reasons for believing that the procedure has not been properly carried out. This may include data bearing on the inherent unlikeliness of non-conformity, such as data, appropriately analysed, obtained by the application of adequate control procedures, from production to sale. Due to the complexity of the operations involved both countries should always bear in mind the possibility of a dispute, and, as part of the procedures mentioned in the last paragraph of Section 5.1, maintain adequate records to enable their compliance with the specified procedures to be assessed. Exchange of information in this situation is of utmost importance and an appropriate analysis of the data has

to be done, avoiding misinterpretation of the results and identifying, precisely, the procedures that were not adequately carried out.

If the CA procedures are not appropriately carried out a situation is created where either compliant product is subject to unacceptable risk of rejection (**α** error) or non-compliant product is subject to unacceptable risk of acceptance (**β** error).

The CA procedure is not based on Codex standards or guidelines, or is absent

Where a CA procedure is not based on Codex standards or guidelines, or is absent, it may also be necessary to deal with a claim that the procedure that was used exposed the exporting country to unacceptable risk of failure for compliant product. This may include, for instance, situations related to the application of statistical tools based on laws of probability, such as sampling plans, when the level of inspection and the AQL or LQ chosen by the importing country were unnecessarily stricter, for evaluation of the specification of interest, than those chosen by the exporting country. An example could be the choice of an AQL of 1.0% for a characteristic that does not pose any health risk to the consumer, where an AQL of 6.5% could be considered appropriate.

Such situations should be avoided, as no basis to assess conformity exists. Decision rules should be agreed before trade takes place, otherwise there is no means to establish fair practices during trade activities, and related costs of failure may be considerable for the two parties. If no CA method has been established, the whole procedure has to be examined in detail whenever a dispute is raised. Such an examination may be lengthy and costly.

The importing country would have to demonstrate that they operated to a fixed sampling plan, and did not, for example, simply go on taking samples until they found one that appeared non-compliant. This fixed plan would have to be assessed statistically to estimate the associated producer's and consumer's risk.

Unless these issues are appropriately considered, the resolution of dispute can not be reached in a balanced way and the international trade practices will not be fair.

5.2 HOW THESE CAUSES MAY AFFECT THE DISPUTE RESOLUTION PROCESS

In a dispute, the aggrieved country presents a formal appeal to the importing country, explaining its grounds for disagreement and requesting the establishment of a dispute resolution process. To elaborate the contents of this section, it is appropriate to make a proposal of what can be understood by 'dispute resolution process' (DRP), in order to identify its main/basic components, to examine how the mentioned causes of disputes affect these components, and to identify the outcome of these effects.

The DRP begins once the importing country accepts the appeal from the aggrieved country and agrees to establish a DRP.

When trade agreements contain procedures for dispute resolution these procedures constitute what can be understood by a dispute resolution process. On the other hand, when trade agreements do not have provisions for dispute resolution, it is necessary to define or to adopt procedures capable of helping the parties to identify the problem, its causes, and how to manage them to solve the identified problem.

Taking these points into consideration it can be said that:

> '**Dispute Resolution Process** (**DRP**) is a set of agreed procedures, between the parties or governments, for the purpose of characterizing the dispute, its causes and its resolution, that is capable of leading to a fair and balanced decision about the fate of a consignment or a lot of product in international trade'.

From the above statement the main/basic components of DRP are:

- dispute characterization;
- identification of the causes of disputes and their resolution;
- the decision on the fate of the consignment or lot.

The identification of the causes of disputes was discussed in the previous Sections of the DP and will only be generally mentioned here. If a dispute is not well characterized the parties cannot move to deal with the next components of DRP. So, in this section of the discussion paper we will focus on how the causes of disputes affect dispute characterization and how they affect the decision on the fate of the consignment or lot in international trade.

Dispute characterization

When the product fails an importer's CA, this failure must be communicated by the importing country to the exporting country, by means of an exchange of information, and this communication has to specify the procedures used in the CA, the analytical results obtained and the finding of non-conformity in a clear way, indicating in particular any variation from agreed procedures and stating the grounds for failure.

The exchange of information is a statement that mentions all the applied procedures that led to the finding of non-conformity. This facilitates the understanding by the exporting country of the causes of disputes and the methods through which the results were obtained. The exchange of information should include the information on the sampling plan, and sampling uncertainty where it is appropriate; sampling technique; preservation, transport, storage, preparation and analysis of samples; the analytical method and its performance characteristics; the analytical results with their measurement uncertainty and recovery factors where necessary; and the statistical method applied to primary data to generate the final outcome. Much of the required material may be covered in a relevant agreement or Codex document, and for this material reference to the relevant documents would normally be adequate.

The importer has to communicate all the information prescribed in the Codex document CAC/GL 25-1997 – "Guidelines for the Exchange of Information Between Countries on Rejections of Imported Food", especially the contents mentioned in the paragraphs 14 to 17 related to 'reasons for rejection'. This information has to follow the model indicated in the Annex named "Standard Format for Exchange of Information between Countries on Rejections of Imported Food".

If the exchange of information is not complete and transparent it creates difficulties for the exporting country, in identifying possible explanations for any apparent discrepancy between the results of the exporting and importing countries and also in identifying the sampling and measurement methods by which the results were obtained. If it is not well documented, this limitation produces a negative effect on the DRP, making the process lengthy and costly. It may also give rise to a lack of trust: an exporting country may feel that details are being withheld in case they may give grounds for dispute if revealed.

Identification of the causes of disputes

A proper investigation leading to correct identification of the causes of disputes produces a positive effect on the DRP and contributes to making the process easier, understandable, speedy, and cost advantageous. It also assists the identification of the scientific, technical and administrative elements that may lead to the resolution of a dispute. It is a key step to be accomplished before moving to the next step of the DRP.

The process employed must ensure that the correct cause is identified before resolution is attempted. For example, care must be taken not to interpret differences in results as being necessarily due to differences in analytical methodology (or its application) when these differences might be due to other reasons such as deterioration of perishable foods during storage or transit, or merely due to the testing of different samples by the exporter and the importer, particularly in lots with large variability (whether homogeneous or non-homogeneous).

The decision on the fate of a consignment or lot of a food

If the exchange of information mentioned before is complete and transparent, this contributes to improving the level of understanding regarding the difference between the methods and results of CA used by the importing and exporting countries. If it is decided to reassess the CA through the generation of new data, it will be much easier to reach an agreement on how to proceed and the methods that will be applied.

To proceed further towards an adequate decision on the fate of a consignment or lot of a food, it is necessary to apply appropriate methodology for resolution of disputes. This methodology must be such that the producer's and consumer's risks associated with the prescribed CA procedure which was (or should have been) applied by the importing country are affected as little as possible.

So the effect of the DRP on the fate of a consignment or lot of a food in international trade is positive if the dispute characterization is clear and transparent, the causes of dispute are unambiguously identified, and the proposal for the resolution is based on sound scientific, technical and administrative elements and based on approved Codex-related texts.

The dispute ends when the grounds for the appeal have been examined by the importing country and a response to each part of the appeal is made and a consensus is reached on the outcomes.

5.3 HOW TO MANAGE THESE CAUSES DURING THE DISPUTE RESOLUTION PROCESS

The ideal scenario is the establishment of a prior agreement concerning the management of possible causes of dispute during the entire chain from production to sale and to avoid these situations occurring. The worst situations arising in the international trade of food are those where a dispute takes place. In situations like this, the DRP must identify the possible causes of dispute as discussed in the previous Section of the DP.

The management of the causes of disputes during the DRP is a cooperative process. The importing and exporting countries must work together and exchange information to reach an agreement on the identification of causes of dispute and how they occurred. Related to this step the management has to consider all the available scientific and technical guidelines and tools to control the causes of disputes. One of the available tools is the HACCP, for instance. A series of questions could be raised to investigate how the dispute may have arisen. The questions could be put in the form of a checklist, such as the following example. This list is not comprehensive because the situations of disputes vary considerably, and have to be approached on a case-by-case basis:

Example of a checklist of information to investigate the causes of dispute during the DRP

Control systems applied by the exporting country

1 Was HACCP appropriately applied?

2 Were all the critical control points correctly identified?

3 Were they appropriately controlled?

4 Were the data adequately obtained to control the critical points?

Sampling

5 Was the sampling plan appropriately chosen and applied?

6 Were the methods used to choose the points in the lot from which samples were taken appropriate?

7 Was the sampling procedure performed adequately?

8 Were the samples appropriately packed, identified, stored and transported?

9 Were the samples adequately prepared for analysis?

Test methods and testing

10 Were the analytical methods fit for purpose (e.g. sufficiently specific and precise for the concentration range of the analyte)?

11 Were the methods fully validated or validated in house, at least?

12 Were the analysts sufficiently trained to run the method according to good laboratory practice?

Results (and communication of results)

13 Were the analytical results appropriately recorded and reported for interpretation?

Interpretation/conformity assessment

14 What CA procedure and associated statistical methods were used? Have these been agreed? If not agreed, were they appropriate to the nature of the product and analyte and the risks associated with them?

15 Was the CA procedure applied correctly and was the conclusion presented in a clear and unambiguous way?

16 Was the conclusion appropriately communicated to the interested party?

17 Were the procedures to allow the interested party to question the obtained results appropriately communicated to that party?

18 Was the exchange of information between the importing and exporting countries done according to the Codex prescribed procedures contained in the document CAC/GL 25-1997 – "Guidelines for the Exchange of Information Between Countries on Rejections of Imported Food"?

It would be appropriate if the aforementioned questions were considered in the context of a broader document that could be named DRP Checklist. The headings for this document could be:

1 Exchange of information;
2 Dispute characterization;
3 Causes of disputes;
4 How to manage the causes of dispute: checklist of information;
5 Consideration and evaluation of the evidence and decision on the fate of consignment or lot;
6 Statement of agreement on the decision reached;
7 List of participants in the DRP.

