

INTERNATIONAL  
STANDARD

ISO  
13528

First edition  
2005-09-01

เพื่อใช้เป็นข้อมูลทางวิชาการ

---

---

**Statistical methods for use in proficiency  
testing by interlaboratory comparisons**

*Méthodes statistiques utilisées dans les essais d'aptitude par  
comparaisons interlaboratoires*

ใช้ประกอบการอบรมวิชาการ



Reference number  
ISO 13528:2005(E)

© ISO 2005



PDF disclaimer

This PDF file may contain embedded typefaces. In accordance with Adobe's licensing policy, this file may be printed or viewed but shall not be edited unless the typefaces which are embedded are licensed to and installed on the computer performing the editing. In downloading this file, parties accept therein the responsibility of not infringing Adobe's licensing policy. The ISO Central Secretariat accepts no liability in this area.

Adobe is a trademark of Adobe Systems Incorporated.

Details of the software products used to create this PDF file can be found in the General Info relative to the file; the PDF-creation parameters were optimized for printing. Every care has been taken to ensure that the file is suitable for use by ISO member bodies. In the unlikely event that a problem relating to it is found, please inform the Central Secretariat at the address given below.

ใช้ประกอบการอบรมเท่านั้น

© ISO 2005

All rights reserved. Unless otherwise specified, no part of this publication may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying and microfilm, without permission in writing from either ISO at the address below or ISO's member body in the country of the requester.

ISO copyright office  
Case postale 56 • CH-1211 Geneva 20  
Tel. + 41 22 749 01 11  
Fax + 41 22 749 09 47  
E-mail [copyright@iso.org](mailto:copyright@iso.org)  
Web [www.iso.org](http://www.iso.org)

Published in Switzerland

## Contents

Page

Foreword.....	v
0 Introduction.....	vi
0.1 The aims of proficiency testing.....	vi
0.2 ISO/IEC Guide 43.....	vi
1 Scope.....	1
2 Normative references.....	1
3 Terms and definitions.....	1
4 Statistical guidelines for the design and interpretation of proficiency tests.....	2
4.1 Action and warning signals.....	2
4.2 Guidelines for limiting the uncertainty of the assigned value.....	3
4.3 Guidelines for choosing the number of replicate measurements.....	3
4.4 Homogeneity and stability of samples.....	4
4.5 Operationally defined measurement methods.....	4
4.6 Reporting of data.....	5
4.7 Period of validity of the results of proficiency tests.....	5
5 Determination of the assigned value and its standard uncertainty.....	5
5.1 Choice of method of determining the assigned value.....	5
5.2 Formulation.....	5
5.3 Certified reference values.....	6
5.4 Reference values.....	7
5.5 Consensus values from expert laboratories.....	8
5.6 Consensus value from participants.....	9
5.7 Comparison of the assigned value.....	14
5.8 Missing values.....	14
6 Determining the standard deviation for proficiency assessment.....	15
6.1 Choice of method.....	15
6.2 Prescribed value.....	15
6.3 By perception.....	15
6.4 From a general model.....	17
6.5 From the results of a precision experiment.....	17
6.6 From data obtained in a round of a proficiency testing scheme.....	18
6.7 Comparison of precision values derived from a proficiency test with established values.....	18
7 Calculation of performance statistics.....	18
7.1 Estimates of laboratory bias.....	18
7.2 Percentage differences.....	22
7.3 Ranks and percentage ranks.....	24
7.4 $z$ -scores.....	25
7.5 $E_n$ numbers.....	27
7.6 $z'$ -scores.....	28
7.7 Zeta-scores ( $\zeta$ ).....	29
7.8 $E_z$ score.....	30
7.9 An example of the analysis of data when uncertainties are reported.....	30
7.10 Combined performance scores.....	35
8 Graphical methods for combining performance scores for several measurands from one round of a proficiency test.....	36
8.1 Application.....	36
8.2 Histograms of performance scores.....	36

8.3	Bar-plots of standardized laboratory biases.....	37
8.4	Bar-plots of standardized repeatability measurements.....	38
8.5	Youden Plot.....	38
8.6	Plots of repeatability standard deviations.....	45
8.7	Split samples.....	47
9	Graphical methods for combining performance scores over several rounds of a proficiency testing scheme.....	52
9.1	Applications.....	52
9.2	Shewhart control chart for $z$ -scores.....	52
9.3	Cusum control chart for $z$ -scores.....	55
9.4	Plots of standardized laboratory biases against laboratory averages.....	56
9.5	Dot plot.....	57
Annex A (normative)	Symbols.....	59
Annex B (normative)	Homogeneity and stability checks of samples.....	60
Annex C (normative)	Robust analysis.....	64
Bibliography	.....	66

กรมการมาตรฐาน  
กระทรวงอุตสาหกรรม

## Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

The main task of technical committees is to prepare International Standards. Draft International Standards adopted by the technical committees are circulated to the member bodies for voting. Publication as an International Standard requires approval by at least 75 % of the member bodies casting a vote.

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights.

ISO 13528 was prepared by Technical Committee ISO/TC 69, *Applications of statistical methods*, Subcommittee SC 6, *Measurement methods and results*.

## 0 Introduction

### 0.1 The aims of proficiency testing

Proficiency testing by interlaboratory comparisons is used to determine the performance of individual laboratories for specific tests or measurements, and to monitor the continuing performance of laboratories. The Introduction to ISO/IEC Guide 43-1:1997 should be consulted for a full exposition of the purposes of proficiency testing. In statistical language, the performance of laboratories can be described by three properties: laboratory bias, stability and repeatability. Laboratory bias and repeatability are defined in ISO 3534-1, ISO 3534-2 and ISO 5725-1. The stability of a laboratory's results is measured by intermediate precision as defined in ISO 5725-3.

Laboratory bias may be assessed by tests on reference materials, when these are available, using the procedure described in ISO 5725-4. Otherwise, proficiency testing by interlaboratory comparisons provides a generally available means of obtaining information about laboratory bias, and the use of data from proficiency tests to obtain estimates of laboratory bias is an important aspect of the analysis of such data. However, stability and repeatability will affect data obtained in proficiency tests, so that it is possible for a laboratory to obtain data in a round of a proficiency test which indicate bias that is actually caused by poor stability or poor repeatability. It is therefore important that these aspects of laboratory performance are assessed regularly.

Stability may be assessed by re-testing of retained samples, or by making regular measurements on a reference material or an in-house reference material (a stock of material established by a laboratory to use as private reference material). These techniques are described in ISO 5725-3. Stability may also be assessed by plotting estimates of laboratory bias derived from proficiency tests in control charts. This can provide information about laboratory performance that is not apparent from the examination of the results of individual rounds of proficiency testing schemes, and is another important aspect of the analysis of such data.

Data suitable for assessing repeatability may be generated by tests carried out in the normal course of the work of a laboratory, or by extra tests carried out within a laboratory specifically to assess repeatability. Consequently, the assessment of repeatability is not necessarily an important aspect of proficiency testing, although it is important that laboratories monitor their repeatability in some way. Repeatability may be assessed by plotting ranges of duplicate measurements on a control chart as described in ISO 5725-6.

The flowchart (Figure 1) illustrates how the techniques described in this International Standard are to be applied.

### 0.2 ISO/IEC Guide 43

ISO/IEC Guide 43-1 describes different types of proficiency testing schemes and gives guidance on the organization and design of proficiency testing schemes. ISO/IEC Guide 43-2 gives guidance on the selection and use of proficiency testing schemes by laboratory accreditation bodies. Those documents should be consulted for detailed information in those areas (the information is not duplicated here). ISO/IEC Guide 43-1 contains an annex that briefly describes the statistical methods that are used in proficiency testing schemes.

This International Standard is complementary to ISO/IEC Guide 43, providing detailed guidance that is lacking in that document on the use of statistical methods in proficiency testing. ISO 13528 is to a large extent based on a harmonized protocol for the proficiency testing of analytical laboratories [1], but is intended for use with all measurement methods.

