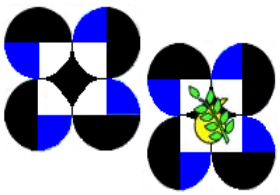
# FNRI Proficiency Testing Scheme

## Feb 2016
## 3rd Edition

# PROTOCOL

**LEAH C. DAJAY**
PTL Head

**MARIO V. CAPANZANA, Ph.D.**
Director



Republic of the Philippines
Department of Science and Technology
**FOOD AND NUTRITION RESEARCH INSTITUTE**
Gen. Santos Ave., Bicutan, Taguig City, Philippines
Website: www.fnri.dost.gov.ph
Tel. Nos.: (632)837-2934; 837-2071 to 82; Fax No.: (632) 837-3164

## PREFACE

This document provides an overview of the FNRI Proficiency Testing (PT) Rounds with the principles and procedures applicable to the organization of the FNRI PT. It also explains how laboratory performance is evaluated. The document does not attempt to cover each step in the proficiency testing process. These are discussed in FNRI's internal procedures, which are in compliance with the requirements of the ISO/IEC 17043:2010.

The Protocol should be read in conjunction with the FNRI Proficiency Testing Supplement on Statistical Procedures which describes the statistical procedures for data analysis and performance assessment based on ISO 13528:2015 and the International Harmonized Protocol for the Proficiency Testing of Analytical Chemical Laboratories (IUPAC Technical Report: 2006 IUPAC).

**CONTENTS**

**Page**

# 1.0 INTRODUCTION

The Food and Nutrition Research Institute (FNRI) is a research and development institute under the Department of Science and Technology (DOST) of the Philippine Government. The Institute provides relevant and sustainable technologies and scientific information on food and nutrition. The FNRI-Proficiency Testing Laboratory (PTL) is a pioneer in providing PT in food on nutrition labeling components and aims to contribute in the development and implementation of an analytical quality assurance (AQA) system.

# 2.0 PROFICIENCY TESTING

Proficiency Testing is an essential element of the quality assurance of laboratories accredited to ISO/IEC 17025:2005 **[1]** to establish technical competence of PT participant laboratories.

Proficiency testing provides an independent means of testing and comparing individual performance of participant laboratories against pre-established criteria by interlaboratory comparison of test results **[2]**. It is way of checking the accuracy of results from laboratories **[3].** It involves the distribution of PT materials to participating laboratories for PT participants to analyze measurands at a prescribed period of analysis, and submit results to the PT provider at a specified date. Proficiency testing may also be referred to as "external quality assessment" (EQA), "round robins", or "ring trials."

With the increasing demands for independent proof of competence from regulatory bodies and customers, proficiency testing is relevant to all food testing laboratories for quality and safety purposes.

# 3.0 QUALITY MANAGEMENT OF PROFICIENCY TESTING SCHEME

## 3.1 Operation

The FNRI PT Rounds are operated according to the principles defined in the ISO/IEC 17043:2010 standard, *"Conformity Assessment – General Requirements for Proficiency Testing"*. PTL also implements and maintains a quality management system on recognized management standards, ISO/IEC 17043:2010 **[2]**, and the ISO 13528:2005 *"Statistical Methods for Use in Proficiency Testing by Interlaboratory Comparison"* **[4]** and IUPAC Technical Report (2006) *"The International Harmonized Protocol for Proficiency Testing of Analytical Chemistry Laboratories"* **[5]** for statistical analysis.

An Advisory Committee is established for every FNRI PT Round, which is composed of the Technical Working Group Members who are chemists and of Consultants –Chemist/PT Provision Specialist and Statistician. Members of the Advisory Committee are selected on the basis of their qualifications and expertise, and not their affiliations. The Advisory Committee recommends and assists in the decision-making in every phase or activity of the Scheme and discusses the scientific issues

arising from the conduct of PT. During discussion, the FNRI PTL staff does not disclose any information pertaining to the PT participant, only scientific information is exchanged. Communications with the Advisory Committee are made through telephone or e-mail and minutes are prepared for every Advisory Committee Meeting.

The day to day running of the PT scheme, including the preparation of PT materials and conduct of PT is the responsibility of the PTL Head. He/she is an expert in the provision of FNRI proficiency testing services. To supplement this expertise, regular updating of knowledge through attendance to trainings/seminar-workshops, study of updated PT references, and consultation with local and foreign experts.

## 3.2 Participation

The FNRI PT Rounds were aimed at providing the participants with the basis to evaluate their laboratory performance in the analysis of nutrition labeling components through an interlaboratory comparison. The PT Rounds were also intended to assist participants in conducting self-help investigative, corrective, and preventive actions to improve their laboratory performance.

The FNRI PT Rounds are made available on the website, www.fnri.dost.gov.ph. Local and foreign PT participant laboratories are invited to participate one (1) month before distribution of a homogeneous and stable PT material. A call for participation, containing information of proficiency test item, measurand, PT timetable, and cost of PT participation is sent through e-mail/fax and the PT participant returns a Registration Form with the following information: a) laboratory contact person; b) address, fax and phone number; c)analytes/measurands to be tested.

The PT participation fee is expected to be received before distribution of PT materials. FNRI PTL reserves the right to withhold distribution of proficiency test item and/or Final PT Reports from participants if payment is not yet settled.

## 3.3 Confidentiality

Each PT Report contains all results and method details, including performance score, of every PT participant in the Round but individual laboratories are not identified within the main body of the PT Report. Each participant is provided with a unique laboratory code number, which identifies their results and all data they submitted to the PT provider. This laboratory code is used throughout the period of the Round to ensure their confidentiality.

Information on the performance of a participant laboratory will be confidential to that laboratory, although in some instances, an agreement may exist for participants' results to be made known to the Philippine Accreditation Bureau (PAB), if requested. This will be done

with the knowledge and consent of the PT participant. A participant laboratory will not have access to the details of other participants.

## 3.4 Typical Timetable

Organization of FNRI PT Round is done at least once a year. The deadline for responding to invitation, date of PT material distribution, and reporting of results are indicated in the Call for Participation/Invitation to Participate.

For each PT Round, the organizer sets the deadline for the return of results or specifies a closure date, wherein after which, results will not be accepted. This is to ensure that sufficient time is available to complete the test and report the results in time for the set timeframe. Participants are informed by email when any delay arise in the schedule.

A typical FNRI PT Round list of activities is as follows:

a. Preparation of proficiency test item;
b. Conduct of homogeneity and stability tests;
c. Call for Participation and Registration of participants;
d. Dispatch of proficiency test item;
e. Analysis of proficiency test item by the participant laboratories and reporting of results to the PT provider – generally, the closing date is one month from the date of dispatch of proficiency test item, but time scale may be reduced in instances of potentially unstable analyte/measurand and/or matrix;
f. Collation, screening and verification, and tabulation of results;
g. Statistical evaluation of PT results;
h. Sending of Interim Report to the PT participants;
i. Preparation of the PT report;
j. Distribution of the PT Report to PT participants; and
k. Conduct of post-PT Meeting/Seminar-Workshop.

## 3.5 Subcontracting

The FNRI PTL does not subcontract proficiency test item l preparation to while homogeneity and stability testing are done by qualified subcontracting laboratory, evaluated based on their accreditation, precision of results, performance in an applicable PT Round and turn-around time.

In necessary instances where additional technical competence and advice in PT operation are needed, the PTL Staff seeks technical assistance from the members of the Advisory Committee (Consultants and members of the Technical Working Group, TWG), particularly on the areas of statistical evaluation of laboratory performance, technical comments on methods used by the participant laboratory, and preparation of the PT Report.

# 4.0 TEST MATERIAL PREPARATION AND QUALITY CONTROL

## 4.1 Choice of PT materials

The criteria for choosing the test materials are presented to the FNRI PT Advisory Committee for comments. The PT material, which resembles the routine food samples tested in laboratories are typically produced by the Provider. The PT material with assigned values from the consensus of PT participants' results will be used as Quality Control Test Material (QCTM)/ Reference Material (RM) for interested laboratories.

## 4.2 Test material supply

The proficiency test item is collected/ purchased from a market/wet market or is requested from a local food manufacturing company. The proficiency test item collected should be of the same species for unprocessed foods and from the same batch and/or lot number and date of manufacture for processed foods.