5.4 RISKS OF WRONG DECISIONS IN THE DISPUTE RESOLUTION PROCESS (DRP)

This section focuses on the risks of wrong decisions being made during the dispute resolution process. It should be emphasized first of all that the most appropriate place for control of risks of wrong decisions by the importing country **in respect of product compliance** is in the choice of the procedure of CA used by that country. If the procedure is well chosen and implemented, and particularly if it has been agreed with the exporting country, ideally the DRP would alter these risks as little as possible, since the only changes to these risks that can result, when the combined effect of this CA and DRP is considered, are a reduction in producer's risk at the expense of the consumer's risk. This is simply because a DRP opens up a second route by which product, whether satisfactory or unsatisfactory, may be passed, and the resulting increase in the probability of acceptance of an unsatisfactory product must be kept small.

The scenario being considered is that the parties in dispute have already taken all the recommended measures to characterize the dispute, all documents have been presented for analysis and exchange of information, and the parties agree that they have all the necessary information for making a decision. If the method of CA that has been used by the importing country is such as to lead, when carried out properly, to an acceptable level of producer's risk, the question at issue in a dispute is whether this CA has in fact been carried out in such a way that this level of risk has been, or may have been, significantly exceeded. It is to be expected that, in the presence of sampling and measurement error, a small number of lots which are in fact satisfactory may in fact fail, by virtue of the random nature of sampling and measurement errors, but this is the price to be paid for a reasonable probability that an unsatisfactory lot will fail (in spite of such random errors). On various occasions, the exporting country may be required, or think it has been required, to pay this price[21]. These occasions may be unwelcome, but unless over the long term the exporting country has to pay it more often than it should (that is, that over the long term, product that the exporting country has good reason to suppose compliant is failing more often than the nominal producer's risk would suggest) they do not give it real cause for complaint.

It may seem reasonable in dispute situations for an exporting country to be allowed a second chance to have *acceptable* product passed. However, whenever this is done, the importing country is at the same time exposed to a second risk that *unacceptable* product will be passed. If it were possible, at reasonable cost, to determine unambiguously during a dispute whether or not the product was in fact acceptable, this problem would not arise, but this scenario is unlikely. If it is granted that the producer's risk to which acceptable product was exposed, under the selected method of CA, was reasonable, and that the importing country was not at fault in the way the test was carried out, it is desirable that the dispute process should increase the consumer's risk as little as possible. Thus it is desirable that the importing country's assessment should be allowed to stand unless either:

a) the method of CA was not properly carried out, or

b) the method of CA, even when properly carried out, exposed the exporting country to undue risk that compliant product will be rejected.

It should be particularly noted that in a DRP the question of whether the lot was compliant in the first place is not the question being considered. (Of course, evidence that the producer may have on this matter may be relevant in deciding whether a disputes procedure may reasonably be invoked.) It is only when a) or b) applies that the true compliance status of the lot becomes an issue.

In summary, there are certainly risks that even after the DRP procedure has been followed, product will still be misclassified. But if neither the CA used nor its manner of implementation is at fault, ideally these risks would remain unchanged from those expected under the original method of CA. It is in the choice of the

[21] The only way of avoiding these occasions is to reduce the producer's risk attached to the importing country's CA, either by allowing an increase in the consumer's risk or by changing the method of CA, e.g. by increasing the number of samples used or by using improved measurement techniques.

original CA procedure that such risks should be managed. Once this has been achieved it should be the aim of a DRP to change these risks as little as possible. This is because if the combined effect of a CA and DRP, as compared to a simple CA, is a reduced risk to the exporting country, there must inevitably be an increased risk to the importing country.

5.4.1 How to manage the risk of wrong decisions in the Dispute Resolution Process (DRP)

As stated above, it is not primarily the role of a disputes process to manage risks of wrong decisions: these are preferably controlled in the method of CA itself. It is only when the method of CA used was not properly carried out, or exposes the exporting country to an inappropriately high risk of failure for acceptable product that consideration should be given to replacing the initial assessment by a reassessment of the lot.

These situations may occur by accident (e.g. laboratory outliers, damage to samples), through insufficient consideration being given to the producer's risk, through inappropriate statistical methods being used to calculate it, or through failure to take into account product related factors that affect the risk.

It is the role of a DRP to assess whether one of these situations a) or b) in fact occurred, and if so to negotiate some remedial action.

The types of wrong decisions that can be made can be summarized as follows:

a) Was the CA properly carried out?

   i) Deciding that it was not when it was.

   ii) Deciding that it was when it was not.

b) Did the method of CA itself expose the exporting country to undue risk of failure for acceptable product?

   i) Deciding that it did when it did not.

   ii) Deciding that it did not when it did.

It would be difficult if not impossible to quantify the probabilities of either of these wrong decisions occurring under a given DRP. All that can be done is try to keep them to a minimum. All that can be done is to take all the preventive measures that could contribute to keeping them to a minimum.

Was the CA properly carried out?

Investigation of this question would seem to be largely a matter of reviewing the taking, storage, transporting, preparing and analysing the samples, and reviewing the calculation of the acceptance criterion from the results obtained (see items 5 to 13 in the investigation checklist, section 5.3).

The sampling procedure (item 7 on the checklist) is often a concern. For example, in the importing country sampling may have been concentrated in parts of a lot that are considered most likely to contain product that is out of specification. However if the specified sampling plan is based on keeping the producer's risk below a certain level provided the AQL is not exceeded, such concentrated sampling will defeat this intention. Indeed if samples are significantly clustered into a few parts of a lot, several samples being taken in each cluster, it would be difficult to maintain that either the producer's risk or the consumer's risk associated with the sampling plan remains valid. This is particularly so in the case of perishable product, where a lot that is initially homogenous may be less so when it enters the importing country.

In regard to cases where the analytical results themselves are in question, GL 70 is the recommended DRP.

Did the method of CA itself expose the exporting country to undue risk of failure for acceptable product?

The answer for this question requires an assessment of the probability of failure to which acceptable product is exposed (see items 14 and 15 in the investigation checklist, section 5.3). In principle, prior agreement between importing and exporting countries on the CA procedure to be used could be taken to imply that the producer's and consumer's risks attaching to the CA had been deemed acceptable by both parties.

Some points which may be relevant, in addition to any points specifically raised by the exporting country on this matter, are listed below:

1) Was the CA procedure well defined? That is, were the procedures followed and the acceptance criteria to be used specified in advance? Were the conditions under which various parts of the consignment would be accepted or rejected specified? If a CA

procedure is not defined, it may well be impossible to assess the risks involved to either party.

2) Is the AQL (acceptable quality level) effectively being enforced appropriate to the product, analyte and specification?

3) In estimating the producer's risk for the test, are assumptions being made about the nature of the product which are not justified? For example, if the estimate is based on an assumed standard deviation for the product, as in certain "inspection by variable" plans in which an assumed standard deviation is used to judge the proportion out of specification from an estimate of the lot mean, the exporting country may be able to demonstrate that a much lower standard deviation is appropriate.

4) In many circumstances, allowances for measurement error may be required. Have these been made, and can it be asserted with reasonable confidence that the allowances made are adequate?

Remedial action in respect of the lot under assessment

It may not necessarily follow that because the DRP identifies one or more areas in which the CA procedure is deficient, the decision made by the importing country on the lot must necessarily be overturned. For example, if it is decided that a required allowance for measurement error was not adequate, or that no allowance had been made, it may yet be that the measurements obtained by the importing country were such as to justify rejection of the lot even had adequate allowance been made. Or if the acceptance criterion for a sampling plan has set too low, effectively enforcing an AQL that is unreasonably low in view of the nature of the product and analyte concerned, acceptance criterion may be exceeded even for a more reasonable plan based on the same sample size.

Equally, it may be clear that had an appropriate allowance for measurement error been made, or a more suitable acceptance criterion been used, the lot would have passed.

In such cases re-sampling of the lot may not be required, and the fate of the lot may be resolved by applying more appropriate acceptance criteria to the results already obtained. In situations dealt with under GL 70, where the analytical results themselves are disputed, there is provision for determining the fate of the lot by independent analysis of duplicate samples under certain circumstances.

There remains the possibility that the problems associated with the CA as carried out may be so severe that the only options that remain are to resample the lot, or to accept the exporter's own assessment of the lot. Section 4.2.1 highlights the importance of cooperation between the parties in trade, leading to the recognition of the food control system in the exporting country by the importing country "including the controls applied during production and manufacture" according to the guidance given by the Codex document *Principles and Guidelines for Imported Food Inspection Based on Risk* (Appendix to the *Guidelines for Food Import Control Systems*, CAC/GL 47). Where such recognition exists, and if the CA in question is simply a "spot check" carried out by the importing country, it may be quite reasonable to accept the exporting country's assessment when the importer's CA has been defective.

General Considerations

It is very desirable that the CA procedure to be used by an importing country should be communicated in advance to, and preferably agreed with, the exporting country. This is desirable for several reasons, among them:

- So that they can implement measures necessary to ensure that disputes are avoided.
- So that factors which may affect the validity of the proposed CA procedure (and thus have the potential to give rise to disputes) may be drawn to the importing country's attention.
- So that the importing country will be committed to a well-defined procedure of assessment, which is open to scrutiny, and which enables an assessment to be made of the relevant probabilities of failure.
- So that the exporting country will be aware of the risk of failure to which its product will be exposed, and in a position to attempt negotiations with the importing country if this seems desirable.
- More generally, to enable problems associated with the CA procedure to be sorted out before actual product is put at risk.

Independence of the DRP

Questions have been raised about whether a dispute should be required to be resolved by a third party independent of the parties to the dispute, particularly when the parties fail to reach or have difficulty reaching agreement. This seems desirable in principle, but there are sometimes difficulties in implementing it in practice. In any case it is desirable that the resolution of the dispute should rest, as far as possible, on scientific rather than on political considerations.

Agreement on the outcome of the DRP

It is necessary that the parties in dispute agree on the outcome of the DRP and take all the necessary measures to prevent recurrence, and to ensure the appropriate disposition of any rejected food product. Section 4.3 points out possible strategies to be adopted concerning rejected food in international trade taking into consideration the provisions of the *Guidelines for Food Import Control Systems* (CAC/GL 47).

5.5 HOW CA PROCEDURES AFFECT DISPUTE RESOLUTION

This section is a summary of the preceding sections. The reasoning developed above demonstrates that correct choice and application of CA procedures is essential for relatively quick and straightforward resolution of disputes when they occur. The advantages of this situation could be envisaged as follows.