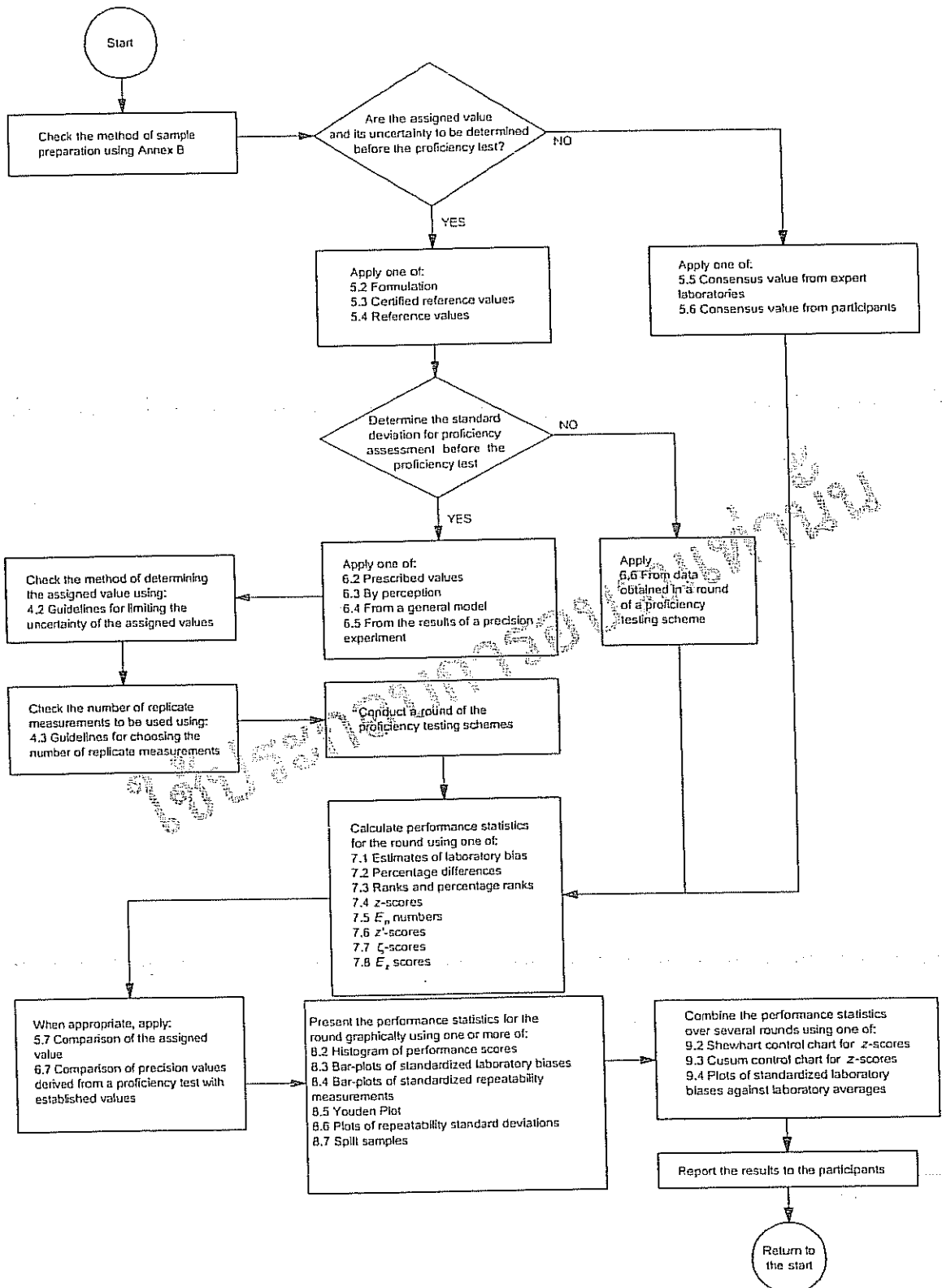


Figure 1 — Flowchart showing the activities requiring the use of statistical methods when operating a proficiency testing scheme



# Statistical methods for use in proficiency testing by interlaboratory comparisons

## 1 Scope

This International Standard complements ISO Guide 43 (all parts) by providing detailed descriptions of sound statistical methods for organizers to use to analyse the data obtained from proficiency testing schemes, and by giving recommendations on their use in practice by participants in such schemes and by accreditation bodies.

This International Standard can be applied to demonstrate that the measurement results obtained by laboratories do not exhibit evidence of an unacceptable level of bias.

It is applicable to quantitative data but not to qualitative data.

## 2 Normative references

The following referenced documents are indispensable for the application of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO 3534-1, *Statistics — Vocabulary and symbols — Part 1: Probability and general statistical terms*

ISO 3534-2:—1), *Statistics — Vocabulary and symbols — Part 2: Applied statistics*

ISO 5725-1, *Accuracy (trueness and precision) of measurement methods and results — Part 1: General principles and definitions*

ISO/IEC Guide 43-1:1997, *Proficiency testing by interlaboratory comparisons — Part 1: Development and operation of proficiency testing schemes*

## 3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO 3534-1, ISO 3534-2, ISO 5725-1 and the following apply.

### 3.1

#### **interlaboratory comparison**

organization, performance and evaluation of tests or measurements on the same or similar test items by two or more laboratories in accordance with predetermined conditions

NOTE Adapted from ISO/IEC Guide 43-1.

### 3.2

#### **proficiency testing**

determination of laboratory testing performance by means of interlaboratory comparisons

1) To be published.

**3.3**

**assigned value**

value attributed to a particular quantity and accepted, sometimes by convention, as having an uncertainty appropriate for a given purpose

**3.4**

**standard deviation for proficiency assessment**

measure of dispersion used in the assessment of proficiency, based on the available information

**3.5**

**z-score**

standardized measure of laboratory bias, calculated using the assigned value and the standard deviation for proficiency assessment

**3.6**

**coordinator**

organization (or person) with responsibility for coordinating all of the activities involved in the operation of a proficiency testing scheme

**4 Statistical guidelines for the design and interpretation of proficiency tests**

(see ISO/IEC Guide 43-1:1997, 5.4.2.)

**4.1 Action and warning signals**

**4.1.1** This International Standard describes some simple numerical or graphical criteria that should be applied to the data obtained in a proficiency test to see if they give rise to action or warning signals. Even in a well-run laboratory, with experienced staff, anomalous results may sometimes be obtained. Also, it is possible that a standardized measurement method, even though it has been validated by a precision experiment, may contain faults that become apparent only after several rounds of a proficiency testing scheme. The proficiency scheme itself may contain faults. For these reasons, the criteria given here shall not be used to condemn laboratories, as being unfit to perform the measurement method under examination. If proficiency testing is used to condemn laboratories, then it shall be necessary to devise appropriate criteria for that purpose.

**4.1.2** The criteria given here are designed so that, when the standard deviation for proficiency assessment is based on observed performance (using one of the methods described in 6.4 to 6.6), the criteria give action signals when results are so exceptional as to merit investigation and corrective action.

**4.1.3** The coordinator should have an understanding of the major sources of variability that can be anticipated in proficiency test data for the measurement in question. The first step in any analysis should be to examine the distribution of results for evidence of unanticipated sources of variability. For example, a bimodal distribution might be evidence of a mixed population of results caused by different methods, contaminated samples or poorly worded instructions. In this situation, the concern should be resolved before proceeding with analysis or evaluation. Accrediting bodies shall have policies for response to unacceptable performance in proficiency testing. Follow-up actions are determined by that policy or by the laboratory's quality procedures. However, there are generally recommended actions when a laboratory produces an unacceptable result in a proficiency test. Guidance for actions by laboratories in response to unsuccessful performance on a proficiency test is given in 4.1.4.

**4.1.4** In schemes where the standard deviation for proficiency assessment is based on observed performance, when a result gives an action signal, the laboratory shall decide what investigations and corrective actions are appropriate, in consultation with the coordinator or an accreditation body if necessary. Unless there is a valid reason not to do so, the laboratory shall examine its procedures and identify one or more corrective actions that, according to staff in the laboratory, are likely to prevent the recurrence of such results. The laboratory may ask the coordinator for advice on possible causes of its problem, or may ask the coordinator to consult other experts. The laboratory shall take part in further rounds of the proficiency testing scheme to assess the effectiveness of the corrective actions. Appropriate corrective actions may be one of the following:

- a) checking that staff understand and follow the measurement procedure;
- b) checking that all details of the measurement procedure are correct;
- c) checking the calibration of equipment and the composition of reagents;
- d) replacing suspect equipment or reagents;
- e) comparative tests of staff, equipment and/or reagents with another laboratory.

The use of the results of proficiency tests by laboratory accreditation bodies is described in ISO/IEC Guide 43-2:1997, Clause 6.

#### 4.2 Guidelines for limiting the uncertainty of the assigned value

The assigned value  $X$  has a standard uncertainty  $u_X$  that depends on the method that is used to derive it, and also, when it is derived from tests in several laboratories, on the number of laboratories and, perhaps, on other factors. Methods for calculating the standard uncertainty of the assigned value are given in Clause 5.