## 4.3 Preparation of test materials

The preparation of the proficiency test item is conducted in FNRI's designated Proficiency Testing Laboratory that is maintained clean and temperature-monitored. Preparation of proficiency test item is specific to the type of matrix used and the relevant measurand. Test materials are usually prepared in bulk and then divided into individual sub-samples.

The proficiency test item is described as follows:

a. species/brand
b. scientific name if unprocessed
c. source of sample
d. weight per pack and total weight of bulk
e. batch number/ lot number
f. ingredients/composition

## 4.4 Measurand

The FNRI-PTL focused its analytes/measurands of interest to nutrition-related parameters commonly conducted by the private food testing laboratories and local government laboratories. The level of the analytes/measurand varies according to the sample matrix.

## 4.5 Handling and storage of test materials

During preparation and analysis, and prior to distribution, the proficiency test item packets are handled and stored under conditions which minimize contamination, damage, and deterioration, and that the test item integrity is preserved. All proficiency test items are stored in appropriate location (freezer, refrigerator, ambient) until use or dispatch to participant laboratories. After taking the samples for homogeneity testing, the remaining samples for contingency for homogeneity testing,

stability testing, PT participants' testing, and for surplus are stored in a refrigerator and/or ambient temperature.

## 4.6 Packaging and labeling

The packaging is designed to protect the proficiency test item from contamination, damage, and deterioration during storage and distribution. The proficiency test item is packed in a suitable packaging material, i.e. laminated aluminum foil, sealed using a vacuum sealer and/or heat sealer to minimize leakage and absorption of moisture during transport. Each proficiency test item is assigned with a computer-generated random number, and a label indicating the following information:

1. Name of PT Provider;
2. Name and code number of PT Scheme;
3. Measurands;
4. Matrix of PT material;
5. Sample number; and
6. Date of preparation (Month YYYY).

A separate label is also pasted on the label indicating the participants' unique lab code.

## 4.7 PT material homogeneity

Non-homogeneous proficiency test items will undergo another series of homogenization (i.e. mixing, packaging, sealing) prior to distribution to the PT participants. The FNRI proficiency test items will not be distributed until testing demonstrates that the sample is of sufficient homogeneity. FNRI PT uses the statistical procedure for testing "sufficient homogeneity" developed by Fearn and Thompson **[6].** Details of the test material homogeneity testing are retained by the FNRI PTL, and may be released on request upon review and approval of the PTL Head as to the purpose of request.

## 4.8 Quality control of test materials

Surplus proficiency test items are stored in the freezer or refrigerator, depending on the matrix, to preserve for future use. Stability tests are conducted on specified period(s) to check integrity of the material. Storage temperature is also controlled and monitored. The surplus proficiency test items/quality control test materials (QCTMs) are given assigned values and range of satisfactory results from the consensus of the PT participants' results. These will be used as internal quality control (IQC) samples and for method validation on specified analyte/measurand by the local food testing laboratories.

# 5.0 CONDUCT OF PROFICIENCY TESTING

## 5.1 Distribution and Receipt of Proficiency Test Item

The proficiency test item is sent to the participant laboratories for analysis in their respective laboratories together with the following documents:

**a. Receipt Form**

The Receipt Form is provided for the participants to acknowledge receipt of the package containing the proficiency test item and pertinent documents, and indicate the condition of the package and completeness of enclosed documents upon receipt.

**b. Instruction to Participants**

The Instructions for Participants is carefully designed for each PT Round, and participants are instructed to adhere to them closely. It is the participants' responsibility to read and follow the instructions. The FNRI PTL will not be held responsible for any problem arising from failure to comply with the instruction.

**c. Method Details Form**

The Method Details Form is provided for the participants to supply and clearly state the steps in the analytical procedure used for each measurand. The information in the form is being used to compare similarities and differences in the treatment of the sample among participants, and to evaluate possible causes of error and/or outlying test results. The information is used to support in the interpretation why participants may have obtained a "Warning" or an "Action" signal (e.g. if the method used is not appropriate or has limitations for the analyte/measurand under test). On occasions, the PT provider may require the complete method details used by the participant in the analysis.

**d. Results Sheet**

The Results Sheet is provided for the participants to write the results of analysis in a prescribed format (e.g. units of measure, number of decimal places). The format is provided in the Results Sheet to ensure consistency of reporting of the results for statistical treatment.

The proficiency test items are distributed through courier or by pick-up at the FNRI-DOST Office for Metro Manila participants.

It is the responsibility of the participants to contact the PT provider if they have not received the test material within the agreed timescales.

The PT documents are also sent electronically via e-mail to the PT participants' contact person.

## 5.2 Analysis of Proficiency Test Item

The participants are instructed to perform the analyses using their own routine test method, unless otherwise stated by the PT provider. The proficiency test item should be treated in the same way samples are routinely tested in their respective laboratories, i.e. no special treatment of the proficiency test item. In the best way possible, participants are expected to use validated/verified standard methods of analysis.

## 5.3 Reporting of Results

The participants are given a minimum of fifteen (15) days to four (4) weeks after receipt of the proficiency test item to finish the analyses, record the results of analyses on the Results Sheet in the prescribed format, and submit to the FNRI PT Provider on or before the set deadline. Part of the challenge of proficiency testing is the ability to perform calculations and transcribe results correctly. The PT provider cannot interpret or calculate results for the participants.

The participants are also requested to report the estimated measurement uncertainty (MU) for each result as expanded uncertainty, expressed in a prescribed unit, confidence level, and coverage factor.

## 5.4 Late return of results

The PT provider maintains strict control over the return of results, and all returned results are filed accordingly. Participant results submitted after the due date for submission will not be included in the statistical analysis.

## 5.5 Collation and verification of results

The PT results, MU, and method details are collated, tabulated, checked for accuracy of entry, and verified before statistical evaluation is performed. Instructions are given to the participants that the reported results are considered final, unless corrected before the set deadline for submission of results. No corrections from participants will be accepted once the statistical evaluation of results started.

## 5.6 Ethical issues

Participant laboratories shall employ methods used in their routine analyses, and shall not subcontract the analyses to another laboratory. Laboratories shall avoid collusion and falsification of results. Participants should not discuss with each other the results of respective laboratories.

In all occasion, participant laboratories are expected to behave in a professional manner. PT Rounds are intended primarily to help participants test their own methods of analysis, learn from their peers, and ultimately improve their laboratory performance. In rare instances, some participants may only be interested in achieving the “correct”

result. In these cases, submitted results may have been obtained as a result of fabrication or collusion.

Certain measures are built onto the Round to try and prevent collusion, for example, assigned values are not made known to anyone before closure date of the PT Round, and no PT participants' results are accepted once the assigned value has been issued. Moreover, accreditation agencies also have a role to play in detecting unethical behavior.

## 6.0 STATISTICAL EVALUATION AND REPORTING OF RESULTS

The object of applying statistics to results is to produce a simple, quantitative, and transparent means of comparing performance for participants and other interested parties to readily understand and interpret. The statistical procedure used in the evaluation of the PT participants' results follows the FNRI Proficiency Testing (PT) Supplement on Statistical Procedures **[7]** based on ISO 13528:2005, *Statistical Methods for Use in Proficiency Testing by Interlaboratory Comparisons* **[4]**.

### 6.1 Analysis of results

Results are evaluated as soon as possible to give assigned values of the analyte for the PT scheme. It is recommended that reported results and calculations are checked thoroughly before submitting to the PT provider as mistakes cannot be rectified once the assigned values have been reported. Laboratories that did not meet the "Satisfactory" range (i.e. | z-score | > 2.0), are encouraged to review, investigate their results, and take corrective action to prevent recurrence of the problem.

### 6.2 PT report content

The FNRI PT Reports contain details of the composition of test materials, which include the following:

1. standard deviation for proficiency assessment ($\sigma_p$);
2. assigned values (X);
3. standard uncertainty ($u_x$); and
4. computed z-scores.

Tabular and graphical presentations of the results are also provided, namely:

1. table of results, methods used by participants, and z-scores/z'-scores;
2. summary statistics;
3. dotplot of all results;
4. plot of ordered test results with expanded uncertainty;
5. plot of ordered test results according to methods used; and
6. plot of ordered z-scores/z'scores.