Advantages of a well-defined CA procedure

In this situation firstly the CA procedure is well defined, including a method of analysis that is fit for the specified conformity assessment. Consumer's and producer's risks have been assessed and are likewise appropriate to the intended purpose. The CA procedure may be based on Codex recommendations or may be agreed between the exporting and importing countries. The CA procedure is followed appropriately.

Disputes are then relatively unlikely, since if the CA procedure is followed appropriately there is no basis for dispute, subject to satisfactory analysis of samples and appropriate allowances being made for analytical test error and bias in the CA. When a dispute occurs it can be resolved on an objective basis and can be confined to checking that the prescribed procedures have been followed correctly and/or checking on validation of measurements using GL70, thus maintaining as far as possible the consumer's risk associated with the CA procedure itself, without increase to the producer's risk. The resolution should therefore be relatively quick and non-acrimonious.

Disadvantages of non-defined or poorly defined CA procedure

On the other hand a situation can be envisaged in which the method of CA is undefined or poorly defined. In this situation it is likely that it will be difficult to calculate the consumer's and producer's risks to determine whether the importer has applied a fair assessment, or to judge whether the risks are appropriate to the parameter that is being assessed. If the method of CA has not been agreed with the exporting country, the importing country is likely to find it difficult to defend a finding of non-compliance if the method is imposing an unduly high or unknown producer's risk. The exporting country may well have good evidence to support its view that the product is compliant.

A dispute will be difficult to resolve since it will be necessary to identify the cause from a greater range of possibilities, and a new CA may need to be found and agreed that is acceptable to both parties, including appropriate control of producer's and consumer's risks. The difficulties may create unwelcome consequences for both the exporting and importing countries: unnecessary delays, costs of storage, sampling and testing, and waste of food.

Requirements for a defined DRP

If a dispute situation arises it is necessary to have clear procedures to characterize:

- how the dispute will be dealt with;
- how the dispute will be ended; and
- how to implement the conclusions of the DRP.

Advantages of a collaboratively-developed prior agreement

This paper emphasizes again the importance of a collaboratively-developed prior agreement between the exporting and importing countries about the appropriate CA procedures associated with Codex specifications and recommendations. As well as giving appropriate control of risks of misclassification of the product

under consideration, such agreements minimize the likelihood of disputes and ensure relatively quick and simple resolution when disputes occur.

6 Conclusions

This document has considered procedures for conformity assessment and its role in reaching a fair assessment of product compliance, including the resolution of subsequent disputes that might occur.

The DP pointed out that both conformity assessment (CA) and the dispute resolution process (DRP) constitute a complex set of activities with uncertainties inherent in them. These uncertainties may contribute to make the DRP lengthy and costly and its consequences are of great importance. As a result of these issues, this document considered the approaches and perspectives for preventing the raising of a dispute, and collaboratively resolving it if one occurs, as the best way to proceed.

PRIOR AGREEMENT

This DP has emphasized that, to minimize the risk of a dispute, it is very desirable that before starting trading activities, the parties concerned should reach an agreement related to the CA procedures that will be applied by the importing country and also on the DRP to be followed in the case of a dispute.

SELECTION OF A METHOD OF CONFORMITY ASSESSMENT

Sections 5.4 and 5.4.1 presented the view that the management of risks of wrong decisions related to the fate of a food in international market should be controlled through the appropriate choice and application of a CA procedure itself. Such a procedure must be well chosen and implemented by the importing and exporting countries.

Before deciding on the most appropriate procedure for CA, the parties entering into trade must give consideration to the fact that variation of foods may be diverse due to the origins of raw materials, varied farming and manufacturing processes, and differences between manufacturers. It may thus be difficult or impossible to determine estimates of sampling variability that are universally applicable, even for a well-defined single food type. This may give rise to a need for lot-specific estimation of variability within the CA procedure itself, normally requiring a multi-sample assessment. The chosen CA procedure should control the producer's and consumer's risks in a way that does not penalize a compliant product with a high risk of rejection while providing adequate consumer protection. Suitable procedures can be sought for this purpose.

The risks of wrong decisions could be considered as having been accepted by both parties if the CA procedure has been agreed by them, as both have had the opportunity to be aware of the associated risks of failure to which the product will be exposed, and to negotiate accordingly. To help countries going into trade, Section 4.1.2 mentions a set of Codex documents that deal with CA which should be consulted prior to a trade agreement being reached.

CONSIDERATIONS IN THE SELECTION OF A CONFORMITY ASSESSMENT PROCEDURE

The following issues have been identified as requiring consideration:

1) Managing producer's and consumer's risks

In regard to managing producer's and consumer's risks, as a matter of principle it is undesirable to expose compliant product to unknown and uncontrolled risks of failure – in other words, importing countries should manage their consumer's risk in a way that does not unduly penalize exporting countries. This implies a maximum producer's risk. Given that such a criterion on producer's risk is agreed, it would then be up to the control authorities to define the adequate measures to control both the producer's and consumer's risks in a balanced way within this constraint.

2) On the role of sampling method

The DP recognizes choice of a statistical sampling method as an important factor in the choice of a CA procedure. Not only is the associated sampling uncertainty affected, but also the final cost and timeliness of the assessment. For non-homogeneous products stratification must be considered when the application of a sampling method is considered, noting that it is not entirely clear what this stratification entails (see (7) below). If these kinds of products are sampled randomly, and the samples are composited, the variability of the food may not be adequately allowed for and can lead to wrong decisions on conformity assessment. In choosing an appropriate sampling plan, due consideration must be given to the following characteristics: whether the lot is considered in isolation or is part of a continuous series; whether the measurement is qualitative or quantitative; and whether it is required to control the percentage non-conforming or the average analyte content.

It was stated in Section 4.1.1 that statistical sampling involves the drawing of conclusions about the status of a lot as a whole, from data obtained from sampling the lot. It was pointed out that it is not possible to determine, without possibility of error, whether a lot is compliant. This limitation has always to be kept in mind when sampling is done at the point of import, particularly for non-stable or perishable foods, whose conditions may change in the course of storage and transportation. Ideally, 100% sampling could solve this problem but it is not practical except for small lot sizes when sampling is non-destructive.

Section 4.1.2 demonstrated that the *General Guidelines on Sampling* (CAC/GL 50-2004) is an important document to be consulted when developing appropriate CA procedures for foods in trade. It needs clarification on some points related to sampling of materials sold in bulk, especially about the use of the terms "producer's risk" and "consumer's risk" (as pointed out in Section 4.2.4), and about significant measurement error and stratified sampling of non-homogeneous lots.

3) Differing approaches to CA

A fundamental difference exists between the traditional approach to acceptance sampling and the approach used in the *Guidelines on Measurement Uncertainty* (CAC/GL 54-2004). This discussion paper favours the traditional approach, as discussed in section 4.2.5, topic "Definition of analytical measurement uncertainty", and Annex 1.

4) Measurement uncertainty

The conditions under which the current concept of measurement uncertainty can be used for satisfactory conformity assessments need to be considered. To control producer's and consumer's risks, the relative frequency with which a measured value exceeds (or falls short of, depending on the context) the relevant true value by more than a given amount is of fundamental importance in compliance testing. This is a parameter of the measurement process used, and it is desirable to have limits, particularly an upper limit, to its plausible values. The approach to measurement uncertainty as presented in the *Guidelines on Measurement Uncertainty* results in uncertainty intervals for which a "best estimate" of this relative frequency takes a specified value (e.g. 5%) rather than intervals for which there is a reasonable assurance that this specified value is actually achieved or bettered. This approach is of very limited value in controlling acceptance probabilities and in designing conformity assessment procedures for food.

In other words, when making allowances for measurement uncertainty it is necessary, as indicated in Section 4.5, to use a worst case scenario rather than a best estimate regarding measurement error.

5) Sampling uncertainty

In many situations the measurement error may be negligible compared to sampling variation, caused by variation in analyte levels, or product quality, within a lot. The fate of a lot will then be determined largely by which particular sample or samples happen to have been selected for the conformity assessment, and measurement error will have a negligible impact on the operating characteristics of the CA. In such a situation the theory of acceptance sampling, on which the General Guidelines on Sampling (GGS) are based, is well developed, and therefore there is little reason to develop a theory of "sampling uncertainty" as part of "measurement uncertainty", to be dealt with under the approach underlying the Guidelines to Measurement Uncertainty with its associated problems in controlling consumer's and producer's risks (see paragraph 4 above). This is particularly so in that such an approach leads to conformity assessments that are limited to single-sample assessment of the mean analyte level in a lot. Although it could be philosophically attractive to consider sampling as part of the measurement process, the practical implications of doing so can be undesirable, both in restricting the types of CA available and requiring even these to be dealt with in a way that is not the most appropriate. The approach to these issues is dealt with in Section 4.2.4.

6) Assessment where both measurement and sampling error may be significant

In the presence of significant measurement uncertainty, the GGS methods are not immediately applicable, and significant work will be needed to modify the acceptance sampling procedures accordingly. A particular problem lies in the "percentage defective" approach to acceptance sampling and to the setting of producer's and consumer's risks. In the presence of significant measurement error, for product close to a compliance limit the levels of 0% defective and 100% defective may be virtually indistinguishable. While this *should* have limited practical implications in testing of food, in practice it creates severe difficulties with the standard approach to acceptance sampling via an AQL and LQ, and the associated producer's and consumer's risks, on which the GGS is based. The DP, in section 4.1.1,

has suggested that these difficulties may be mitigated if the AQL and LQ are set with respect to different analyte levels, well separated by the measurement method used. For instance the AQL may be set with reference to an analyte level reflecting good manufacturing practice, and the LQ with reference to a somewhat higher analyte level, still considerably short of the level at which the product would start to pose a threat to the consumer. Further development of this idea may be warranted, although it would apparently require commodity committees to consider two "specification limits" instead of just one.

7) Heterogeneous lots

The GGS recommends "stratified sampling" for heterogeneous lots. While it is possible to interpret this literally, as implying the acceptance or rejection of an entire lot using an estimate of the mean analyte level of percentage defective derived from a stratified random sample, it is not clear that this is the intended meaning. Another possible interpretation is that the heterogeneous lot should be divided into more homogeneous sub lots, each to be sampled, and accepted or rejected, independently. CCMAS could consider further work on this issue.