The standard deviation for proficiency testing  $\hat{\sigma}$  is used to assess the size of estimates of laboratory bias found in a proficiency test. Methods for obtaining the standard deviation for proficiency testing are given in Clause 6 and criteria that compare it with estimates of laboratory bias are given in Clause 7.

If the standard uncertainty  $u_X$  of the assigned value is too large in comparison with the standard deviation for proficiency testing  $\hat{\sigma}$ , then there is a risk that some laboratories will receive action and warning signals because of inaccuracy in the determination of the assigned value, not because of any cause within the laboratories. For this reason, the standard uncertainty of the assigned value shall be established and shall be reported to laboratories participating in proficiency testing schemes (see ISO/IEC Guide 43-1:1997, A.1.4 and A.1.6).

If

$$u_X \leq 0,3\hat{\sigma} \quad (1)$$

then the uncertainty of the assigned value is negligible and need not be included in the interpretation of the results of the proficiency test.

If these guidelines are not met, then the coordinator shall consider the following.

- a) Look for a method for determining the assigned value such that its uncertainty meets the above guideline.
- b) Use the uncertainty of the assigned value in the interpretation of the results of the proficiency test (see 7.5 on  $E_n$  numbers or 7.6 on the  $z$ -score).
- c) Inform the participants in the proficiency test that the uncertainty of the assigned value is not negligible.

**EXAMPLE** Suppose that the assigned value  $X$  is determined as the average  $\bar{x}$  of the results of tests in 11 laboratories, and that the standard deviation for proficiency testing is determined as the standard deviation  $s$  of these same 11 results, so  $\hat{\sigma} = s$ . As a first approximation, the standard uncertainty of the assigned value in this situation may be estimated by  $u_X = s/\sqrt{11} = 0,3s$ , so that the requirement appears to be met. However, the requirement cannot be met in this situation with fewer than 11 laboratories. Further, the uncertainty of the assigned value will be larger than  $s/\sqrt{11}$  if the samples suffer from non-homogeneity or instability, or if there is a factor that causes a common bias in the results of the laboratories (e.g. if they all use the same reference standard).

#### 4.3 Guidelines for choosing the number of replicate measurements

Repeatability variation contributes to the variation between the laboratory biases in a proficiency test. If the repeatability variation is too large in comparison with the standard deviation for proficiency testing, then there

is a risk that repeatability variation will cause the results of the proficiency test to be erratic. In this situation, a laboratory could have large bias in one round, but not the next, and they will have difficulty identifying the cause.

For this reason, when it is considered desirable to limit the influence of repeatability variation, the number of replicate measurements  $n$  made by each laboratory in a proficiency test shall be chosen so that:

$$\sigma_r / \sqrt{n} \leq 0,3\hat{\sigma} \quad (2)$$

where  $\sigma_r$  is the repeatability standard deviation that has been established in a previous interlaboratory experiment.

The justification for the factor of 0,3 is that when this criterion is met, the repeatability standard deviation contributes no more than about 10 % of the standard deviation for proficiency testing.

Further, all laboratories shall carry out the same number of replicate measurements. (The methods of analysis given later in this International Standard assume that this requirement is met.) If the requirement inequality of (2) is not met, then the number of replicate measurements shall be increased, or the results of the proficiency test shall be interpreted with caution.

This approach assumes that laboratories have generally similar repeatability. Cases can arise where this is not so. In such cases, for the methods described in this International Standard to be applied, the following device may be used. The coordinator should fix the number of replicate measurements  $n$ , using a typical value for the repeatability standard deviation. Then, each laboratory should check that it satisfies Inequality (2) with its own repeatability standard deviation. If it does not, then it should modify its measurement procedure so that it obtains a test result as the average of some number of determinations chosen so that Inequality (2) is satisfied.

#### 4.4 Homogeneity and stability of samples (see ISO/IEC Guide 43-1:1997, 5.6.2 and 5.6.3)

Methods are given in Annex B for checking that the samples to be used in a proficiency test are adequately homogeneous and stable.

When a method of sample preparation is used such that the homogeneity criterion in Annex B is not met, then replicate samples shall be tested by the participants, or the standard deviation for proficiency testing shall include an allowance for the heterogeneity of the samples, as described in Annex B.

#### 4.5 Operationally defined measurement methods

With an operationally defined measurement method, the measurement result is defined by the measurement procedure. For example, the size distribution of a particulate material may be determined using either square-holed or round-holed sieves. There may be no good reason why one type of sieve should be preferred, but unless the type of sieve is specified, laboratories that use different types of sieves may obtain differing results. If a participant uses a different method from that used to establish the assigned value, then their results may show a bias when no fault in execution is present.

If participants are free to choose between operationally defined methods, no valid consensus may be evident amongst them. Two recourses are available to overcome this problem:

- a) when a standardized method is in routine use by the participants, it is used to establish the assigned value, and participants are instructed to use it in the proficiency test;
- b) a separate value of the assigned value is produced for each method used.

A similar situation arises when the measurand is specified, but not the procedure, and the same choice has to be made.

#### 4.6 Reporting of data (see ISO/IEC Guide 43-1:1997, 6.2.3)

For the purposes of the calculations required in proficiency testing, it is recommended that individual measurement results should not be rounded by more than  $\sigma_r/2$ .

Participants shall be asked to report the actual values of their measurement results. Measurement results shall not be truncated (i.e. results shall not be reported in the form "<0,1" or "less than the detection limit"). Likewise, when a negative result is observed, the actual negative value shall be reported (i.e. results shall not be reported as zero even when logically the measurement result cannot be negative). Participants shall be informed that if they report truncated results on a sample, or zeros when results are negative, then all the data for that sample will be excluded from the analysis. If necessary, the form used to report results may contain a box to allow a participant to indicate that a result is below the detection limit.

#### 4.7 Period of validity of the results of proficiency tests

The period of validity of the result obtained by a laboratory in a single round of a proficiency testing scheme is limited to the time that the laboratory performed the test. Thus, if a laboratory achieves a satisfactory result in a single round, the result shall not be used to support a claim that the laboratory obtained reliable data on any other occasion.

A laboratory that operates a quality system and achieves a history of satisfactory results in many rounds of a proficiency testing scheme shall be entitled to use the results as evidence that it is able to obtain consistently reliable data.

### 5 Determination of the assigned value and its standard uncertainty

#### 5.1 Choice of method of determining the assigned value

Five ways of determining the assigned value  $X$  are described in 5.2 to 5.6. The choice between these methods shall be the responsibility of the coordinator following consultation with technical experts as described in ISO/IEC Guide 43-1. The methods described in 5.5 and 5.6 are unlikely to be applicable when the number of laboratories participating in the scheme is small. The methods of calculating the standard uncertainty,  $u_X$ , of the assigned value given in this clause will usually be adequate for the applications for which they are used in this International Standard. Alternative methods may be used provided that they have a sound statistical basis and that the method used is described in the documented plan for the scheme.

The determination of the assigned value shall be the responsibility of the coordinator. The assigned value shall not be disclosed to the participants until they have reported their results to the coordinator. The coordinator shall prepare a report giving details of how the assigned value was obtained, the identities of laboratories involved in its determination, and statements of the traceability and measurement uncertainty of the assigned value.

The *Guide to the expression of uncertainty in measurement* gives guidance on the evaluation of measurement uncertainties.

This International Standard recommends the use of robust statistical methods when it is considered that they are the most appropriate methods to use (for example, as in 5.5 and 5.6). Alternatively, procedures that involve the detection and removal of outliers may be used provided that they have a sound statistical basis and the method that is used is reported. Guidance on the use of tests for outliers is given in ISO 5725-2.

#### 5.2 Formulation [see ISO/IEC Guide 43-1:1997, A.1.1, item a)]

##### 5.2.1 General

The test material may be prepared by mixing constituents in specified proportions, or by adding a specified proportion of a constituent to a base material. In this case, the assigned value  $X$  is derived by calculation from the masses used.

The approach is especially valuable when individual samples may be prepared in this way, and it is the proportion of the constituents or of the addition that is to be determined: there is then no need to prepare a bulk quantity and ensure that it is homogeneous. However, when formulation gives samples in which the addition is more loosely bonded than in typical materials, or in a different form, it may be preferable to use another approach.

### 5.2.2 Standard uncertainty $u_X$ of the assigned value

When the assigned value is calculated from the formulation of the test material, the standard uncertainty is estimated by combination of uncertainties using the approach described in the *Guide to the expression of uncertainty in measurement*. For example, in chemical analyses the uncertainties will usually be those associated with gravimetric and volumetric measurements.