## 7.0 PERFORMANCE EVALUATION

The FNRI-PTL uses z-scores/z'scores to demonstrate performance of the PT participants and computes performance based on a fitness-for-purpose criterion.

The standard deviation for proficiency assessment ($\sigma_p$), which defines the scale of acceptable variability and determines the limits of "Satisfactory" (S) performance in participants' results, is determined based on a fit-for-purpose criterion (i.e. by collaborative trial, by perception of how laboratories should perform based on CV, or by Horwitz equation).

Computation of z-score/z'score using $\sigma_p$ based on a fitness-for-purpose criterion addresses both the trueness and the precision of the result. The standard of accuracy required is based on an uncertainty that is determined independently to be appropriate for the conduct of analyses. Moreover, scores can be computed only when the consensus value is suitable to be used as the assigned value (X).

## 8.0 TECHNICAL COMMENT

The FNRI PT Round provides expert commentary on the performance of participants with regards to one or more of the following:

a. different methods used by the PT participants and variation between methods or procedures;
b. possible sources of error (with reference to outliers) and suggestions for improving performance;
c. advice and educational feedback to participants as part of the continuous improvement procedures of participants;
d. situations where unusual factors make evaluation of results and commentary on performance impossible;
e. any other suggestions, recommendations, or general comments; and
f. conclusions.

An Interim Report is provided to the PT Participants prior to the release of the Final PT Report. In the Interim Report, an indication wherein the participant laboratory obtained a high or low result than the assigned value (consensus value) is reflected.

## 9.0 CORRECTIVE ACTION

In an instance a PT participant laboratory obtains "Warning" (W) or "Action" (A) Signal, the participant will be advised to check and investigate their results.

## 10.0 INFORMATION DISTRIBUTED TO PARTICIPANTS

The timescales for distribution of results and reports, and format of reports distributed to participants is similar for every PT Round with some modifications, if required. Participants are advised on the date of distribution of the Final PT Report. PT scheme reports are electronically distributed in portable default file (PDF) format.

# 11.0 LIAISON WITH PARTICIPANTS

## 11.1 PT provider's advice and feedback

Communication from the PT provider to participants is usually done through the PT Reports, Protocols, and Supplements. Additional information may be communicated through electronic mail. Post-PT Meeting for local PT participants is conducted after distribution of PT Report to assist participants to identify the different sources of errors and conduct investigative and corrective action to improve laboratory performance.

The PT provider will always use its best endeavor to ensure that advice and information given is accurate and relevant, but the recipient must rely on his/her own judgment to determine if and how the information is to be used.

Advice is available at all times to participating laboratories on all aspects of the Round, and on more general questions of performance, accreditation, and analytical methods. Where appropriate, this may be followed by a written reply, for instance where a literature search has been undertaken and supporting references are needed.

## 11.2 Participant's feedback, complaints and appeals

The PT provider aims to run all PT Rounds in a manner that gives participants satisfaction in every aspect of the Round. It also aims to monitor and improve the quality of FNRI PTL services. Feedback and/or complaints and appeals from participants are regarded as source of valuable information, possibly highlighting shortcomings of the Scheme or its management, and will be investigated promptly.

For feedback, surveys/questionnaires/evaluation forms are distributed during post-PT Meeting. Laboratories have the option not to take part in such surveys although every contribution is valuable.

Where complaints and appeals are received, they will be fully investigated, according to FNRI PT Laboratory's management quality system, to determine the underlying cause, and to decide a course of action. This course of action, together with results of any investigations carried out, will be communicated to the complainant. The PT participant must inform the Head of the Institute in writing and all appeals may be sent via e-mai, post mail, or fax, and must be addressed to The Director.

## 12.0 POLICY GROUPS

FNRI's overall policy for the PT Round is to always consider the following:

a. best interest of the participants;
b. to be fair and objective in its assessment of results;
c. to pass information onto and listen (and react, where appropriate) to information from participants; and
d. to be transparent in PT operations.

Under FNRI PTL's management quality system, there is a dedicated staff for all Round, which is responsible for overseeing and reviewing the operation and management of each Scheme. Qualified external personnel are invited to become members of the Advisory Committee (Technical Working Group and consultants) and contribute to the progress of PT Round. Each external member of policy groups is bound during the term of their membership and with a confidentiality agreement with the PT provider.

## 13.0 ACKNOWLEDGMENT

Advice, comment, and technical inputs on the organization of the PT Round were sought from different institutions and individuals. The FNRI PTL gratefully acknowledges the following for generously providing the funds and their expertise necessary to realize this program:

a. Philippine Council for Industry, Energy and Emerging Technology Research and Development (PCIEERD) of the Department of Science and Technology (DOST) for funding assistance,
b. Dr. Aida R. Aguinaldo, Consultant – Analytical Chemist,
c. Ms. Teresita R. Portugal – Consultant – PT Provision,
d. Ms. Majella M. Vegafria, Consultant – Statistician,
e. The Project Technical Working Group (TWG),
f. The Food Quality and Safety Section Staff, and
g. The local manufacturers of food samples used in the development of the FNRI Food Reference Materials

## 14.0 REFERENCES

(1) ISO/IEC 17025:2005. General requirements for the competence of testing and calibration laboratories.

(2) ISO/IEC 17043:2010. Conformity assessment – General requirements for proficiency testing.

(3) Analytical Methods Committee (AMC) Technical Brief 13, 2003. Terminology – the key to understanding analytical science. *Part 1: Accuracy, Precision, and uncertainty.*

(4) ISO 13528:2005. Statistical methods for use in proficiency testing by interlaboratory Comparisons.

(5) Thompson, M., Ellison, S.L.R., and Wood, R., 2006. The international harmonized protocol for the proficiency testing of analytical chemistry laboratories, *Pure Applied Chemistry,* 78(1), 145 – 196

(6) Fearn, T., and Thompson, M., 2001. A new test for sufficient homogeneity, *Analyst,* 126, 1414 – 1417

(7) Horwitz, W. (1982). Analytical Chemistry. 54 77A-76A.

(8) ILAC G13:2000 'Guide to the Requirements for the Competence of Providers of Proficiency Testing Schemes'.

(9) AMC Technical Brief, 2002. 'Understanding and acting on scores obtained in proficiency testing schemes. Royal Society of Chemistry.'

# APPENDIX A

# *GLOSSARY OF TERMS*

## GLOSSARY OF TERMS

For the purpose of this FNRI PT Protocol, the following definitions apply.

**Accuracy**
The closeness of agreement between a test result and the accepted reference value

**Assigned Value**
The value attributed to a particular quantity and accepted, having an uncertainty appropriate for a given purpose. It is the best estimate of the true value of the measurand in the matrix.

**Average**
The sum of all values divided by the number of values

**Certified Reference Material (CRM)**
A reference material, one or more of whose property values are certified by a technically valid procedure, accompanied by, or traceable to, a certificate or other documentation issued by a certifying body.

**Error**
The difference between a reported result and the assigned value.

**Interlaboratory Comparisons**
The organization, performance and evaluation of tests on the same, or similar items, or materials, by two or more different laboratories in accordance with predetermined conditions.

**Internal Quality Control (IQC)**
A set of procedures undertaken by the laboratory staff for the continuous monitoring of operations and results in order to judge whether the results are sufficiently reliable. It is a system by which a laboratory is able to monitor the day-to-day consistency of results.

**Outlier**
A member of a set of values that is inconsistent with the other members of that set.

**Precision**
The closeness of agreement between independent test results
.

**Proficiency Testing (PT) Round**
A system that objectively checks laboratory results. It aims to compare a participant laboratory's results with those of other laboratories in the purpose of establishing a consensus value (assigned value), and to assess the accuracy of a laboratory's result.

**PT Provider**
The organization which spearheads all the activities related to the PT Scheme

**Quality Assurance Program**
The sum of a laboratory's activity aimed at achieving the required standard of analysis. While IQC and proficiency testing are very important components of an analytical quality assurance program, it must also include administration, staff training, management structure, auditing, and other important activities that support analytical testing.