8) Lot sizes

The indexing of sampling plans to lot sizes used in the GGS, and derived from similar indexing in ISO, should be reviewed. Such indexing avoids the need for explicit (and possibly unrealistic) consideration of consumer's risk in choosing a sampling plan, but its arbitrary nature creates difficulties in applying attribute sampling plans to bulk product, leaving control of the mean as the only option for such lots. CCMAS could consider further work on this issue.

9) Sampling plans for control of the mean analyte level

The choice of a sampling plan for control of the mean is not very well dealt with in GGS, which concentrates largely on control of the percentage defective. There is ambiguity regarding the term "discrimination difference" and the prerequisites for applying the procedures to bulk product seem unrealistically demanding. Some clarification and expansion of the relevant material may be desirable. CCMAS could consider further work on this issue.

10) Product specifications

Currently a compliance limit is given, and the GGS gives some guidance on the appropriate AQL for products and defects of various types. The interpretation is that the compliance limit is a limit to which all samples should conform, and the sampling plans are geared to yield appropriate probabilities of acceptance and rejection as the percentage above this limit rises. In food testing, situations occur where the percentage above the limit is not the critical factor, and more important is how far above this limit the non-conforming samples actually are: in many cases, limits are set with reference to good manufacturing practice, and they are considerably below levels at which health risks start to appear.

11) On the fitness for purpose of an analytical method for use in a CA procedure.

The contents of this DP highlight the need for a deeper understanding about the impact of analytical methods on the producer's and consumer's risks. This led to an explanation on the meaning of fitness for purpose and how it relates to the activities of CA.

Consideration could be giving as regarding fitness for purpose of an analytical method as it could be expressed as the effect on consumers' and producers' risks (defined in terms of the relative frequencies of the possible outcomes for lots of given quality) from the specified conformity assessment system (sampling plan and the use of that analytical method). Thus a test method and a sampling plan for a parameter specified in a commodity standard is an implied statement of the required fitness for purpose of that parameter, since the test method and sampling plan in combination imply the levels of consumers' risk and producers' risk that are deemed appropriate.

In the light of this clarification and the interpretation presented at the Section 4.2.6, a concept of fitness for purpose in the context of Codex standards might be described as follows:

> An analytical method is fit for use in a given conformity assessment procedure if the method and CA procedure used in conjunction meet specified criteria for consumer's or producer's risk.

This new concept specifically related to conformity assessment, highlights the types of consideration involved and their implications and could be considered by CCMAS.

12) On preventive measures in exporting countries

Note that the exporting country will very likely use different CA procedures, in particular because it needs to assure the importing country that the consumer's risk is controlled, and because it is likely to have greater knowledge of the product's variability. The exporting country's CA procedures generally include a combination of end-product testing and a range of other controls. Effective management of the combined procedures is vital, but has not been considered in the DP as it is outside the scope. However further work by Codex may be desirable.

13) Need for overall risk management strategy in importing countries

Assuming that producer's risk is to be adequately controlled, obtaining satisfactory consumer's risk at moderate cost (that is, using small numbers of samples) could be difficult or impossible (see Annex 4). The DP has recommended that importing countries should have an overall risk management strategy of which compliance testing at the border is only one of a number of measures used to manage risk. Such strategies are likely to require a degree of cooperation between importing and exporting countries. Codex could provide guidance for importing countries in developing such strategies.

GENERAL COMMENTS ON DISPUTES AND THEIR RESOLUTION

Dispute

In this DP the word "dispute" (defined in section 3) has been reserved for situations in which an exporting country appeals against an assessment of non-conformity based on a CA carried out by an importing country.

Other types of dispute (for example, what sanctions should result from a finding of non-conformity, or to whom they should be applied) have not been covered, as they were considered out of scope.

Reporting a failure

When product fails an importer's CA, this failure must be communicated by the importing country to the exporting country, by means of an exchange of information, and this communication has to specify the procedures used in the CA, and the outcome obtained in a clear way, indicating in particular any variation from agreed procedures and stating the grounds for failure.

Initiating the dispute situation

The exporting country may then appeal against the finding of non-conformity, giving the grounds for appeal and asking for a dispute resolution process (DRP) to be invoked.

The appeal may be focused on the conformity assessment procedure itself, or on the way in which it was carried out. It is desirable that the CA procedure itself, with its associated consumer's and producer's risks, should have been agreed by both parties: this being so, the DRP should change the agreed risks by as little as possible. The only changes to these risks that can result, when the combined effect of this CA and DRP is considered, are a reduction in producer's risk and an increase in consumer's risk. This is because a DRP opens up a second route by which product, whether satisfactory or unsatisfactory, may be passed, and the resulting increase in the probability of acceptance of an unsatisfactory product must be kept small.

The conditions under which a dispute can reasonably be raised need to be considered in a general way, particularly in respect of the type of data that can be submitted by the exporting country to justify an appeal. Some suggestions for dealing with this question have assumed that each party is in possession a result from a single sample from a lot, which have been effectively treated as analytical duplicates. A more general treatment is advisable, particularly in respect of the type of data that can be submitted by the exporting country to justify an appeal. For example they may have extensive data on lots manufactured under the same conditions, but none on the particular lot in question, which may have been formed more or less arbitrarily from a larger consignment. In other words the exporting country may have data from "production lots" or "delivery lots" but not necessarily relating to the "inspection lot" that has failed in the importing country.

Need for cooperation during a dispute

Section 5.3 demonstrated that a DRP must be a cooperative process between the parties to avoid a lengthy and costly process and to guarantee a fair decision on the fate of a food in international trade. Exchange of information during the DRP is a key step to facilitate the advancement of the process. It is recommended that the exchange of information be done according to the Guidelines *for the Exchange of Information Between Countries on Rejections of Imported Food* (CAC/GL 25-1997), especially the points mentioned in the paragraphs 14 to 17 related to 'reasons for rejection'. It is necessary that the parties in dispute agree on the outcome of the DRP, and take all the necessary measures to prevent recurrence, and ensure the appropriate disposition of any rejected food product.

The dispute resolution process

The main question to be addressed is whether the CA was such as to expose acceptable product to an unacceptably high probability of failure (producer's risk.) This may have arisen either through the CA procedure chosen or through incorrectly applying the procedure (e.g. laboratory outliers, systems failure, improper sample selection, mistakes in data transcription, etc.) A checklist is a useful way to identify possible causes of a dispute and their effects.

In the absence of a prior agreement between the importing and exporting countries, it may be necessary to establish during the DRP whether the CA procedure was appropriate to the nature of the product and to the risks associated to its consumption.

Provided the probability of failure attached to the CA procedure itself is deemed acceptable (as evidenced by, for example, prior agreement of the procedure by the parties themselves, recommendation by Codex or its commodity committees, or by a specific assessment of the procedure in the course of the dispute) it should be the aim of the procedure to rectify or allow for deficiencies in carrying out the CA on the lot in question, in a way that increases as little as possible the probability of passing unacceptable product (i.e. increases the consumer's risk).

Finally, the fate of the lot concerned is decided, and relevant follow-up actions to prevent recurrence of the dispute are taken.

Actions following a dispute

Following a dispute, where possible the cause of the dispute should be identified, and control measures improved to prevent its recurrence. Data for the entire chain from production to import is potentially relevant.

7 Recommendations and future work

During the elaboration of this DP several important issues were identified that deserve further consideration by CCMAS in order to provide appropriate guidance. This section enumerates the identified points, making proposals for solutions where possible and leaving others to be proposed by CCMAS.

MEASUREMENT UNCERTAINTY AND CONFORMITY ASSESSMENT

1. The concept of measurement uncertainty used in the "Guidelines to Measurement Uncertainty" has serious limitations for the purposes of controlling consumer's and producer's risks, (see for, example, those identified in section 6 (CONSIDERATIONS IN THE SELECTION OF A CONFORMITY ASSESSMENT PROCEDURE, paragraph 4)).

   CCMAS is recommended to consider the development of an appropriate concept of measurement uncertainty for the purposes of conformity assessment, such as an upper limit to its plausible values.

CLARIFICATION OR EXPANSION OF THE GENERAL GUIDELINES ON SAMPLING (GGS)

2. CCMAS is recommended to consider specific amendments to the GGS in respect of:

   a) the treatment of bulk materials;
   b) methods for control of the lot mean;
   c) indexing of sample sizes to lot sizes.

3. Advice on methods for practical application of the General Guidelines on Sampling (GGS) is necessary as it has been noted that the guidelines may be considered complex for routine use, and the sample sizes involved may have an important impact on the costs of sampling, testing and administration. Consideration could be given to simplifying the way in which sampling plans are presented. On the other hand, it has been noted that any alternative, simpler, sampling plans should be formulated taking account of the consumers' and producers' risks, and the costs of wrong decisions.

   Consideration could be given to alternatives to GGS plans where single sample tests are concerned for CA. As an alternative to the procedures suggested in the GGS, conformity assessment procedures have been proposed for assessing the conformity of a lot on the basis of the test result from one sample

together with uncertainty information[22]. Such procedures have been proposed as a way of reducing sampling and testing costs, and may reflect the reality of import control procedures. However, details on how these procedures are to be used to assess the conformity of individual lots are currently unclear. Clarification is needed and should include:

- Both analytical measurement uncertainty and uncertainty of sampling, and the allowance, if any, to be made for uncertainty in the estimation of parameters of the distributions of measurement and sampling error.
- Whether the procedures are confined to assessment of conformity using the lot mean, as a consequence of a restriction to single (possibly composite) sample assessment.
- The effects on producer's and consumer's risks of the method of estimation of within-lot variation (including the possible impact of underestimation) and the uncertainty of the estimates.
- The way method or laboratory bias should be allowed for.
- Clarification of the criteria for applying a particular set of estimates of bias and within-lot variation to a particular lot (e.g. criteria might be: same manufacturer, same consignment, same time of year, or other relevant factor).
- Managing producers' and consumers' risks appropriately, including the level of rejection of compliant product, and including all possibilities of rejection to which the lot is exposed.

International standardization organizations may be able to assist with this work.

4. Development of methods for dealing with situations where both analytical and sampling uncertainty play a significant role. The GGS do not cover the control of homogeneous goods in cases where measurement error is not negligible compared to sampling error.