The limitation of this method (in chemical analysis) is that care is needed to ensure that:

- a) the base material is effectively free from the added constituent, or that the proportion of the added constituent in the base material is accurately known;
- b) the constituents are mixed together homogeneously (where this is required);
- c) all sources of error are identified (e.g. it is not always realized that glass absorbs mercury compounds, so that the concentration of an aqueous solution of a mercury compound can be altered by its container);
- d) there is no interaction between the constituents and the matrix.

### 5.2.3 Example: Determination of the cement content of hardened concrete

In this case, concrete specimens may be prepared by weighing out quantities of the constituents (cement, aggregates and water) and mixing them together to form each concrete sample. The approach is satisfactory because the accuracy with the specimens can be prepared is far superior to that of the analytical method used to determine the cement content.

## 5.3 Certified reference values [see ISO/IEC Guide 43-1:1997, A.1.1 item b)]

### 5.3.1 General

When the material used in a proficiency test is a certified reference material (CRM), its certified reference value is used as the assigned value  $X$ .

### 5.3.2 Standard uncertainty $u_X$ of the assigned value

When a certified reference material is used as the test material, the standard uncertainty of the assigned value is derived from the information on uncertainty provided on the certificate.

The limitation of this approach is that it can be expensive to provide every participant in a proficiency test with a sample of a certified reference material.

### 5.3.3 Example: Los Angeles value of aggregates

The "Los Angeles value" is a measure of the mechanical strength of aggregates that are used for road construction, and results of the test are measured in "LA units". In an exercise to certify a reference material, a large number of samples of a particular aggregate were prepared, and some of these samples were used in an interlaboratory experiment involving 28 laboratories, allowing an assigned value of  $X_{\text{CRM}} = 21,62$  LA units to be established with a standard uncertainty of  $u_{X_{\text{CRM}}} = 0,26$  LA units. The remaining samples of this aggregate could be used in proficiency tests.

## 5.4 Reference values [see ISO/IEC Guide 43-1:1997, A.1.1 item c)]

### 5.4.1 General

In this approach, samples of the test material that is to be the reference material (RM) are prepared first, ready for distribution to the participants. A number of the samples are then selected at random and tested along with certified reference materials, in one laboratory, using a suitable measurement method, and under repeatability conditions (as defined in ISO 3534-2). The assigned value  $X_{RM}$  of the test material is then derived from a calibration against the certified reference values of the CRMs.

### 5.4.2 Standard uncertainty $u_X$ of the assigned value

When the assigned value of a test material is derived from the results of a series of tests on that material and on CRM, the standard uncertainty of the assigned value is derived from the test results, and the uncertainties of the certified reference values of the CRM. If the test material and the CRM are not similar (in matrix, composition and level of results), then the uncertainty arising from this is also to be included.

This method allows the assigned value to be established in a manner that is traceable to the certified values of the CRMs, with a standard uncertainty that can be calculated, and avoids the cost of distributing the CRM to all the participants. These are good reasons for preferring it to other methods. However, the method assumes that there are no interactions between the materials used and the test conditions.

The example in 5.4.3 illustrates how the required uncertainty may be calculated in the simple case when the assigned value of a test material is established by direct comparison with a single CRM.

### 5.4.3 Example: Los Angeles value of aggregates

The CRM described in the example in 5.3 may be used to determine the assigned value for an RM which is another, similar, aggregate. This determination requires a series of tests to be carried out, in one laboratory, on samples of the two aggregates, using the same measurement method, and under repeatability conditions. Let

$X_{CRM}$  is the assigned value for the CRM

$X_{RM}$  is the assigned value for the RM

$D_i$  is the difference (RM – CRM) between the average results for the RM and the CRM on the  $i^{\text{th}}$  samples

$\bar{D}$  is the average of the differences  $D_i$

Then

$$X_{RM} = X_{CRM} + \bar{D} \quad (3)$$

The standard uncertainty of the assigned value of the RM may be calculated as:

$$u_{X;RM} = \sqrt{u_{X;CRM}^2 + u_D^2} \quad (4)$$

Table 1 gives an example of data that might be obtained in such a series of tests, and shows how the standard uncertainty  $u_D$  of the difference is calculated.

With these results,

$$X_{RM} = 21,62 + 1,73 = 23,35 \text{ LA units} \quad (5)$$

and

$$u_{X,RM} = \sqrt{0,26^2 + 0,24^2} = 0,35 \text{ LA units} \quad (6)$$

where 0,26 is the standard uncertainty of the assigned value of the CRM (given in the example in 5.3), and 0,24 is the standard uncertainty of  $\bar{D}$ .

Table 1 — Calculation of the average difference between a CRM and a RM, and of the standard uncertainty of this difference

Sample	RM		CRM		Difference in average values RM – CRM LA units
	Test 1 LA units	Test 2 LA units	Test 1 LA units	Test 2 LA units	
1	20,5	20,5	19,0	18,0	2,00
2	21,1	20,7	19,8	19,9	1,05
3	21,5	21,5	21,0	21,0	0,50
4	22,3	21,7	21,0	20,8	1,10
5	22,7	22,3	20,5	21,0	1,75
6	23,6	22,4	20,3	20,3	2,70
7	20,9	21,2	21,5	21,8	-0,60
8	21,4	21,5	21,9	21,7	-0,35
9	23,5	23,5	21,0	21,0	2,50
10	22,3	22,9	22,0	21,3	0,95
11	23,5	24,1	20,8	20,6	3,10
12	22,5	23,5	21,0	22,0	1,50
13	22,5	23,5	21,0	21,0	2,00
14	23,4	22,7	22,0	22,0	1,05
15	24,0	24,2	22,1	21,5	2,30
16	24,5	24,4	22,3	22,5	2,05
17	24,8	24,7	22,0	21,9	2,80
18	24,7	25,1	21,9	21,9	3,00
19	24,9	24,4	22,4	22,6	2,15
20	27,2	27,0	24,5	23,7	3,00
Average difference, $\bar{D}$					1,73
Standard deviation					1,07
Standard uncertainty of $\bar{D}$ (standard deviation / $\sqrt{20}$ )					0,24
NOTE The data are measurements of the mechanical strength of aggregate, obtained from the Los Angeles (LA) test.					

## 5.5 Consensus values from expert laboratories [see ISO/IEC Guide 43-1:1997, A.1.1 item d)]

### 5.5.1 General

As with the *reference values* approach (5.4), samples of the test material are prepared first, ready for distribution to the participants. Some of these samples are then selected at random and analysed by a group of expert laboratories. Alternatively, the group of expert laboratories may be participants in a round of a proficiency testing scheme, when the assigned value and its uncertainty will be derived after the round is completed. The assigned value  $X$  is calculated as the robust average of the results reported by the group of expert laboratories, calculated using Algorithm A in Annex C.



Other calculation methods may be used in place of Algorithm A, provided that they have a sound statistical basis and the report describes the method that is used.

### 5.5.2 Standard uncertainty $u_X$ of the assigned value

When each of  $p$  expert laboratories reports a measurement  $x_i$  on the test material together with an estimate  $u_i$  of the standard uncertainty of the measurement, and the assigned value  $X$  is calculated as a robust average using Algorithm A, the standard uncertainty of the assigned value  $X$  is estimated as:

$$u_X = \frac{1,25}{p} \times \sqrt{\sum_{i=1}^p u_i^2} \quad (7)$$

When the expert laboratories do not report standard uncertainties, or when the uncertainties are not validated independently (e.g. by a laboratory accreditation body), the standard uncertainty of the assigned value shall be estimated as described in 5.6.

**NOTE** The factor 1,25 represents the ratio of the standard deviation of the median to the standard deviation of the arithmetic mean, for large samples ( $p > 10$ ) from a normal distribution. For normally distributed data, the standard deviation of a robust average calculated using the algorithm in Annex C is not known, but will fall somewhere between the standard deviation of the arithmetic mean and the standard deviation of the median, so the formula gives a conservative estimate of the standard uncertainty  $u_X$ . For  $p < 10$ , the appropriate factor is less than 1,25, so the formula is then doubly conservative.

The limitations of this approach are that there may be an unknown bias in the results of the group of expert laboratories, and the claimed uncertainties may not be reliable.

### 5.5.3 Example: Petrographic analysis of aggregates

This approach may be used when samples of aggregates are to be distributed, and the participants are to determine the petrographic composition of the samples. Classification of aggregates requires skill and experience, and there are no reference materials available, so in this case the consensus of a small group of experts may well be the best way to establish assigned values.