**Reference Material (RM)**
A material or substance with one or more properties of which are sufficiently homogeneous and well-established to be used for the calibration of an apparatus, assessment of a measurement method, or for assigning values to other materials.

**Robust Statistical Techniques**
Techniques to minimize the influence that extreme results can have on estimates of the mean and standard deviation

**Standard Deviation**
A measure of the dispersion of data about the mean value

**Target Standard Deviation**
A numerical value that is designated by the organizer of a proficiency testing (PT) Scheme as a realistic goal for measurement quality

**Test Method**
A defined technical procedure to determine one or more specified characteristics of a material or product

**Test Material**
The material or sample provided for the purposes of a proficiency testing (PT) scheme.

**Testing Laboratory**
A laboratory that measures, examines tests, calibrates or otherwise determines the characteristics or performance of materials or products.

**Trueness**
The closeness of agreement between the average value obtained from a large series of test results and an accepted reference value

**True Value**
The actual concentration of the measurand under test in the test material

**z-score**
A standardized measure of performance, calculated using the participant result, assigned value and the standard deviation for proficiency assessment

**z'-score**
A common variation to z-score formed by combining the uncertainty of the assigned value with the standard deviation for proficiency assessment before calculating the z-score

# APPENDIX B

# *FNRI PROFICIENCY TESTING SUPPLEMENT ON STATISTICAL PROCEDURES*

# I. EVALUATION OF PROFICIENCY TESTING (PT) RESULTS

## A. INTRODUCTION

Proficiency test results are assessed by comparison with assigned values derived from the consensus of results (consensus value) from participants, or values determined by a reference laboratory.

The consensus values are estimated using robust procedures. Robust procedures are used in the estimation of consensus values because the most commonly used measures of location and dispersion – **arithmetic mean and standard deviation** – are highly influenced by the presence of extreme outliers and their interpretation depends on an implicit assumption that they are a random sample from a normal distribution. The mean and standard deviation are the optimal estimators of location and dispersion, respectively, for a normal distribution but they can be substantially sub-optimal for distributions close to the normal.

It is very common in many fields to encounter data that have skewed distributions or contain outliers. Analytical data from testing laboratories often depart from the assumption that the data are a random sample from a normal distribution. It is often heavy tailed – contains a higher than expected proportion of results far from the mean – and sometimes contains outliers.

***Outliers*** are values that are so far in value from the rest of the data that they may be viewed as coming from a different population, or the result of a measurement error. One way of coping with outliers is to exclude them from the calculation of the statistics. But when is it justifiable to exclude outliers in the calculation? The decision to exclude or retain an outlier depends on the understanding of the cause of the outlier and its impact on the results.

Employing tests such as Grubbs' test or the boxplot usually identifies suspect outliers. The use of the Grubbs' test presumes that the distribution of the variable is approximately normal. A boxplot, on the other hand, can be used in identifying outliers for both normal and non-normal distributions.

On the basis of some simple assumptions, outlier tests identify where it is likely to have a technical error but it does not assess or judge that the point is "wrong". In a data set, the value may be extreme but it could be the correct one. Only with experience or by identification of a certain cause can data be declared "wrong" and excluded from the computations. Generally, if more than 20% of the data are identified as outlying, the assumption about the data distribution and/or the quality of the data collected becomes questionable.

A convenient way of coping with outliers is to use **robust statistics**. Robust statistics includes methods that are largely unaffected by the presence of extreme values. "It provides an alternative way of summarizing results when

they include a small proportion of outliers, without the requirement to identify specific observations as outliers or exclude them." [**1**].

Examples of robust statistics are the median and the mode for they are not highly influenced by the presence of outliers. "The ***median*** is the value in an ordered data set that has an equal number of data points on either side while the ***mode*** is the value of the peak of the distribution." [**2**].

Among the three statistics – mean, median and mode – the mode is least affected by the presence of outliers. However, because the calculation of the mode is more difficult than that of the mean or median, the mode has limited application.

## B. SETTING THE STANDARD DEVIATION FOR PROFICIENCY ASSESSMENT

"The standard deviation for proficiency assessment ($\sigma_p$) is a parameter that is used to provide a scaling for the laboratory deviations from the assigned value and thereby define a z-score. The value is determined by "fitness-for-purpose" as it does not represent a general idea of how laboratories are performing, but how they ought to perform to fulfill their commitment to their clients." [**3**].

**Fitness-for-purpose** is the ability of a value to satisfy a set of conditions given by the application. "The **uncertainty of measurement** is a parameter associated with the results of a measurement that characterizes the dispersion of the value that could reasonably be attributed to the measurand." [**4**].

Most common approaches in setting the $\sigma_p$, of a measurand are the following:

(i) by collaborative trial data calculated using the formula:

$$\sigma_p = RSD_R \times x_{pt}$$

where:

$RSD_R$ is the relative standard deviation of reproducibility from collaborative trials

$x_{pt}$ is the assigned value from consensus of PT participants' results derived as a robust average using Alogarithm A of ISO 13528 expressed in appropriate units

(ii) by perception of how laboratories should perform, based on CV of previous PT results on appropriate (same or similar) matrix

$$\sigma_p = \frac{(CV \times x_{pt})}{100}$$

where:

CV is the coefficient of variation

$x_{pt}$ is the assigned value from consensus of PT participants' results, derived as a robust average using Algorithm A of ISO 13528 expressed in appropriate units

(iii) by Horwitz equation in the absence of collaborative trial data for minerals using any of the formula:

$$\sigma_p = 0.02(X_c^{0.8495})$$

where:

$X_c$ is the consensus value, $\sigma_p$ and $X_c$ are expressed as mass fraction

The standard deviation of reproducibility found in collaborative trials is generally considered an appropriate indicator of the best agreement that can be obtained between laboratories [**5**].

## C. EVALUATION PROCESS FOR PT RESULTS

The evaluation of proficiency test results proceeds as follows:

- **Exclusion of invalid data**

  There may be instances where a participant's test result will be excluded from the calculation of a measurand's consensus value and its associated standard uncertainty. Reasons for exclusion are the following:

  a. method used for the measurand is not applicable to the food matrix (e.g. fat analysis using direct solvent extraction instead of acid hydrolysis);
  b. removal of extreme results or results that are identifiably invalid, (e.g., results caused by calculation errors or used wrong unit of measurements). Results which are out of range of the median ± 5 * $\sigma_{pt}$ will be excluded based on the General Protocol of LGC standards Proficiency Testing **[4]**: and
  c. extremely low or high values identified by the boxplot rule.

- **Determination of the assigned value ($x_{pt}$) and its standard uncertainty ($u_x$)**

  - Calculation of the robust average (x*) or median (med (x)) for use as consensus value and the corresponding robust standard deviation (s*) or *MADe(x)*, whichever is applicable, of the test results:
  - Computation of the standard uncertainty of the consensus value (u) using the formula:

$$U_x = \frac{1.25 \times MADe(x) \text{ or } s^*}{\sqrt{n_2}}$$

where:

*MADe(x)* is the scaled median absolute deviation calculated using the formula:

$MADe(x) = 1.483 * med(d)$

| | |
|---|---|
| s* | is the robust standard deviation computed using Algorithm A of ISO 13528: 2005 |
| $N_2$ | is the number of data included in the computation of consensus value |

- **Calculation of performance statistics**

  z-scores are typically used in the evaluation of performance. The z-scores are calculated, using the consensus value and $\sigma_{pt}$, only when the consensus value is suitable for use as an assigned value.

  Section I.C.2 gives the details on the calculation of performance statistics.

- **Evaluation of performance**

  Section I.C.3 describes the steps in evaluating the performance of participating laboratories.