   This is major topic, and has so far proved fairly intractable. At least part of the problem lies in the conventional assessment of product in terms of "percentage defective" with respect to a single analyte level (upper limit or lower limit) or range of analyte levels (upper and lower limits.) Investigation along alternative lines, such as considering different analyte levels in specifying producer's and consumer's risks, may be worthwhile.

5. Concerning the use of composite samples, the GGS also advises that, except for economic reasons that must be very well evaluated, preparing composite samples is not to be recommended given the loss of information on sample-to-sample variation within a lot. There are other implications of composite sampling that also need to be considered: it leads to less precise estimates of sampling variation, and increases the impact of measurement error of the repeatability type, compared to averaging an equal number of samples analysed separately. On the other hand, if there are economic reasons for reducing the number of analyses, it has been suggested that it may be possible to use composites but compensate for the loss of information by other means. This would require further work.

6. Stratified sampling is a method of obtaining a more precise estimate of a population characteristic (e.g. mean analyte level, percentage defects) where it is possible that the characteristic is distributed differently in different, and identifiable, parts of the population. The population may be, for instance, a lot or a consignment. Note that stratified sampling does not in itself imply that the various strata into which the lot is divided are then to be accepted or rejected individually, depending on the samples taken within each stratum. If in fact it is desired to formulate a decision rule that involves accepting/rejecting various combinations of sub-lots (corresponding to strata) in various circumstances, the statistical characteristics of the decision rules should be evaluated. In either case, CCMAS could decide on preparing guidance related to this issue.

THE WHOLE-OF-FOOD-CHAIN APPROACH

CCMAS is recommended to consider further work in order to provide practical guidance that will assist producers, manufacturers, and exporting and importing countries in the area of sampling, testing and the understanding of their relationship to producer's and consumer's risks. Possible items of work include:

[22] Note: It may be worthwhile to note that the tests of Section 5 of GGS (which is however put forward only for bulk materials), or even Section 4.4.2, can be applied for single sample assessments (which may be composite), provided that the relevant standard deviations are known and stable. Substantial investigations are however prescribed in Section 5 to assure this. To some extent, the single sample proposals seem orientated towards methods of conducting, or replacing, these investigations. However, as usual, the GGS methods are only applicable in the absence of significant measurement error.

7. Guidance on effective control of food production and manufacture to ensure conformity to regulatory standards upon receipt. The work could include consideration of processes in a state of statistical process control, processes under statistical control, and good practices.

8. Consideration of how to improve cooperation when implementing control and conformity assessment procedures. This should include taking advantage of procedures that are already considered well established, such as evaluating competence of laboratories, and transparency of procedures for conformity assessment and procedures for verification, particularly focused on non-stable or perishable products like unprocessed and minimally processed foods (fruit and vegetables; meat and meat products; cocoa; milk and milk products; fish and fish products; eggs and eggs products; and honey). FAO could be invited to cooperate with this effort.

9. Guidance for verification of food control procedures on the basis of sampling and testing, (as distinct from lot-by-lot conformity assessment covered by this DP), emphasizing managing risks in the medium to long term. Guidance would be needed on establishing levels of acceptable risk.

10. Guidance would also be needed for exporting countries to ensure that exported foods will meet requirements when subject to import inspection. Codex Recommended Codes of Practice (CAC/RCP) and the Standards (Codex Stan), updated as necessary, would be helpful in this respect. Controls on food production and manufacture (as noted in section 4.2.1) are an example of where such guidance could be developed.

11. Recommendations for risk management strategies by importing countries not exclusively reliant on border testing.

    The problem of obtaining adequate consumer's risk at reasonable cost (e.g. single sample assessment) suggests that such strategies need to be considered. It should also be noted that in many cases, it may not be possible to test all lots presented for import. Guidance is needed on how importing countries should use conformity assessment based on food testing in conjunction with other tools as part of a wider risk management strategy.

PRINCIPLES AND GUIDELINES FOR CONFORMITY ASSESSMENT AND RESOLUTION OF DISPUTES

12. This discussion paper prepared by an electronic working group has covered many points that could be developed into principles and guidelines. In summary, the document would set out principles and guidance on the determination, by sampling and testing, of the conformity of consignments of food in trade to official specifications, and provide guidance for governments on procedures to resolve disputes which arise between food control authorities about the implications of product test results on the status of a food consignment.

    CCMAS is recommended to consider new work on such a document.

REVIEW OF CRITERIA FOR SELECTION OF METHODS OF ANALYSIS

13. When considering alternative analytical test methods for use in a given CA procedure, the effect on fitness for purpose for that conformity assessment should be considered. The decision process on the use of alternative test methods should include consideration of whether, to maintain adequate producer's and consumer's risks, modifications to the sampling plan are necessary. The number of samples taken and the product acceptance criterion may both need to be reviewed. Codex procedures for the adoption of test methods and criteria for test methods could be considered in this light.

8 Bibliography

WTO - TBT – Agreement (http://www.wto.org/english/docs_e/legal_e/17-tbt_e.htm)

EURACHEM/EUROLAB/CITAC/Nordtest *Guide on the Estimation of Measurement Uncertainty Arising from Sampling*. Downloadable from: http://www.eurachem.org/guides/pdf/UfS_2007.pdf

Nordtest handbook for sampling planners on sampling quality assurance and uncertainty estimation *Uncertainty from sampling* (Based on the EURACHEM/EUROLAB/CITAC/Nordtest *Guide on the Estimation of Measurement Uncertainty Arising from Sampling)*. Downloadable as Report 604 from: http://www.nordicinnovation.net/nordtestfiler/tr604.pdf

Bakker, Arthur *et al*. 2008. "Statistical Inference at Work: Statistical Process Control as an Example." *Statistics Education Research Journal, 7(2), 130-145, http://www.stat.auckland.ac.nz/serj*

ANNEX 1.

COMMENTS ON PERFORMANCE CHARACTERISTICS OF TEST METHODS

It should be noted that in estimating the performance characteristics of a measurement method, we are in fact making measurements on that method. These measurements themselves will be subject to uncertainty, which may be considerable. Just as users of the analytical results need to have information on their uncertainty, so do users of a method's performance characteristics (e.g. repeatability, reproducibility) need to have information on the uncertainty of the measured performance characteristics. These latter users include those responsible for formulating methods of conformity assessment.

Documents so far presented to CCMAS on the topic of measurement uncertainty have appeared to treat it by Bayesian arguments, often using an "uninformative prior" for the true value of the measurand. The true value of the measurand is treated as though it were distributed about the measured result, according to some measure of relative plausibility, which is interpreted as a probability distribution, and used to make statements about the probability that a true value lies in a certain range, given the measured result. The difficulty is that this Bayesian probability will often have no resemblance to the relative frequency with which samples yielding the given measured result do in fact have true values within the range. The Bayesian probability quantifies a degree of belief, rather than a frequency of occurrence. This is suggested both in the original definition of measurement uncertainty[23] and the methods given for calculating it when the standard deviation of the distribution of measurement errors is unknown[24].

However, in formulating sampling schemes to control risks of wrong decisions such methods are inappropriate. What determines an exporter's profit and loss, and an importer's risks, is the frequencies with which compliant and non-compliant lots at various analyte levels are accepted and rejected, not the degree of belief someone has in the compliance or non-compliance of the various lots. To control risks of wrong decisions what is needed is information on the distribution of the measured result about the true value. We need to know "given that a lot is compliant, what is the probability that it will fail," not "given that it has failed, what is the (undefinable except in Bayesian terms) probability that it is compliant."

In order to deal with these issues it is recommended that CCMAS should (a) develop guidelines for estimating uncertainties related to the performance characteristics of a method; and (b) develop guidelines for a more appropriate approach to the treatment of measurement uncertainty, i.e. appropriate for formulating sampling schemes to control risks of wrong decisions. International standardization organizations could be invited to develop guidance on these issues.

---

[23] ***Measurement uncertainty:*** Non-negative parameter characterizing the dispersion of the values being attributed to a measurand, based on the information used. [Guidelines on Analytical Terminology (CAC/GL 72-2009), with reference to VIM, International Vocabulary of Metrology – Basic and general concepts and associated terms, 3rd edition, JCGM 200: 2008]

[24] E.g. EURACHEM/CITAC Guide CG 4, Quantifying Uncertainty in Analytical Measurement, section 8.3, Expanded uncertainty. Here it is recommended that the expanded measurement uncertainty be calculated from a sample standard deviation by multiplying by the relevant percentile of the Student's t distribution.

ANNEX 2.

COMMENTS ON "HOMOGENEOUS" GOODS

The meaning of homogeneous should be carefully understood here. The GGS covers cases where some items in a lot are compliant and some are not, without qualification, by means of simple attribute sampling plans25. Goods that are not homogeneous in that, say, one part of the lot has a large percentage of non-compliant items and another a small percentage can still be validly sampled using these plans provided that the sampling is random. It is only when sampling is non-random that an insistence on homogeneity is necessary. The GGS does not give permission for non-random sampling.

If it is known or suspected that different, identifiable, parts of the lot may be contaminated at different rates, sampling it as a single lot may not be the best approach. It may be better to treat each part as a separate lot, sampling the parts independently and making independent decisions on acceptance in relation to each part. But if you do decide to sample it as a single lot, under a pure attribute sampling plan you will achieve the risks associated with the plan. They will be stated in terms of the percentage non-conforming in the combined lot.

In addition to simple attribute sampling plans, further plans involving smaller sample sizes are given that are appropriate in cases where analyte levels are normally distributed within a lot. This technique is called attribute sampling by variables, or using the terminology used in the guidelines "Inspection by variables for percent non-conforming." The exclusion from the guidelines of lots that are not homogeneous presumably relates to lots that are going to be treated as having a normal distribution of analyte levels, but in fact do not, perhaps because different parts of the lot have different normal distributions, e.g. because they were manufactured in different runs. In such a case, the tests for control of the mean analyte level are still approximately valid, but the plans for attribute sampling by variable are not. This is because the attribute sampling-by-variables schemes estimate the percentage non-conforming items on the assumption that the true analyte level is normally distributed, whereas tests of the mean analyte level make this assumption only in respect of the mean of several samples taken at random from the lot.

[25] The word "simple" is used to distinguish these plans from those which postulate a normal (or other parametric) distribution of true analyte levels within a lot, attempting to estimate the "percent defective" from estimated parameters of the distribution (e.g. mean and SD). The GGS refers to the two different types of plan as "control of the percentage of defectives by attribute" and "control of the percentage of defects by variable" respectively.