## 5.6 Consensus value from participants [see ISO/IEC Guide 43-1:1997, A.1.1 item e)]

### 5.6.1 General

With this approach, the assigned value  $X$  for the test material used in a round of a proficiency testing scheme is the robust average of the results reported by all the participants in the round, calculated using Algorithm A in Annex C.

Other calculation methods may be used in place of Algorithm A, provided that they have a sound statistical basis and the report describes the method that is used. For example, the calculation in C.1 may be stopped at Equation (53) when the median is obtained, and the median absolute deviation may be used in place of C.2.

This approach may be particularly useful with an operationally defined measurement method, provided that the method is standardized.

### 5.6.2 Standard uncertainty $u_X$ of the assigned value

When the assigned value is derived as a robust average calculated using Algorithm A, the standard uncertainty of the assigned value  $X$  is estimated as:

$$u_X = 1,25 \times s^* / \sqrt{p} \quad (8)$$

where  $s^*$  is the robust standard deviation of the results calculated using Algorithm A in Annex C. (Here a "result" for a participant is the average of all their measurements on the test material.)

The limitations of this approach are that:

- a) there may be no real consensus amongst the participants;
- b) the consensus may be biased by the general use of faulty methodology and this bias will not be reflected in the standard uncertainty of the assigned value calculated as described above.

Neither of these conditions is rare in the determination of trace constituents.

### 5.6.3 Example: Antibody concentrations

Table 2 gives data from a round of a proficiency test in which concentrations of three allergen-specific IgE (immunoglobulin E) antibodies were determined. Figure 2 displays these same data in histograms.

To apply Algorithm A, the data are first sorted into ascending order, initial robust estimates of the average and standard deviation are calculated, and then the iterative method of the algorithm is applied. Table 3 gives these calculations for the results for the allergen specific IgE antibody d1 from Table 2.

The calculations required by Algorithm A may be carried out in a spreadsheet as follows.

- a) **Step 1:** Enter the data in a column, in ascending order, as shown in Table 3 for Iteration 0. Calculate their average and standard deviation (10,91 and 3,13 in Table 3). Calculate initial values for the robust average and robust standard deviation (10,85 and 3,53 in Table 3) using the formulae given in C.1.
- b) **Step 2:** Copy the data into the next column, as shown in Table 3 for Iteration 1. Use the initial values for the robust average and robust standard deviation to calculate the cut-off values (5,56 and 16,15 in Table 3) using the formulae given in C.1. Replace data outside the cut-off values by the cut-off values (2,18 is replaced by 5,56 and 16,30 is replaced by 16,15). Calculate the new average and standard deviation of the altered data (11,03 and 2,81 in Table 3). According to the formulae given in Annex C, the robust average is the same as this average (11,03) and the robust standard deviation (3,19) is obtained by multiplying the standard deviation by 1,134.
- c) **Step 3:** With a spreadsheet, there is now no need to create further columns of data. Instead, change the calculation of the cut-off values at the top of the second column of data so that they use the robust average and robust standard deviation from the bottom of the same column. This will give the cut-off values (6,24 and 15,82) shown in Table 3 under Iteration 2. The calculation can then be completed by continuing to replace data outside the cut-off values by the cut-off values until the iterations converge. When data are replaced, the spreadsheet will automatically update the averages and standard deviations and cut-off values, but the changes in these values will become progressively smaller until they are no longer significant.

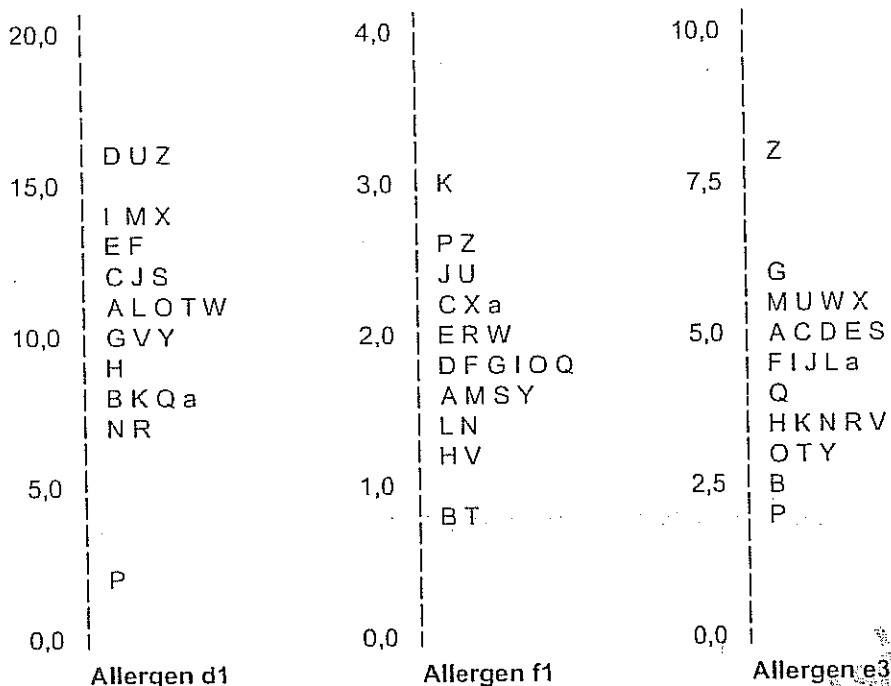
The robust averages and standard deviations for the other two allergen-specific IgE antibodies are calculated similarly.

It is interesting to note that the result for Laboratory P for d1 is not classed as an outlier or a straggler by Grubbs' test. Hence if one relies on the use of outlier tests as described in ISO 5725-2 in this example, the average and standard deviation are calculated from all the data, giving the values shown for *Iteration 0* in Table 3. With the robust method, the low result for Laboratory P and the high results for Laboratories D, U and Z have no influence on the values of the robust estimates. The bar-plots (see Figure 9 in 8.3) would identify the results of some laboratories as being worth investigation. For example, in Figure 9, Laboratory Z can be seen to give large positive *z*-scores at all three levels.

Table 2 — Concentrations of three allergen specific IgE antibodies (d1, f1 and e3) —  
Data as reported by  $p = 27$  laboratories

Laboratory	Concentrations		
	d1 kU/l	f1 kU/l	e3 kU/l
A	11,30	1,69	5,02
B	8,29	0,74	2,52
C	11,90	2,23	5,15
D	15,60	1,76	5,15
E	13,40	1,91	4,84
F	12,50	1,71	4,54
G	10,40	1,88	5,94
H	9,38	1,14	3,50
I	14,20	1,74	4,48
J	12,10	2,39	4,75
K	8,10	3,10	3,70
L	10,80	1,39	4,70
M	13,80	1,52	5,59
N	7,00	1,50	3,40
O	10,85	1,80	2,80
P	2,18	2,52	1,88
Q	8,39	1,83	3,80
R	6,95	1,92	3,52
S	11,80	1,58	4,86
T	10,90	0,80	2,80
U	16,30	2,39	5,60
V	9,71	1,21	3,33
W	10,50	1,93	5,35
X	13,60	2,23	5,53
Y	10,10	1,63	3,18
Z	16,07	2,69	8,22
a	8,47	2,16	4,64
$p$	27	27	27
Robust average $x^*$	11,03	1,83	4,35
Robust standard deviation $s^*$	3,04	0,50	1,25

NOTE The data are numbers of units (U) in thousands (k) per litre (l) of sample, where a unit is defined by the concentration of an international reference material.



NOTE 1 The data are numbers of units (U) in thousands (k) per litre (l) of sample, where a unit is defined by the concentration of an international reference material.

NOTE 2 The numerical values given in Table 2 are those that will be obtained when the calculation is carried out manually working to two decimal places.