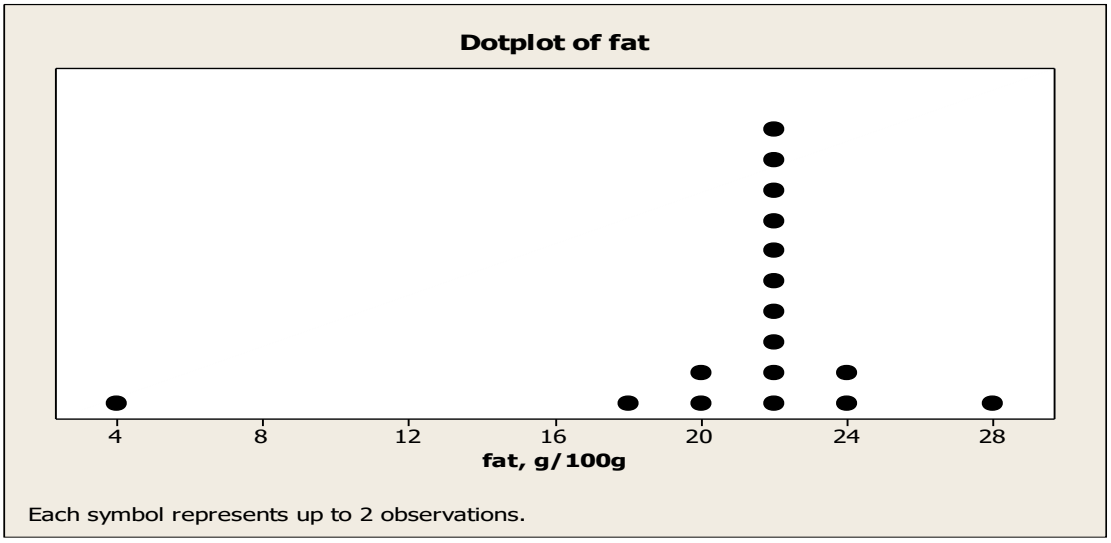
Graphical representations of some evaluation outputs are in Figures 1 to 5. The flowchart of the statistical evaluation process is presented in Figure 6 and the detailed steps and explanatory notes are given below.

## 1. Determination of the Assigned Value (X) and its Standard Uncertainty ($u_X$)

1.1. *Construction of a dotplot or histogram of the results*

When the number of data points in the plot is less than 50, a dotplot is constructed. When there are 50 or more data points, the histogram is used instead.

**Figure 1**



1.2. *Calculation of initial statistics (e.g. mean, median, minimum value, maximum value) of all data*

1.3. *Exclusion of data obtained using inapplicable method or expressed in wrong units of measure (e.g. fat analysis using direct solvent extraction instead of acid hydrolysis prior to solvent extraction)*

1.4. *Identification and exclusion of laboratories with extremely low or high reported results.*

Test results that satisfy the following boxplot criteria are considered as extreme outliers:

$x > (Q3 + (3 \times IQR))$ or $x < (Q1 - (3 \times IQR))$

where

x is the laboratory's average result
Q1 is the first quartile of the laboratory averages
Q3 is the third quartile of the laboratory averages and
IQR is the interquartile range and computed as follows:

$IQR = (Q3 - Q1)$

1.5. *Removal of extreme results or results that are identifiably invalid, (e.g., results caused by calculation errors or used wrong unit of measurements). Results which are out of range of the median ± 5 * $\sigma_{pt}$ will be excluded based on the General Protocol of LGC standards Proficiency Testing:*

1.6. *Calculation of the robust average (x*) for use as consensus value and the corresponding robust standard deviation (s*) of the test results with outlying data excluded using Algorithm A of ISO 13528:2005*

1.7. *Calculation of the standard uncertainty of the consensus value (*u*) using the following formula:*

$$u = \frac{1.25 \times s^*}{\sqrt{n}}$$

where:

n is the number of data included in the computation of the robust average, and
s* is the robust standard deviation computed using Algorithm A

1.8. *Determination of the suitability of the consensus value to be used as assigned value based on the ISO 13528:2015 criteria [3]:*

if $u \leq 0.3\sigma_{pt}$ - u is negligible, z-scores can be issued;
if $u > 0.3\sigma_{pt}$ - u is high, use the uncertainty of the assigned value

in the interpretation of performance, i.e. z'-scores can be issued,
if $\mu_x^2 + \sigma_{pt}^2 \leq \sigma_{rob}^2$

where:

$\sigma_{pt}$ is the standard deviation for proficiency assessment
$\mu_x$ is the standard uncertainty of the assigned value
$\sigma_{rob}$ is the robust standard deviation

1.9 *Abandonment of attempt to determine a consensus value*

The attempt to determine a consensus value is abandoned if the uncertainty of the consensus value is not negligible or is too high, i.e. $\mu_x^2 + \sigma_{pt}^2 \leq \sigma_{rob}^2$. There is no real consensus of results, thus no z-score is issued. However, the participants are provided with summary statistics (e.g. mean, median) of the data set as a whole.

## 2. Calculation of Performance Statistics

z-scores are the basis for evaluating the performance of participating laboratories. The z-scores are calculated using the consensus value and the standard deviation for proficiency assessment ($\sigma_{pt}$), only when the consensus value is suitable for use as the assigned value. The performance of individual PT participant laboratories was evaluated using the formula:

$$z = \frac{x - x_{pt}}{\sigma_{pt}}$$

where
- $x$ is the participant's reported result
- $x_{pt}$ is the assigned value from the consensus of the PT participants' results derived as a robust average or median
- $\sigma_{pt}$ is the standard deviation for proficiency assessment

The laboratory z-scores are interpreted as follows:

|z-score| ≤ 2.0: "Satisfactory" **(S)** performance
2.0 < |z-score| < 3.0: "Warning" **(W)** signal
|z-score| ≥ 3.0: "Action" **(A)** signal

If the uncertainty of the assigned value is greater than $0.3\sigma_{pt}$, then the uncertainty can be taken into account by expanding the denominator for the calculation of performance score, such that:

$$z' = \frac{x - x_{pt}}{\sqrt{(\sigma_{pt}^2 + \mu_x^2)}}$$

where:
- $x$ is the participant's reported result
- $x_{pt}$ is the assigned value from the consensus of PT participants'
- $\sigma_{pt}$ is the standard deviation for proficiency assessment
- $\mu_x$ is the standard uncertainty of the assigned value

z'-scores are interpreted in the same way as z-scores and using the same critical values of 2.0 and 3.0.

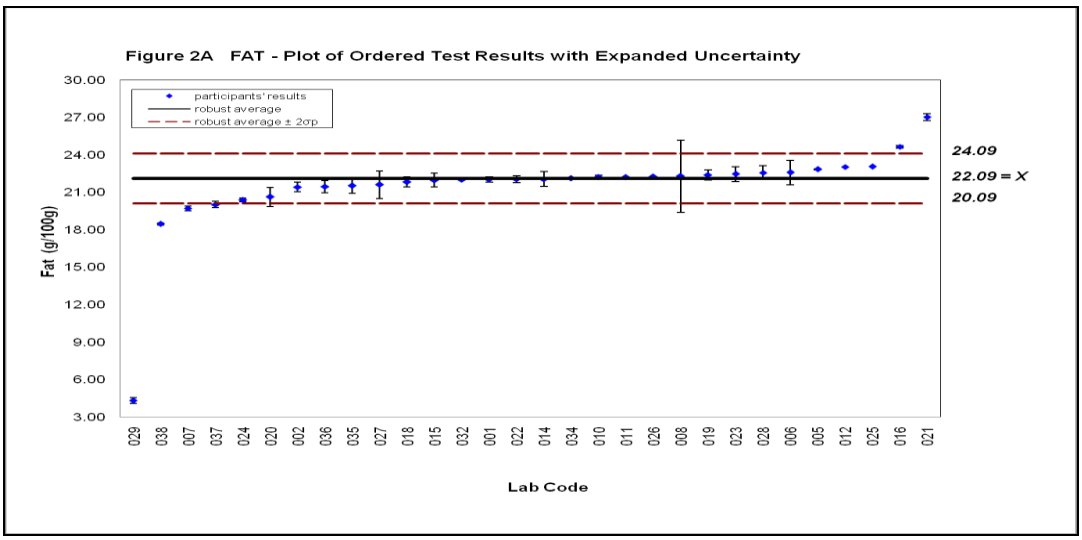
The plots of ordered test results with expanded uncertainty, ordered test results according to methods used and ordered z-scores are also used in evaluating performance. These, including the dotplot, are graphical means by which a participating laboratory can be readily compare its performance relative to the other laboratories.