ANNEX 3.

COMMENTS ON STRATEGIES FOR MANAGING CONSUMERS' RISK

It is a consequence of the way the universe is constructed and reflected in the international trade of food, that an importing country can't get a given level of assurance without taking the appropriate number of samples. It is also a consequence of human nature that a buyer is not prepared to pay for the samples unless there is evidence of real risk (which usually there is not.) The problem arises when the buyer insists on a low consumers' risk in respect of **each lot**.

Consider an extreme case where the importing country decides to take one sample from each lot and reject if it is non-compliant. The producers' risk may be acceptable (<5% probability of rejection if less than 5% of the lot is non-compliant.) But the consumers' risk is appalling, e.g. 50% chance of acceptance if 50% of the lot is non-compliant. But this strategy is adopted, and after, say 300 lots, none have been rejected, the importing country can be reasonably happy that **overall** less than 1% of the items were not OK, although there may well have been some lots in which the non-compliance rate was very high. So as long as things tick along with none of the lots being rejected, the importing country may well have an assurance of an average level of compliance with which it may be reasonably satisfied. Of course, some lots are rejected, it cannot have this level of confidence e.g. If 299 out of 300 are rejected, it certainly can't have any confidence in the one that snuck through. But presumably, after having, say, 5 lots in a row that failed, some action would have been taken.

This thinking leads to consideration of over-all strategy for managing average non-compliance rates over the medium to long term, rather than paying for a high premium for a high level of assurance on a lot-by-lot basis.

While the producers' risk must always be kept adequately low, it may be that in view of the high costs involved consumers may have to be satisfied with what seems a less than adequate control of risk in respect of individual lots, seeking rather to control the overall rate of non-compliance using a wider risk management strategy, in which lot-by-lot testing is just one of a number of tools that may be used.

This may seem to put the burden of accepting high risk on the side of consumer. However it would not be justifiable, at least in normal circumstances, to adopt any procedure that gives compliant product a high probability of rejection. If plans which share out the risk equally, are too expensive to implement because of the number of samples required, the consumer has other strategies available to manage the risk. The producer does not, other than by working to an AQL, or a maximum analyte level, well below the nominal one. In such a case, the "true" AQL to which the producer is being required to work should be made explicit.

ANNEX 4.

NOTES ON THE EFFECT OF CLASSIFICATION ERROR, AND IMPRECISION IN ITS ESTIMATION, IN SIMPLE ATTRIBUTE SAMPLING SCHEMES

J.H. Jowett, 27-August-10

Summary

The setup considered in this Annex will be a considerable over-simplification for many of the scenarios arising in the compliance testing of foods. However, it brings out several points which apply more generally, including:

- The serious impact of even low rates of false positives when substantial numbers of samples are taken;
- The high potential impact on risks when the true values of parameters associated with the distribution of measurement errors are not accurately known, and the need to allow for "worst case" estimates of such to avoid serious under-statement of risks;
- The difficulties inherent in situations where it is unacceptable to accept lots in which any apparently defective items are found.

We have considered three cases:

1) There are no classification errors. The sampled items are correctly classified without any risk of error.
2) There are risks of misclassification, but the associated probabilities are exactly known.
3) There are risks of misclassification, but the associated probabilities are known only approximately, although possibly with high precision.

The first two cases probably represent an ideal situation in most circumstances. They are dealt with separately to present the argument in a step by step manner, and to illustrate the increase in sample size as each factor is successively taken into account.

Introduction

Attribute sampling is sampling to estimate or control the "percentage defective" in a lot. Within this category there are two main types of sampling: the type where a "defective" is defined in terms of the concentration of some analyte, and the type where the distinction is essentially qualitative. In the former type of sampling measurement uncertainty gives rise to the possibility of false positives and false negatives. This possibility of misclassification is also present when the distinction is purely qualitative. Such a misclassification occurs when the classification given to a sampled item during inspection differs from the "true value" of this classification: that is, the classification that would be given to an item if it were inspected closely enough, or with enough precision, to determine its status unambiguously. This type of error can be considered as arising from measurement uncertainty in a general sense.

For control of percentage defectives in the qualitative case type the *General Guidelines on Sampling* give plans for "Inspection by Attribute" in which a specified number of samples is taken and the product accepted or rejected according to the number of defectives found.

The purpose of this paper is to examine the effect of false positives and negatives on producer's and consumer's risk, and the implications on sample sizes, for this type of sampling plan. The papers that CCMAS has so far considered restrict themselves explicitly to cases where a continuous variable is being measured, and this, together with some of the assumptions underlying our analysis restrict the scope for its direct application. A yes/no criterion defined by whether a continuous variable falls within a specified range preferably needs a more complex statistical analysis than that given here. However, some features that have come through may be expected to persist when more general types of sampling plan are considered. Not the least of these is the dramatic impact on sample sizes of even modest uncertainty regarding the true rates at which false positives and false negatives occur.

Simple attribute sampling

We consider a scheme where a certain number of items (n) is taken at random, and the lot rejected if the number of "defective" items, or "positives" exceeds the "acceptance number" (c).

We refer (using statistical terminology) to the entire group of items taken as "the sample" and the number of items (n) as the "sample size." The individual items selected will be described as "sampled items." While in practice the definition of a defective item may involve the (true) analyte level in the sampled item, in the case of simple attribute sampling this complication is ignored: defective items are effectively taken to be indistinguishable, as are non-defective items. The sole characteristic of an item that is of relevance is whether it is defective or non-defective.

Classification errors may need to be taken into account when classifying items in that there may be a non-zero probability that a non-defective item may mistakenly be classified as defective (a false positive) or that a defective item may be mistakenly classified as non-defective (a false negative). In this paper we consider only an idealised case: the probability of misclassification of a positive is the same for all positives, and the probability of misclassification of a negative is the same for all negatives. Thus the probability of acceptance of a lot is the same for all lots containing the same numbers of positives and negatives. This is of course implausible when the distinction is based on analyte levels, in which case the probability of misclassification will depend not only on how many positives and negatives are present, but also how close to borderline they are. The same may be true even when essentially qualitative characteristics are considered: some cases may be clear cut, some borderline. But at least the simplified model gives some indication of the impact on sample size if false positives and negatives may occur, and their impact on consumers' and producer's risks.

Classification error is the analogue of measurement error, in a situation where "measurement" leads to an estimate of which of two categories (defective or non-defective) an item belongs to, rather than an estimate on a continuous scale.

Case 1): No classification error

This is the standard acceptance sampling scenario.

The selection of an appropriate scheme (choice of n and c) is made by requiring a low probability of rejection where the proportion of defective items in the lot is less than a certain proportion known as the AQL (acceptable quality level), and a low probability of acceptance where the proportion of defectives exceeds a certain higher proportion known as the LQ (limiting quality).

p : the (unknown) proportion of true defectives in the lot.

p_0 : the AQL.

p_1 : the LQ.

α : the probability of *rejection* when $p = p_0$. This is called the "producer's risk."

β : the probability of *acceptance* when $p = p_1$. This is called the "consumer's risk."

Specification of p_0, α, p_1 and β is sufficient to determine n and c uniquely. Large sample sizes are required if p_0 and p_1 are close together, or if the risks α and β are very small. For a given n and c, the probability of acceptance is a function of p. This function is known as the operating characteristic (OC).

Other more complex schemes (e.g. two phase sampling, sequential sampling) are possible, and other measures of performance (e.g. the AOQL or average outgoing quality limit) can be used. The theory is well developed.

In the remainder of the paper we will consider three sampling schemes as examples, so that the impact of classification errors can be shown.

Risk Specification 1.

$p_0 = 5\%, \alpha = 5\%, p_1 = 10\%, \beta = 10\%$. In words, as long as the producer provides lots with less than 5% defectives, his risk of rejection is 5% or less. If the percentage of defectives rises above 10%, there is a 90% chance of rejection.

Sampling scheme: For this rather tight specification, n=233, c=16.

Risk Specification 2.

$p_0 = 1\%, \alpha = 5\%, p_1 = 5\%, \beta = 10\%$: Sampling scheme n=132, c=3

Risk Specification 3.

$p_0 = 0, \alpha = ?, p_1 = 5\%, \beta = 10\%$: Sampling scheme n=45, c=0

This is a special case, where any positive at all in the lot is unacceptable. For instance, a single defect could cause very serious problems to the consumer, as in melamine contamination in baby food. As there is assumed to be no classification error, then a producer who succeeds in producing a lot at the AQL of zero, (i.e. with no defectives at all) can be assumed to run a zero risk of rejection, and only the consumer's side of the specification needs to be considered and agreed by the parties. This situation changes if there is any risk of a false positive. Then the producer's risk becomes non-zero, and must be controlled.

Case 2): Classification error with probabilities of false positives and negatives exactly known.

We use the following notation:

p_F : probability of measured positive for a true negative

In the long run, a proportion p_F of true negatives are misclassified as positive[26].

q_F : probability of measured negative for a true positive

In the long run, a proportion q_F of true positives are misclassified as negative.

We assume that p_F and q_F are known exactly.

For the classification to have any value at all, a true positive must have a better chance of being classified as positive than a true negative, so $1 - q_F > p_F$, that is,

$p_F + q_F < 1$.

We will not know how many of the observed "positives" in the sample are true positives. We will only know how many sampled items were classified (possibly mistakenly) as positive and negative. Essentially the problem is to reject a lot only when the number of items classified positive exceeds that which would be expected from classification errors alone.

If the rate of true positives is *p*, the probability p' that a randomly sampled item will be classified as positive is

$p' = p(1 - q_F) + (1 - p)p_F$.

Thus the sampling scheme has to reject with probability at most α when

$p' = p'_0 = p_0(1 - q_F) + (1 - p_0)p_F$

and accept with probability at most β when

$p' = p'_1 = p_1(1 - q_F) + (1 - p_1)p_F$

Here we use p'_0 and p'_1 to represent the probabilities of an observed positive when the probabilities of a true positive are p_0 and p_1 respectively.

The condition $p_F + q_F < 1$ guarantees that $p'_0 < p'_1$.