Figure 2 — Concentrations of three allergen specific IgE antibodies — Histograms of data as reported (data from Table 2)

Table 3 — Concentrations of three allergen specific IgE antibodies —  
Calculation of the robust average and standard deviation for antibody d1

Iteration	0	1	2	3	4	5
$\delta = 1,5 s^*$	—	5,30	4,79	4,62	4,58	4,56
$x^* - \delta$	—	5,56	6,24	6,41	6,45	6,47
$x^* + \delta$	—	16,15	15,82	15,65	15,61	15,59
P	2,18	5,56	6,24	6,41	6,45	6,47
R	6,95	6,95	6,95	6,95	6,95	6,95
N	7,00	7,00	7,00	7,00	7,00	7,00
K	8,10	8,10	8,10	8,10	8,10	8,10
B	8,29	8,29	8,29	8,29	8,29	8,29
Q	8,39	8,39	8,39	8,39	8,39	8,39
a	8,47	8,47	8,47	8,47	8,47	8,47
H	9,38	9,38	9,38	9,38	9,38	9,38
V	9,71	9,71	9,71	9,71	9,71	9,71
Y	10,10	10,10	10,10	10,10	10,10	10,10
G	10,40	10,40	10,40	10,40	10,40	10,40
W	10,50	10,50	10,50	10,50	10,50	10,50
L	10,80	10,80	10,80	10,80	10,80	10,80
O	10,85	10,85	10,85	10,85	10,85	10,85
T	10,90	10,90	10,90	10,90	10,90	10,90
A	11,30	11,30	11,30	11,30	11,30	11,30
S	11,80	11,80	11,80	11,80	11,80	11,80
C	11,90	11,90	11,90	11,90	11,90	11,90
J	12,10	12,10	12,10	12,10	12,10	12,10
F	12,50	12,50	12,50	12,50	12,50	12,50
E	13,40	13,40	13,40	13,40	13,40	13,40
X	13,60	13,60	13,60	13,60	13,60	13,60
M	13,80	13,80	13,80	13,80	13,80	13,80
I	14,20	14,20	14,20	14,20	14,20	14,20
D	15,60	15,60	15,60	15,60	15,60	15,59
Z	16,07	16,07	15,82	15,65	15,61	15,59
U	16,30	16,15	15,82	15,65	15,61	15,59
Average	10,91	11,03	11,03	11,03	11,03	11,03
Standard deviation	3,13	2,81	2,72	2,69	2,68	2,68
New $x^*$	10,85	11,03	11,03	11,03	11,03	11,03
New $s^*$	3,53	3,19	3,08	3,05	3,04	3,04

NOTE The data are numbers of units (U) in thousands (k) per litre (l) of sample, where a unit is defined by the concentration of an international reference material.

## 5.7 Comparison of the assigned value

When the methods described in 5.2 to 5.4 are used to establish the assigned value  $X$ , after each round of a proficiency testing scheme, the robust average  $x^*$  derived from the results of the round shall be compared with the assigned value. When the methods described in 5.5 and 5.6 are used to establish the assigned value, the value shall, where possible, be compared with a reference value obtained by a competent laboratory. The standard uncertainty of the difference  $x^*-X$  shall be estimated as:

$$\sqrt{\frac{(1,25s^*)^2}{p} + u_X^2}$$

where

$s^*$  is the robust standard deviation;

$p$  is the number of laboratories.

If the difference is more than twice its uncertainty, the reason shall be sought. Possible reasons are:

- bias in the measurement method,
- a common bias in the results of the laboratories,
- failure to appreciate the limitations of the method when using the formulation method described in 5.2 approach,
- bias in the results of the "expert laboratories" when using the "consensus value from expert laboratories" approach, and
- biased participant method(s) or several biased laboratories when the robust consensus mean is used as the assigned value.

## 5.8 Missing values

When the number of replicate measurements  $n$  in a proficiency test is 2 or more, the first step in the analysis of the results will be to calculate the average and standard deviation of each laboratory's results. The averages are then used, for example, to calculate performance statistics as described in Clause 7 and to prepare histograms or bar-plots as described in 8.2 and 8.3. The standard deviations are used, for example, to prepare the plots of repeatability measures as described in 8.4 and 8.6.

Although all the participants may intend to obtain the same number of replicate measurements, they may not all report this number of measurements, for example, if some tests are spoilt and cannot be repeated. The following procedure is recommended when this happens.

If a laboratory reports at least  $0,59n$  replicate measurements, then the average and standard deviation of their measurements shall be included in the calculations and treated as if they had reported  $n$  measurements. The report shall state how many measurements they reported.

If a laboratory reports less than  $0,59n$  replicate measurements, then their results shall not be included in the calculation of statistics that affect other laboratories. For example, their results shall not be included in the calculation of the assigned value as described in 5.6, or the calculation of the standard deviation for proficiency assessment as described in 6.6. Their results may be used to calculate their own performance statistics as described in Clause 7, or included in the graphs described in Clause 8, but the report shall state how many measurements they reported and that this was fewer than the number required by the scheme.

**NOTE** The justification for the multiplier of 0,59 is as follows. The standard deviation of the average of  $n$  replicate measurements is  $\sigma_r/\sqrt{n}$ . If the number of replicate measurements is reduced this standard deviation increases, so if the actual number of replicate measurements is reduced from  $n$  to  $0,59n$ , the standard deviation is increased by a factor of 1,3.

This may be considered to be on the borderline of an acceptable increase in the standard deviation. Using the limit of 0,59 $\sigma$  will thus prevent the increase in the standard deviation being greater than this. There is clearly a degree of arbitrariness in the criterion used to derive this rule, so a coordinator may vary it in consultation with the members of the scheme if they wish to.

## 6 Determining the standard deviation for proficiency assessment (see ISO/IEC Guide 43-1:1997, A.2.1.3)

### 6.1 Choice of method

Five approaches to the problem of determining the standard deviation for proficiency assessment  $\hat{\sigma}$  are described in 6.2 to 6.6. The choice between these methods shall be the responsibility of the coordinator, in consultation with the members of the scheme and any relevant accreditation bodies, and taking into account any relevant regulations. The method described in 6.6 is unlikely to be applicable when the number of laboratories participating in the scheme is small. The determination of the standard deviation  $\hat{\sigma}$  shall be the responsibility of the coordinator. He shall prepare a report giving details of how the standard deviation was obtained.

NOTE ISO/IEC Guide 43-1 uses the symbol  $s$  as the standard deviation for proficiency assessment. This is consistent with normal usage when it represents a sample standard deviation. In this International Standard, the standard deviation for proficiency assessment is sometimes derived by other methods, so it is more appropriate to represent it by another symbol. Here  $\hat{\sigma}$  has been chosen.

### 6.2 Prescribed value

#### 6.2.1 General

The standard deviation for proficiency assessment may be set at a value required for a specific task of data interpretation, or it may be derived from a requirement given in legislation.

This approach has the advantage that the standard deviation for proficiency assessment is related directly to a "fitness for purpose" statement for the measurement method.

#### 6.2.2 Example: Aflatoxins in nuts, nut products, dried figs and dried fig products

There is legislation that states that a method used to test for aflatoxins should have a reproducibility coefficient of variation not larger than 50 % when the statutory limit is 10  $\mu\text{g}/\text{kg}$ . Thus, if a test material is used in a proficiency testing scheme with a content of aflatoxins of 10  $\mu\text{g}/\text{kg}$ , then the legislation implies that the reproducibility standard deviation with this material should be no more than 5  $\mu\text{g}/\text{kg}$ . In this case, it would be appropriate to set the standard deviation for proficiency assessment at 5  $\mu\text{g}/\text{kg}$  also.

### 6.3 By perception

#### 6.3.1 General

The standard deviation for proficiency assessment could be set at a value that corresponds to the level of performance that the coordinator and members of the scheme would wish laboratories to be able to achieve.

With this approach, the standard deviation for proficiency assessment becomes equivalent to a "fitness for purpose" statement for the measurement method.

When the standard deviation for proficiency assessment  $\hat{\sigma}$  is chosen by prescription or by perception, it is possible that the value chosen is not realistic in relation to the reproducibility of the measurement method. The following method may be used to check that the chosen value of  $\hat{\sigma}$  is realistic, provided that information on the repeatability and reproducibility of the method is available. Given

$\sigma_R$  is the reproducibility standard deviation, and

$\sigma_r$  is the repeatability standard deviation,

calculate the between-laboratory standard deviation as:

$$\sigma_L = \sqrt{\sigma_R^2 - \sigma_r^2} \quad (9)$$

and then calculate the value of the factor  $\phi$  by substituting values of  $\sigma_L$  and  $\sigma_r$  and the chosen value of  $\hat{\sigma}$  in Equation (10).

$$\hat{\sigma} = \sqrt{(\phi \times \sigma_L)^2 + (\sigma_r^2 / n)} \quad (10)$$

where  $n$  is the number of replicate measurements each laboratory is to perform.

If the value found for  $\phi$  is small (say  $\phi < 0,5$ ), it implies that the chosen value of  $\hat{\sigma}$  corresponds to a level of reproducibility that laboratories are unable to achieve in practice.