## 3. Construction of Plots

### 3.1 *Construction of plot of ordered test results with expanded uncertainty*

The plot of ordered test results with expanded uncertainty (Figure 2) is a graphical display of each laboratory's test result with the reported expanded uncertainty. It shows the performance of each laboratory relative to the other laboratories. For example, in Figure 2, the test results starting from **Lab 024** to **Lab 025** are within the range of values of "Satisfactory" range: 20.37 to 23.06 g/100g. However, the test results of **Labs 029, 038, 007** and **037** are below the lower limit of the value for "Satisfactory" range, while **Labs 016** and **021** obtained test results above the upper limit of the value for "Satisfactory" range.

**Figure 2**

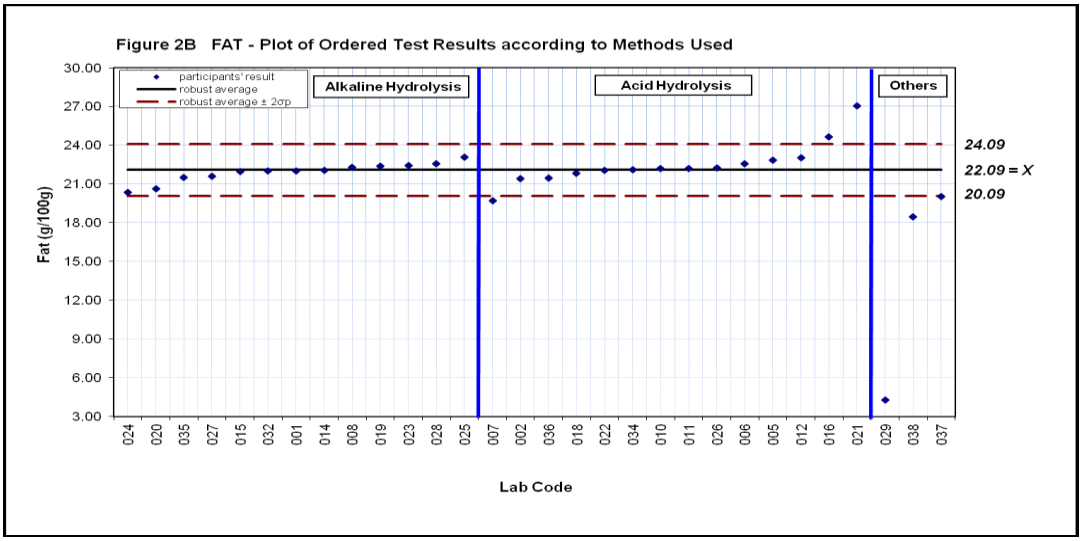



### 3.2 *Construction of plot of ordered test results according to methods used*

This plot is a graphical display of the participant's performance according to methods used. It shows if there are differences and clustering of results by method used. For example in Figure 3, comparable behavior of results was observed for both alkali and acid hydrolysis for fat (i.e. no clustering

of data). When there is clustering of results, there is a need to conduct separate evaluation of participants' results by method used.

**Figure 3**




## 3.3 *Construction of plot of ordered z-scores*

The plot of ordered z-scores is a graphical display of the participants' performance. This plot shows each participant laboratory's performance relative to that of the other laboratories. From this plot, results outside the "Satisfactory" range (i.e. (| z-score | > 2.0) can be quickly identified. As illustrated in Figure 4, **Labs 029, 038, 007, 037, 016,** and **021** have results outside the "Satisfactory" range.