However, $p'_1 - p'_0 = (1 - p_F - q_F)(p_1 - p_0)$, so the effective separation between the AQL and LQ is reduced by the classification errors, and larger sample sizes would be expected to be required.

To illustrate this we consider the following classification error scenario, and its effect on the three schemes already considered: $p_F = 1\%$, $q_F = 10\%$.

Before doing this, we shall have to specify a non-zero producer's risk for specification 3. Obviously, in 45 samples of true negatives, there will be a probability approaching 0.45[27] of at least one false positive, leading to a producer's risk that is definitely unacceptable even when $p = 0$. There are two options:

i) reduce the sample size, for example to n=5 so that there is only a 5% risk of a false positive, or to n=1 so that there only a 1% chance, or

---

[26] This re-formulation of the definition is to make clear the type of probability being considered: it is a classical probability defined in terms of long run proportions (relative frequencies) and is conditioned on the measured sample being a true negative.

[27] The exact figure for this probability is calculated as $1 - (1 - p_F)^{45} = 0.36$ when $p_F = 0.01$.

ii) increase the sample size, but use a non-zero value of c. This may not be possible if the defect is one with severe implications.

Option i) will of course lead to very large consumer's risks, option ii) will sometimes be undesirable in principle.

We must also consider that there is no way a manufacturer can do better than produce no defectives at all, so that in specification 3 whatever producer's risk we settle on for $p_0 = 0$ is inescapable. Even in perfect product, a producer is stuck with it. This suggests that producer's risks around 5% may not really be appropriate in this scenario, especially in cases where the impact of a failure on the producer is severe. In cases where c=0 must be retained, to be fit for purpose, that is to be able to deliver a low producer's risk and at the same time a reasonable consumer's risk, a measurement method must have a very low rate of false positives. For example, for a sample size of 45 and a producer's risk of 1% we must have $(1 - p_F)^{45} > 0.99$, that is $p_F < 0.022\%$, or less than1 false positive per 4,500 true negatives. If "every item must comply" is taken literally, false positive rates must be kept extremely low.

For the purpose of continuing the example with $p_F = 1\%$, $q_F = 10\%$, we assume that option ii) is feasible, and set the producer's risk at 5%.

The consequences of this classification error scenario on the three risk specifications are given below.

Table 1: Sampling schemes with and without classification error

Risk specification	p_0	p_1	No classification error		With classification error: $p_F = 1\%$, $q_F = 10\%$			
			n	c	p_0'	p_1'	n	c
$\alpha = 5\%, \beta = 10\%$	5%	10%	233	16	5.45%	9.90%	293	22
$\alpha = 5\%, \beta = 10\%$	1%	5%	132	3	1.89%	5.45%	214	7
$\alpha = 5\%, \beta = 10\%$	0	5%	45	0	1.00%	5.45%	121	4

Case 3): If the false positive and false negative rates are not exactly known.

Consider the following scenario:

The false positive probability p_F is estimated by testing 1000 known negatives. 10 false positives are found. From 1000 known positives, 100 false negatives are obtained. This gives *estimates* of p_F and q_F, $\hat{p}_F = 0.01$, $\hat{q}_F = 0.10$.

Upper and lower one-sided 95% confidence limits for the true values of p_F and q_F are:

	Lower limit	Upper limit
False positive rate: p_F	0.0054	0.0169
False negative rate: q_F	0.0848	0.1170

The uncertainty regarding the true values of p_F and q_F has serious implications. For example, we showed above that the scheme n=214, c=7 delivered a producer's risk of 5% when $p_0 = 0.01$ under the classification error scenario $p_F = 1\%$, $q_F = 10\%$. What if p_F were not 1%, but 1.69%, a value just consistent, at the 95% level of confidence, with the experimental data? Then p_0' would rise from 0.0189 to 0.0257, and the producer's risk would rise to 18.2%, an unacceptable level. It could be argued that the producer may be able to adjust for this, over the long term at least, by manufacturing better product. But it would have to be a great deal better. To bring his risk back down to 5%, he would need $p_0' = 0.0189$, and with $p_F = 0.0169$ and $q_F = 0.10$ this means $p_0 = 0.0023$. In other words, while claiming to enforce an AQL of 1% we would actually be enforcing one of 0.2%. This sort of thing needs to be avoided. If things were just a little worse, the producer may be unable to avoid a risk substantially above 5% even with perfect product.

(This example is fairly extreme, because the AQL is comparable with the rate of false positives. Improvement of quality thus has a limited impact on the number of observed positives. In this connection a policy of setting cut-offs based on a 1% rate of false positives may need to be reconsidered.)

It is clear from this example that it is undesirable to base a testing scheme on simple estimates of p_F and q_F and assume that these will give approximately the correct risks. We need to consider the imprecision with which p_F and q_F are estimated and allow for worst case scenarios, so that producers and consumers are not inadvertently exposed to absurd risks.

This means considering the range of true values of p_0' and p_1' that are consistent with the experimental data used to estimate p_F and q_F. It is unsatisfactory to simply "estimate" that the producer's risk is "about" 5%: as we have seen, it may not be anything like it. We need to be **confident,** on the basis of the experimental data, that the producer's risk will be at most 5%, and that the consumer's risk will be at most 10%.

In considering the producer's risk, we have to allow for the fact that he may suffer from false positives more than we have allowed for, and benefit from false negatives less. We need to find the maximum plausible value for p_0' consistent with a true rate of p_0 under the type of uncertainty regarding p_F and q_F shown in the table above. In other words, to find an upper 95% confidence limit for p_0'. This is not a problem with an exact solution, even under the simple experimental scenario described. However, denoting upper and lower limits by the subscripts U and L respectively, it seems reasonable to set the upper limit for p_0' at

$$p_{0U}' = \hat{p}_0 + \sqrt{p_0^2 (q_{FL} - \hat{q}_F)^2 + (1 - p_0)^2 (p_{FU} - \hat{p}_F)^2} \quad .$$

> Note: This is the same sort of approximation that is used in the modified large sample (MLS) estimation of variance components. It takes the difference between the relevant confidence limit and the estimate itself as a type of "expanded uncertainty", and combines these uncertainties using a root-mean-square procedure. It will be exact when the estimates involved are normally distributed, or when one of the two terms is negligible compared to the other. In other circumstances it seems to me likely to be conservative (i.e. to over-estimate an upper limit and under-estimate a lower one) as it does not allow for the approach to normality when the two estimates are combined. In fact it was found to be conservative in the MLS procedure. The problem is equivalent to that of finding an upper confidence limit for the prevalence of a rare characteristic using a stratified sample with different sampling fractions in each stratum; there has been work done on this problem but I am not familiar with it.

For the consumer's risk we must allow for the smallest value of p_1' consistent with the table, resulting in

$$p_{1L}' = \hat{p}_1 - \sqrt{p_1^2 (q_{FU} - \hat{q}_F)^2 + (1 - p_1)^2 (p_{FL} - \hat{p}_F)^2}$$

Applying these considerations to the risk scenarios we are considering yields:

Risk specification	p_0	p_1	No classification error		With estimated classification error: $\hat{p}_F = 10/1000$, $\hat{q}_F = 100/1000$			
			n	c	p_{0U}'	p_{1L}'	n	c
$\alpha = 5\%, \beta = 10\%$	5%	10%	233	16	6.11%	9.45%	529	41
$\alpha = 5\%, \beta = 10\%$	1%	5%	132	3	2.57%	5.00%	515	19
$\alpha = 5\%, \beta = 10\%$	0	5%	45	0	1.69%	5.00%	234	7

The schemes shown in this table give 95% confidence that the producer's risk specification will be met, and 95% confidence that the consumer's risk specification will be met. There would certainly be another way of approaching the problem. For instance, one can consider what risk specifications could be met, given appropriate sample sizes.

ANNEX 5.

EFFECTS OF BIAS, BETWEEN-LABORATORY VARIATION AND REPEATABILITY VARIATION

It may be useful to illustrate the different effects of bias, between-laboratory (or, where the laboratory in held constant, between-run) variation and repeatability variation as follows.

Repeatability type variation results in white noise being added to the true values.

Adding in bias shifts every measurement, in every lot tested, upward or downward by the same amount. The effect is the same as if all lots were tested against the wrong compliance limit.

Adding between-laboratory variation results in lots being tested against a compliance limit that is wrong, but which varies from lot to lot about the wrong compliance limit set by the bias.

CONFORMITY ASSESSMENT IN THE PRESENCE OF SIGNIFICANT MEASUREMENT UNCERTAINTY

Part of the difficulty is the form in which consumer's and producer's risk tend to be specified, and also the different impact that the various components of measurement uncertainty have.

For example, a common type of risk specification is to specify that product of which a certain fraction is below the limit should have a certain probability of rejection, say 5%, and product of which a certain fraction is above the limit should have a 90% probability of rejection; the limit being the same in both cases. Measurement error of the between-laboratories type in effect randomly moves the compliance limit being tested against around from run to run. In the presence of significant measurement uncertainty, it is clearly impossible to distinguish product of which 100% is just below the limit, from product of which 100% is just above the limit. In food testing it does not seem sensible to try to do so anyway because the food is essentially the same in the two situations, at least when this involves significant cost. Yet the risk specification requires that the first type of product should have a 5% probability of rejection and the second a 90% probability of rejection. This is clearly impossible, given finite resources.

The resolution of this difficulty may come from a recognition that the true analyte level which is being effectively tested against does in fact move around from lot to lot: it is not always equal to the "legal limit". If reasonable limits can be assigned to this potential movement, it may be possible to give assurances to producer and consumer, but in terms of two different analyte levels. Particularly where there is a significant margin between the analyte levels that can be reliably achieved by good manufacturing practice and those that will pose significant hazard to a consumer, this may be an acceptable situation. It may however involve a change in attitude, both by producers and consumers, in the interpretation of a "specification limit."

Given this, the range in which the "de facto" limit being tested against could reasonably be expected to lie could be estimated as part of the method assessment, perhaps expressed as a statistical tolerance interval. This is illustrated further in Annex 7.

ANNEX 7.