### 6.3.2 Example 1: Glucose measurement in human serum

Suppose that it is accepted that medical laboratories should be able to determine blood glucose levels within  $\pm 10\%$  of the assigned value, although for extremely low concentrations (below 60 mg/dl) a tolerance of  $\pm 6$  mg/dl is acceptable. This information may be used to calculate the standard deviation for proficiency assessment as:

a) for assigned values  $X$  below 60 mg/dl:

$$\hat{\sigma} = 6,0 / 3,0 = 2,0 \text{ mg/dl,}$$

b) for assigned values  $X$  above 60 mg/dl:

$$\hat{\sigma} = 0,1 X / 3,0 = 0,033 X \text{ mg/dl}$$

The factor of 3,0 introduced here corresponds to the critical value of 3,0 used in the interpretation of z-scores (see 7.4).

### 6.3.3 Example 2: Determination of the cement content of hardened concrete

The cement content of concrete is usually measured in terms of the mass in kilograms of cement per cubic metre of concrete (i.e. in  $\text{kg/m}^3$ ). In practice, concrete is produced in grades of quality that have cement contents 25  $\text{kg/m}^3$  apart, and it is desirable that laboratories should be able to identify the grade correctly. For this reason, it is desirable that the chosen value of  $\hat{\sigma}$  should be no more than one-half of 25  $\text{kg/m}^3$ . A precision experiment produced the following results, for a concrete with an average cement content of 260  $\text{kg/m}^3$ :  $\sigma_R = 23,2 \text{ kg/m}^3$  and  $\sigma_r = 14,3 \text{ kg/m}^3$ .

So

$$\sigma_L = \sqrt{23,2^2 - 14,3^2} = 18,3 \text{ kg/m}^3 \quad (11)$$

Taking  $n$  as 2, and substituting  $\sigma_L = 18,3 \text{ kg/m}^3$ ,  $\sigma_r = 14,3 \text{ kg/m}^3$  and  $\hat{\sigma} = 12,5 \text{ kg/m}^3$  into Equation (10) gives:

$$12,5 = \sqrt{(18,3 \phi)^2 + (14,3^2 / 2)} \quad (12)$$

from which one may calculate that  $\phi = 0,40$ . Thus choosing  $\hat{\sigma} = 12,5 \text{ kg/m}^3$  implies that laboratories are able to achieve a between-laboratory standard deviation lower by a factor of 0,4 than that found in the precision experiment. This is clearly unrealistic.



## 6.4 From a general model

### 6.4.1 General

The value of the standard deviation for proficiency testing may be derived from a general model for the reproducibility of the measurement method.

A disadvantage of this approach is that the true reproducibility of a particular measurement method may differ substantially from the value given by the model as the use of a general model implies that the reproducibility depends only on the level of the measurand, and not on the measurand, the measurement procedure, or the sample size.

### 6.4.2 Example: Horwitz curve

Horwitz [3] gives a general model for the reproducibility of analytical methods that may be used to derive the following expression for the reproducibility standard deviation:

$$\sigma_R = 0,02c^{0,8495} \quad (13)$$

where  $c$  is the concentration of the chemical species to be determined in percent (mass fraction).

## 6.5 From the results of a precision experiment

### 6.5.1 General

When the measurement method to be used in the proficiency testing scheme is standardized, and information on the repeatability and reproducibility of the method is available, the standard deviation for proficiency assessment  $\hat{\sigma}$  may be calculated using this information, as follows. Given

$\sigma_R$  is the reproducibility standard deviation, and

$\sigma_r$  is the repeatability standard deviation,

calculate the between-laboratory standard deviation as:

$$\sigma_L = \sqrt{\sigma_R^2 - \sigma_r^2} \quad (14)$$

and then calculate the standard deviation for proficiency assessment as:

$$\hat{\sigma} = \sqrt{\sigma_L^2 + (\sigma_r^2/n)} \quad (15)$$

where  $n$  is the number of replicate measurements each laboratory is to perform in a round of the scheme.

When the repeatability and reproducibility standard deviations are dependent on the average value of the test results, functional relations will have been derived by the methods described in ISO 5725-2. These relations should then be used to calculate values of the repeatability and reproducibility standard deviations appropriate to the assigned value that is to be used in the proficiency test.

### 6.5.2 Example: Determination of the cement content of hardened concrete

With the same data used in the example in 6.3, Equation (15) produces a standard deviation for proficiency testing of

$$\hat{\sigma} = \sqrt{18,3^2 + (14,3^2/2)} = 20,9 \text{ kg/m}^3 \quad (16)$$

assuming that  $n = 2$  replicate measurements are to be made.

## 6.6 From data obtained in a round of a proficiency testing scheme

### 6.6.1 General

With this approach, the standard deviation  $\hat{\sigma}$  used to assess the proficiency of participants in a round of a scheme is derived from the results reported by the participants in the same round. The standard deviation shall be the robust standard deviation of the results reported by all the participants, calculated using Algorithm A in Annex C. In this context, the result reported by a participant shall be the average of the  $n$  replicate measurements obtained by the participant in the round.

Other calculation methods may be used in place of Algorithm A, provided that they have a sound statistical basis and the report describes the method that is used.

A disadvantage of this approach is that the value of  $\hat{\sigma}$  may vary substantially from round to round, making it difficult to use values of the  $z$ -score for a laboratory to look for trends that persist over several rounds. This disadvantage may be overcome in an established scheme by using a robust pooled value of the standard deviations derived from a number of rounds, calculated using Algorithm S in Annex C.

### 6.6.2 Example: Antibody concentrations

Tables 2 and 3 provide an example of this approach.

## 6.7 Comparison of precision values derived from a proficiency test with established values

As a check on the performance of the participants, and to measure the benefit of the scheme to the participants, it is recommended that the coordinator applies the following procedure. The results obtained in each round of a proficiency testing scheme should be used to calculate estimates of the repeatability and reproducibility standard deviations of the measurement method, using the robust methods described in ISO 5725-5. These estimates should be plotted on graphs as time-series, together with values of the repeatability and reproducibility standard deviations obtained in precision experiments (if available).

These graphs should then be examined by the coordinator. If they show that the precision values obtained in the proficiency test differ by a factor of two or more from the values obtained in the precision experiment, then the coordinator should investigate why. If they show that the precision of the measurement method is not improving with time, then it suggests that:

- the participating laboratories are not investigating the causes of action and warning signals, or not implementing corrective actions properly;
- the participating laboratories are not able to identify the causes of action and warning signals;
- the method is in a state of statistical control and reliable conclusions may be based on the data given by the method.

## 7 Calculation of performance statistics

### 7.1 Estimates of laboratory bias [see ISO/IEC Guide 43-1:1997, A.2.1.4 item a)]

#### 7.1.1 General

Let  $x$  represent the result (or the average of the results) reported by a participant for the measurement of one characteristic of the test material in one round of a proficiency testing scheme.

Then an estimate of the bias  $D$  of the laboratory, when measuring that characteristic, may be calculated as:

$$D = x - X \quad (17)$$

where  $X$  is the assigned value.

Performance statistics that involve the absolute value  $|D|$  of the bias of the laboratory or  $D^2$  should not be used, because they conceal the sign of the bias.

### 7.1.2 Interpretation of laboratory biases

When a participant reports a result that gives rise to a laboratory bias greater than  $3,0 \hat{\sigma}$  or less than  $-3,0 \hat{\sigma}$ , then the result shall be considered to give an "action signal". Likewise, a laboratory bias above  $2,0 \hat{\sigma}$  or below  $-2,0 \hat{\sigma}$  shall be considered to give a "warning signal". A single "action signal" in one round, or two "warning signals" in successive rounds, shall be taken as evidence that an anomaly has occurred that requires investigation. This criterion is equivalent to that given in 7.4 for  $z$ -scores in the sense that it will give the same action and warning signals.