**Figure 4**

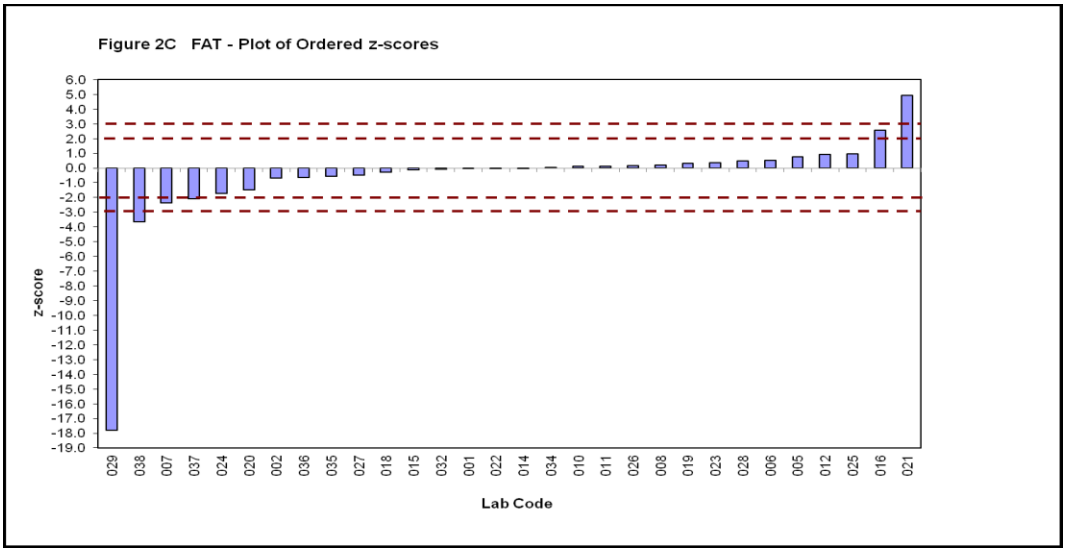
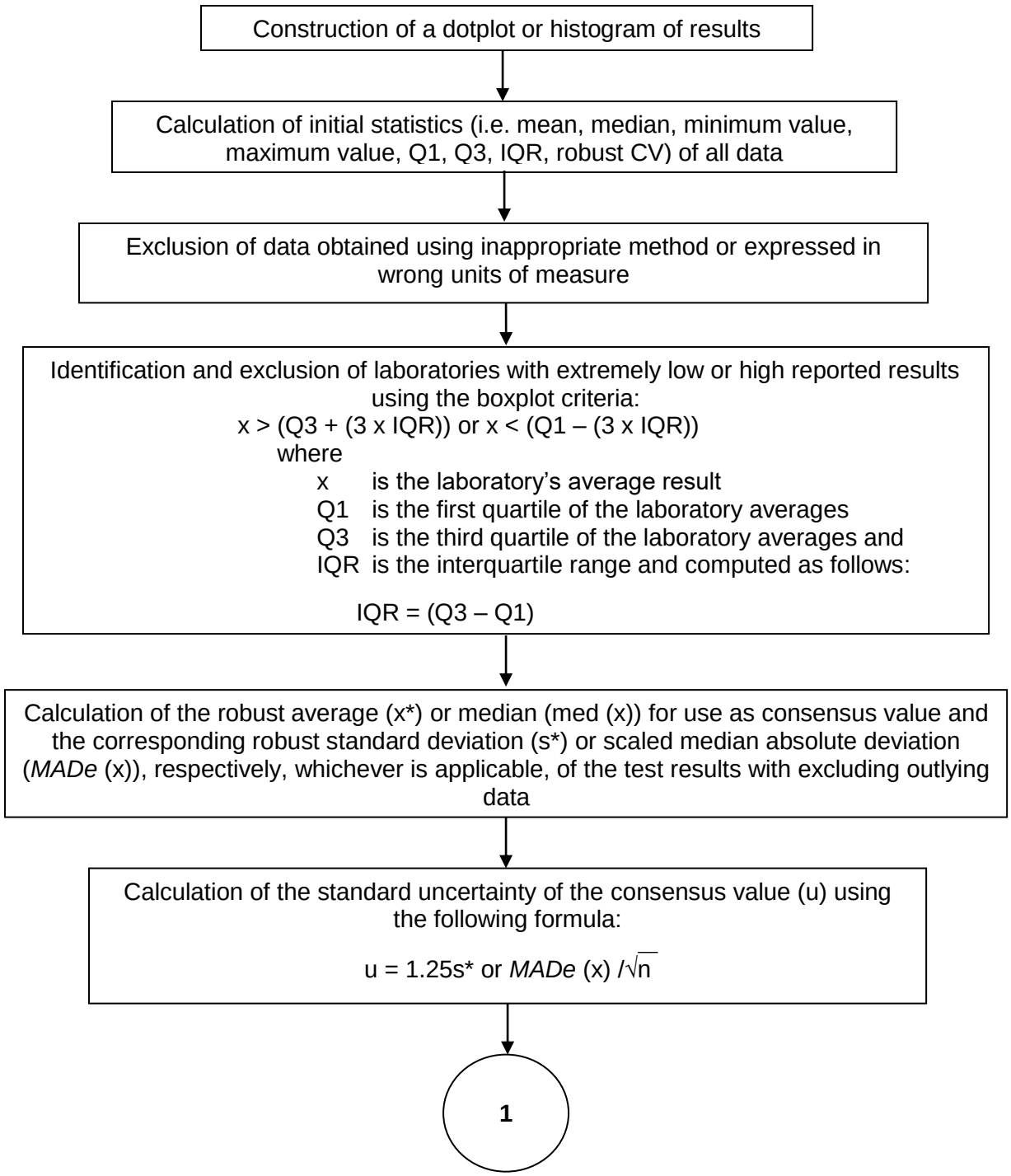
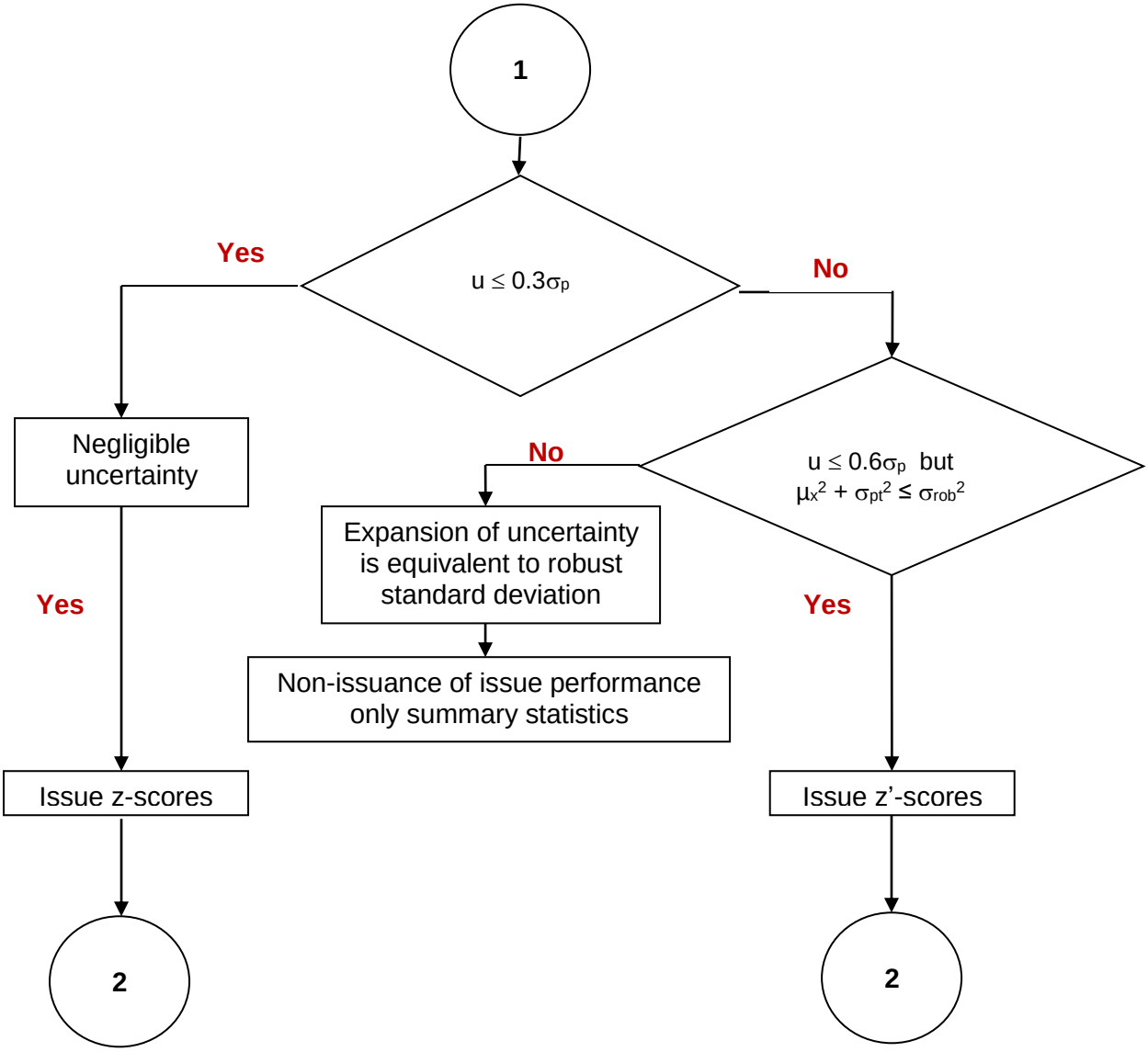
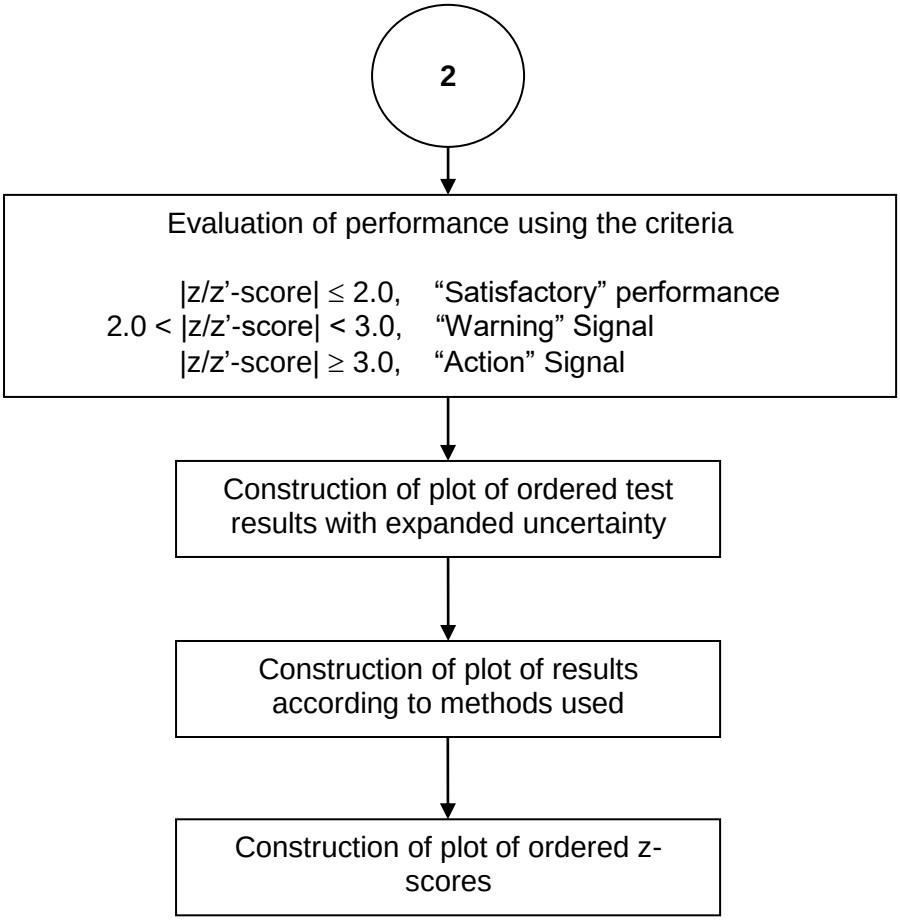


**Figure 5**
**STATISTICAL EVALUATION PROCESS**




1
$u \leq 0.3\sigma_p$
Yes
No
Negligible uncertainty
$u \leq 0.6\sigma_p$ but $\mu_x^2 + \sigma_{pt}^2 \leq \sigma_{rob}^2$
No
Expansion of uncertainty is equivalent to robust standard deviation
Non-issuance of issue performance only summary statistics
Yes
Yes
Issue z-scores
Issue z'-scores
2
2


2
Evaluation of performance using the criteria
|z/z'-score| ≤ 2.0, "Satisfactory" performance
2.0 < |z/z'-score| < 3.0, "Warning" Signal
|z/z'-score| ≥ 3.0, "Action" Signal
Construction of plot of ordered test results with expanded uncertainty
Construction of plot of results according to methods used
Construction of plot of ordered z-scores

## II. EVALUATION OF TEST MATERIAL HOMOGENEITY

The statistical methods used in testing a material for homogeneity are:

- Cochran's test procedure for duplicate results
- Test for "adequate" homogeneity using ISO 13528 assessment criterion for homogeneity check
- Test for "sufficient" homogeneity by Fearn and Thompson (2001)

The following are the steps in conducting the homogeneity test:

1. **Selection** of 10 test samples in their final packaged form using systematic sampling using Microsoft Excel software,

2. **Separate homogenization** of the contents of each of the 10 selected packages by the appropriate techniques, to minimize within-package variability,

3. **Preparation** of two (2) sub-samples from each test sample using techniques appropriate to the test material, to minimize between-test-portion differences,

4. **Obtaining of a measurement result** on each of the twenty (20) sub-samples in a random order as in Step 1 of this Section, where applicable, and completing the whole series of measurements under repeatability conditions (i.e. same analyst, equipment, glassware, etc.).