SIMPLIFIED MEASUREMENT ERROR SCENARIO ILLUSTRATING FITNESS FOR PURPOSE CONSIDERATIONS

Summary

This annex shows the effect of bias and between-laboratories type variation, in separating the analyte levels at which given consumer's and producer's risks apply, in the simplest situation, that of completely uniform product. The size of this separation is relevant in judging the fitness for purpose of an analytical method. The modifications needed when the characteristics of the measurement error distribution are not precisely known are briefly discussed

The simplified scenario

The product is all at the same analyte level. It is required to accept the product with probability at least 95% if this is less than some specified analyte level L_P and reject it with probability at least 90% if this exceeds another specified level L_C.

L_p is the analyte level at which a producer's risk of 5% is required;

L_c is the analyte level at which the consumer's risk 10% is required.

The only measurement error is a randomly varying laboratory bias normally distributed with mean μ and standard deviation σ_L, both of which are known exactly.

We need to choose a cut-off x_C for the measured analyte level so that both producer's and consumer's risk specifications are met.

From tables of the normal distribution, we see that the producer's risk specification will be met if $x_C > L_P + \mu + 1.645\sigma_L$. The consumer's risk specification will be met if $x_c < L_C + \mu - 1.282\sigma_L$. A suitable cut-off, meeting both requirements will be available when both inequalities are satisfied. This requires that

$L_P + \mu + 1.645\sigma_L < L_C + \mu - 1.282\sigma_L$, that is $\sigma_L < \dfrac{L_C - L_P}{1.645 + 1.282} = 0.342(L_C - L_P)$.

If this inequality is not satisfied, the method would be deemed not fit for purpose; we need one with a smaller σ. However, this may not be a final decision. It may be appropriate in the context to raise L_C or lower L_P. This would require further consideration of the likely impact on consumer or producer respectively.

While this example is unrealistic, it may be valuable because it illustrates a best case. Any additional variation (repeatability error, uncertainty about the true values of μ and σ_L, or variation in analyte level within the lot) will only increase the uncertainty, requiring a larger difference between L_C and L_P.

Effect of uncertainty concerning μ and σ_L

These will almost never be exactly known. This gives rise to uncertainty about the range in which x_C should be located. Suppose M_P and M_C are upper 95% confidence limits for $\mu + 1.645\sigma$ and $-\mu + 1.645\sigma$ respectively, so that that it could be stated with 95% confidence that $\mu + 1.645\sigma_L$ were less than M_P, and with 95% confidence that $-\mu + 1.282\sigma_L$ were less than M_C. If 95% confidence were deemed adequate assurance, a suitable cut-off could be found in the range: $L_P + M_P < x_C < L_C - M_C$, provided that $M_C + M_P < L_C - L_P$. Otherwise the fitness for purpose of the method would be questionable.

DRAFT

PROJECT DOCUMENT FOR NEW WORK ON THE DEVELOPMENT OF "PRINCIPLES AND GUIDELINES FOR CONFORMITY ASSESSMENT BASED ON SAMPLING AND TESTING OF FOODS IN TRADE AND IMPLICATIONS FOR RESOLUTION OF DISPUTES"

1. Purpose and scope of the new work

The purpose of the new work is to develop a new document, "Principles and Guidelines for Conformity Assessment Based on Sampling and Testing of Foods in Trade and Implications for Resolution of Disputes", in order to provide general principles and guidance for governments and Codex committees on these issues.

The scope of the new work will be the same as the discussion paper on the subject presented to the 32nd session of CCMAS. It will set out principles and guidance on the determination, by sampling and testing, of the conformity of consignments of food in trade to official specifications, and provide guidance for governments on procedures to resolve disputes which arise between food control authorities about the implications of food product sampling and test results on the status of a food consignment.

2. Relevance and timeliness

Many food standards include specifications that are verified by sampling and testing. Such specifications need to include procedures for conformity assessment, so that the level protection afforded to consumers and the risks carried by producers are known. Without such procedures both importing and exporting countries may use *ad hoc* procedures, and disputes are more likely to occur and be difficult to resolve.

Several Codex committees responsible for such specifications have developed conformity assessment procedures, using different approaches, but in many cases the procedures are lacking. CCMAS has developed elements of what is needed for conformity assessment, particularly methods of analysis and sampling plans, and it has developed guidelines (GL 70) for settling a specific type of dispute, namely over analytical (test) results. A more comprehensive framework is needed to set out the principles of conformity assessment based on Sampling and Testing, approaches that can be recommended, and effective general procedures for resolving disputes when they occur.

A discussion paper prepared by an electronic working group has covered the points that need to be included in the principles and guidelines.

3. Main aspects to be covered

The principles and guidelines will cover the principles for conformity assessment, based on sampling and testing of foods in trade, and implications for resolution of disputes.

The principles would include:

- Prior agreement on conformity assessment procedures and a dispute resolution process.
- The chosen CA procedure should control the producer's and consumer's risks in a way that does not penalize a compliant product with a high risk of rejection while providing adequate consumer protection.

The guidelines will provide advice on practical matters such as:

- Choice of an appropriate sampling procedure
- Allowing for measurement uncertainty and sampling uncertainty
- Considerations on producers´ and consumers´ risks when setting product specifications and choosing methods of analysis
- Preventive measures in exporting countries to ensure exported foods meet requirements, and risk management strategy in importing countries to avoid sole reliance on border testing
- The steps involved in a dispute resolution process, from initiating the dispute, through identifying causes, to its resolution.

4. Assessment against the Criteria for the establishment of work priorities

General criterion

This work is directed towards both consumer protection and ensuring fair trading practices, according to provisions in product specifications that are verified on the basis of sampling and test results. The work will ensure that conformity assessment procedures and dispute resolution processes provide appropriate and clear levels of protection for consumers, and treat both importing and exporting countries fairly, pointing out also the importance of the inherent risks in these activities. The guidance will assist developing and developed countries in establishing appropriate conformity assessment procedures for both imported and exported food, and in resolution of disputes. On a global scale the work will contribute to reduction of adverse human health effects through appropriate control of food-borne risks, will advance fair trade, and will help to avoid costly, disruptive and wasteful disputes.

Criteria applicable to general subjects

(a) Diversification of national legislations and apparent resultant or potential impediments to international trade. This new work will assist all member countries in establishing appropriate conformity assessment procedures for both imported and exported food, resulting in more effective conformity assessment, less likelihood of disputes, and more effective resolution of disputes when they occur.

(b) Scope of work and establishment of priorities between the various sections of the work. The work will provide Codex with a new document that demonstrates the implications of conformity assessment procedures on the resolution of disputes and how to balance the producers´ and consumers´ risks to guarantee fair trade practices and appropriate protection of consumers´ heath. The work will provide a comprehensive framework for a whole-food-chain approach as a strategy for effective conformity assessment and resolution of disputes in international trade. The work should be relatively simple since a great amount has already been done by the eWG. Both aspects of the work, conformity assessment and resolution of disputes, are equally important.

(c) Work already undertaken by other international organizations in this field and/or suggested by the relevant international intergovernmental body(ies). This work does not duplicate any work already undertaken by other international organizations. The eWG has taken account of relevant work by other international organizations on conformity assessment, measurement uncertainty and uncertainty of sampling. The work arises out of material developed by the eWG.

5. Relevance to the Codex strategic objectives

The proposed work contributes to all five goals of the Codex Strategic Plan 2008-2013.

Goal 1: Promoting sound regulatory frameworks

This work emphasizes a horizontal approach to conformity assessment and resolution of disputes, and is not overly prescriptive nor more trade restrictive than necessary, while respecting the basic objectives of the Codex. It takes into consideration the technical and economic implications for all members as well as the special needs of developing countries including infrastructure, resources and technical and legal capabilities.

Goal 2: Promoting widest and consistent application of scientific principles and risk analysis

This work applies scientific principles to conformity assessment and resolution of disputes, focusing on sampling and testing of foods. The guidelines procedures are based on the need to manage the risks of making wrong compliance decisions, and aims to be consistent with the *Working Principles for Risk Analysis for Application in the Framework of the Codex Alimentarius*.

Goal 3: Strengthening Codex work-management capabilities

The principles and guidelines will streamline Codex work by providing guidance for all Codex committees that develop specifications for food that are verified by sampling and testing.

Goal 4: Promoting cooperation between Codex and relevant international organizations

Codex will be able to assist the work of other international bodies by clarifying what is needed for conformity assessment of food and resolution of disputes.

Goal 5: Promoting maximum and effective participation of Members

The new work affects all members of Codex and may trigger further participation of both developing and developed countries with interests in trade of foods and food ingredients. (It is noted that 22 member

countries from the five continents and three international organizations with the status of observers, participated in the eWG that developed the discussion paper that will form the basis for this proposed new work.)

6. Information on the relation between the proposal and other existing Codex documents

The eWG has reviewed a wide range of Codex documents that include information on conformity assessment and resolution of disputes. This includes the *Procedural Manual*; *General Guidelines on Sampling* and the *Guidelines for Settling Disputes over Analytical (Test) Results* prepared by CCMAS; guidelines on food control systems and their operation, prepared by CCFICS; guidance on conformity assessment prepared by CCRVDF, CCPR, CCCF and CCFH; and specifications prepared by commodity committees.

The new work develops beyond these existing Codex documents by providing overall horizontal principles and guidance, while referring to existing documents where appropriate.

7. Identification of any requirement for and availability of expert scientific advice

The new work can be undertaken without a major effort related to scientific advice, since most of the advice has already been presented in the eWG discussion paper. Some aspects will benefit from further development from targeted discussion in a physical or electronic working group. Additional scientific advice could be provided by international organizations occupying themselves with standardization.

8. Identification of any need for technical input to the standard from external bodies so that this can be planned for

None identified.

9. Proposed time-line for completion of the new work

The following timeline is proposed for the completion of the work, preferably for final adoption in 2013. The timeline should not exceed five years (2014).

Timetable	Meeting	Progress
March 2011	32nd session of CCMAS	Agree on project document and submit it to 34th CAC for approval of new work
July 2011	34th CAC	Approval of new work
		Preparation of Proposed Draft Principles and Guidelines. Circulation for comments at step 3.
March 2012	33rd session of CCMAS	Consideration of the Proposed Draft Principles and Guidelines and advance to 35th CAC for adoption at step 5
July 2012	35th CAC	Adoption at step 5
		Circulation for comments at step 6
March 2013	34th session of CCMAS	Consideration of the Draft Principles and Guidelines at step 7 and advance for adoption at step 8
July 2013	36th CAC	Final adoption.