5. **Examination of data** for pathologies,

    5.1. *Construction of a simple plot of duplicate results*

    Use any software that has the capability to construct a scatterplot of duplicate results, e.g., Excel.

    5.2. *Visual examination of a simple plot of the duplicate results and searching for diagnostic features such as:*

    - trends or discontinuities
    - nonrandom distribution of differences between first and second test results
    - excessive rounding; and
    - outlying results within samples

    5.3. *Testing for outlying results within samples using Cochran's test procedure for duplicate results*

    The Cochran's test procedure is as follows:

    5.3.1. Calculation of the sum, $S_i$, and difference, $D_i$, of each pair of duplicates, for $i = 1, ..., m$, where $m = 10$

5.3.2. Calculation of the sum of squares, $S_{DD}$, of the 10 differences

$$S_{DD} = \sum D_i^2$$

5.3.3. Calculation of the ratio, C, and comparison of the result with the appropriate critical value

The Cochran's test statistic is the ratio of $D^2_{max}$, the largest squared difference, to this sum of squared differences

$$C = D^2_{max}/S_{DD}$$

For 10 test samples analyzed in duplicate, the critical values at 95% and 99% levels of confidence are 0.602 and 0.718, respectively. Refer to the IUPAC Technical Report for other values.

5.3.3.1. Close inspection of outlying pairs detected at the 95% or higher level of confidence for transcription or other errors. An outlying pair is rejected when there are irremediable analytical errors or if the difference between duplicate results is significant at the 99% level.

5.3.3.2. Deletion of duplicate results from a single test sample if they are significantly different from each other at the 99% level of significance.

5.3.3.3. Discarding of data if they contain discrepancies in two or more test samples. However, pairs of results with outlying mean (average) value but with no evidence of extreme variance (difference) are not discarded.

## 6. **Testing for homogeneity**

### 6.1. *Testing for "adequate" homogeneity*

6.1.1. Use of the same sum of squared differences in Step 5.3.2 to calculate the analytical variance, $s^2_{an}$

$$s^2_{an} = \sum D_i^2/2m$$

where:
$D_i^2$ is the difference of each pair of duplicates
m is the total number of samples, i.e. 10

6.1.2. Calculation of the variance $V_S$ of the sums, $S_i$

$$V_S = \frac{\sum (S_i - \overline{S})^2}{(m-1)}$$

where:

$S_i$ is the sum of each pair of duplicates

$\overline{S}$ is the mean of the $S_i$, $(1/m)\sum S_i$

6.1.3. Calculation of the sampling variance, $s^2_{sam}$

$$s^2_{sam} = \frac{(V_S/2 - s^2_{an})}{2}$$

or

$s_{sam}^2 = 0$, if the above estimate is negative.

*Note*: The quantities $V_S/2$ and $s^2_{an}$ may be extracted from the analysis of variance table as the "between" and "within" mean squares, respectively.

6.1.4. Calculation of the sampling standard (between-samples) deviation, $s_{sam}$, and comparing it with $0.3\sigma_p$

6.1.4.1. If $s_{sam} > 0.3\sigma_p$, the test for adequate homogeneity has failed.

6.1.4.2. If $s_{sam} \leq 0.3\sigma_p$, the test for adequate homogeneity has been passed.

6.2. *Testing for "sufficient" homogeneity*

6.2.1. Calculation of the allowable sampling variance, $\sigma^2_{all}$, as

$$\sigma^2_{all} = (0.3\sigma_p)^2$$

where $\sigma_p$ is the SD for PT assessment

6.2.2. Calculation of the critical value for the test as

$$c = F_1\sigma^2_{all} + F_2 s_{an}^2$$

where

$F_1 = 1.88$ and $F_2 = 1.01$ (for 10 test samples measured in duplicate; 95% level of confidence). Refer to the IUPAC 2006 Technical Report for other values.

6.2.3. Use of the calculated sampling variance, $s^2_{sam}$, in Step 6.1.3 and making decision based on the following criteria:

6.2.3.1. If $s_{sam}^2 > c$

there is evidence at the 95% level of confidence that the sampling standard deviation in the population of samples

exceeds the allowable fraction of $\sigma_p$; therefore, the test for homogeneity has failed.

6.2.3.2. If $s_{sam}^2 \leq c$

there is no evidence at the 95% level of confidence that the sampling standard deviation in the population of samples exceeds the allowable fraction of $\sigma_p$; therefore, the test for homogeneity has been passed.

## BIBLIOGRAPHY


1. Analytical Methods Committee. "Robust statistics: a method of coping with outliers", AMC Technical Brief No. 6. 2001. Available online at <http://www.rsc.org/images/ brief6_tcm18-25948.pdf> [April 30, 2007].

2. Hedges, S. and Shah, P. 2003. *Comparison of mode estimation methods and application in molecular clock analysis*Available online at <http://www.biomedcentral.com/ content/pdf/1471-2105-4-31.pdf> [April 30, 2007]

3. Thompson, M., Ellison, S.L.R. and Wood, R. 2006. The International Harmonized Protocol for the proficiency testing of analytical chemistry laboratories. *Pure Applied Chem.* **78** (1), 145-196.

4. International Organization of Standardization. *ISO/IEC Guide 43-1: Proficiency testing by interlaboratory comparisons – Part 1: Development and operation of proficiency testing schemes*, 2nd ed., ISO, Geneva, Switzerland.1997.

5. International Organization of Standardization. *ISO 13528:2005. Statistical methods for use in proficiency testing by Interlaboratory comparisons,* ISO, Geneva, Switzerland.

6. Food Analysis Performance Assessment Scheme (FAPAS). 2002. *Proficiency Testing Protocol: Organization and Analysis of Data*, 6th ed., FAPAS, York, UK.

7. LGC Standards Proficiency Testing. (September 2014). General Protocol: Proficiency Testing Schemes. p 14.

8. _____, *Random Error and Systematic Error*. Available online at <http://www.phys.selu.edu/rhett/plab193/labinfo/Error_Analysis/05_Random_vs_systematic.html> [April 30, 2007]

9. Burke, S. *Missing values, outliers, robust statistics & non-parametric methods.* Available online at <http://www.lgceurope.com/lgceurope/data/articlestandard/lgceurope/502001/4509/article.pdf> [April 30, 2007]

10. _____, *Box and Whisker Plots.* 2003. Available online at <http://www.mathcentre.ac.uk/resources/leaflets/mathcentre.business/box_and_whisker_plots.pdf > [April 30, 2007]

11. Thompson, M. and Fearn, T. 2001. A new test for 'sufficient homogeneity. *Analyst*, **126**, 1414-1417.

12. Hall, J.M., Bean, V.E. and Mattingly, G.E. 1999. *Preliminary Results from Interlaboratory Comparisons of Air Speed Measurements Between 0.3 m/s and 15 m/s.* 1999 NCSL Workshop & Symposium. Available online at <http://www.cstl.nist.gov/ div836/836.01/PDFs/1999/NCSL_056.pdf> [April 30, 2007].