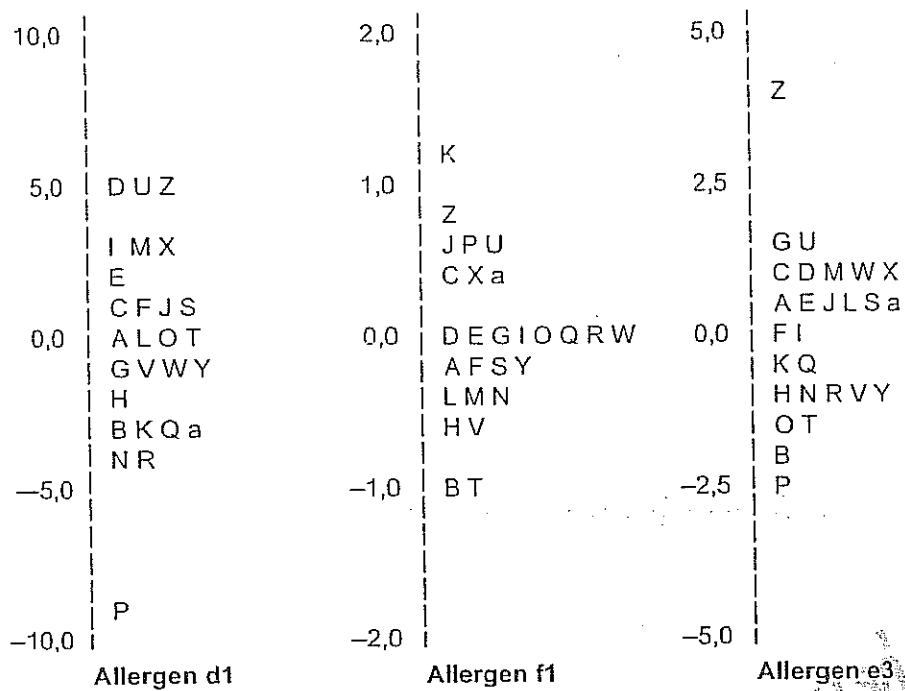
The justification for the use of the multipliers of 2,0 and 3,0 here (and in the other similar criteria given later) is as follows. If  $X$  and  $\hat{\sigma}$  are good estimates of the mean and standard deviation of the population from which the  $x$  values are derived, and the underlying distribution is normal, then the  $D$  values will be approximately normally distributed with a mean of zero and a standard deviation  $\hat{\sigma}$ . Under these circumstances only about 0,3 % of estimated laboratory biases would be expected to fall outside the range  $-3,0 \hat{\sigma} < D < 3,0 \hat{\sigma}$ , and only about 5 % would be expected to fall outside the range  $-2,0 \hat{\sigma} < D < 2,0 \hat{\sigma}$ . Because these probabilities are so low, it is unlikely that action signals will occur by chance when no real problem exists, so there is a reasonable chance of identifying the reason for an anomaly when an action signal is given.

When the standard deviation for proficiency assessment is fixed by either of the methods described in 6.2, 6.3 or 6.4, it may differ substantially from the reproducibility standard deviation, and the probabilities of 0,3 % and 5,0 % will then no longer apply.

When the standard deviation for proficiency assessment is fixed by either of the methods described in 6.2 or 6.3, it may be helpful to the participants for the appropriate performance statistic to be used, so that a direct comparison can be made with the prescribed or perceived performance requirement. For example, in the example in 6.3, where the goal for error in glucose samples is  $\pm 10$  % of the assigned value, one may follow the example, derive a standard deviation for proficiency assessment of 3,33 %, calculate  $z$ -scores and follow 7.4. Equivalently, one may calculate the laboratory bias as a percentage difference (as described in 7.2) and compare it directly with the goal of 10 %.

### 7.1.3 Example: Antibody concentrations

Table 4 shows the results of applying this method to the data from Table 2, and Figure 3 shows histograms of the estimates of laboratory biases. Comparison of Figures 1 and 2 shows that the laboratory biases have the same distribution as the original data (apart from small effects due to rounding), but they are centred on zero.



NOTE The data are numbers of units (U) in thousands (k) per litre (l) of sample, where a unit is defined by the concentration of an international reference material.

Figure 3 — Concentrations of three allergen specific IgE antibodies — Histograms of estimates of laboratory biases (data from Table 4)

Table 4 — Concentrations of three allergen specific IgE antibodies (d1, f1 and e3) — Estimates of laboratory bias

Laboratory	Concentrations		
	d1 kU/l	f1 kU/l	e3 kU/l
A	0,27	-0,14	0,67
B	-2,74	-1,09 W	-1,83
C	0,87	0,40	0,80
D	4,57	-0,07	0,80
E	2,37	0,08	0,49
F	1,47	-0,12	0,19
G	-0,63	0,05	1,59
H	-1,65	-0,69	-0,85
I	3,17	-0,09	0,13
J	1,07	0,56	0,40
K	-2,93	1,27 W	-0,65
L	-0,23	-0,44	0,35
M	2,77	-0,31	1,24
N	-4,03	-0,33	-0,95
O	-0,18	-0,03	-1,55
P	-8,85 W	0,69	-2,47
Q	-2,64	0,00	-0,55
R	4,08	0,09	-0,83
S	0,77	-0,25	0,51
T	-0,13	-1,03 W	-1,55
U	5,27	0,56	1,25
V	-1,32	-0,62	-1,02
W	-0,53	0,10	1,00
X	2,57	0,40	1,18
Y	-0,93	-0,20	-1,17
Z	5,04	0,86	3,87 A
a	-2,56	0,33	0,29
3,0 $\hat{\sigma}$	9,12	1,50	3,75
2,0 $\hat{\sigma}$	6,08	1,00	2,50
-2,0 $\hat{\sigma}$	-6,08	-1,00	-2,50
-3,0 $\hat{\sigma}$	-9,12	-1,50	-3,75

NOTE 1 When following concentration data, A = Action signal and W = Warning signal.

NOTE 2 The laboratory biases in this table have been derived from the data given in Table 2, using the robust averages in Table 2 as the assigned values for the three levels. The action and warning limits shown at the bottom of the table have been calculated, using the robust standard deviations in Table 2 as the standard deviations for proficiency assessment.

NOTE 3 The data are numbers of units (U) in thousands (k) per litre (l) of sample, where a unit is defined by the concentration of an international reference material.

7.2 Percentage differences [see ISO/IEC Guide 43-1:1997, A.2.1.4 item b)]

7.2.1 General

With the notation as in 7.1, percentage differences are calculated as:

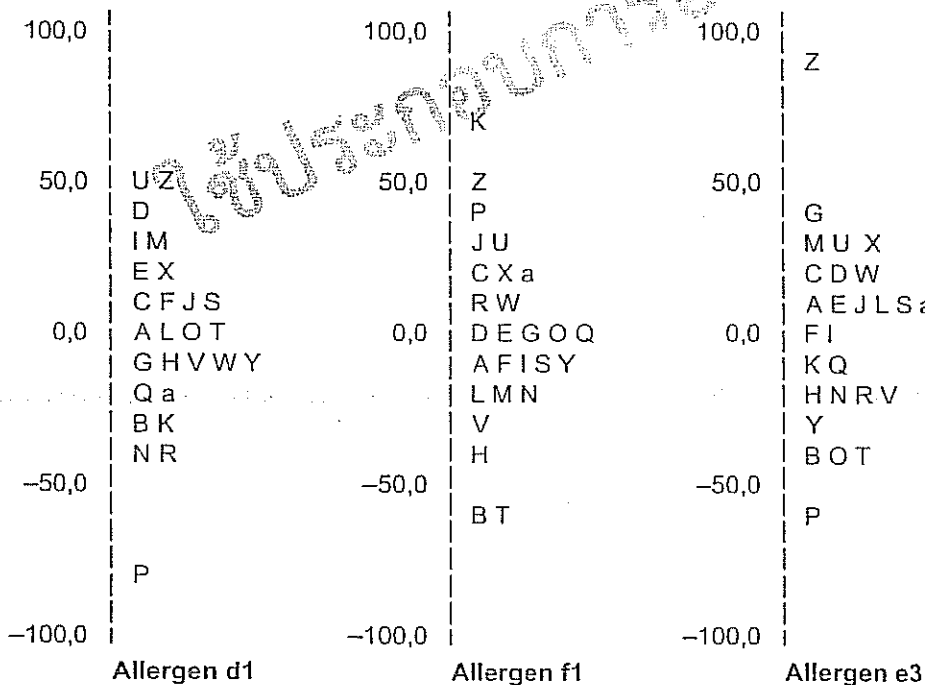
$$D_{\%} = 100 (x - X) / X \tag{18}$$

7.2.2 Interpretation of percentage differences

Percentage differences shall be interpreted using equivalent guidelines to those given for laboratory biases in 7.1, i.e. when a participant reports a result that gives rise to a percentage difference above  $300 \hat{\sigma} / X$  % or below  $-300 \hat{\sigma} / X$  %, then the result shall be considered to give an "action signal". Likewise, a percentage difference above  $200 \hat{\sigma} / X$  % or below  $-200 \hat{\sigma} / X$  % shall be considered to give a "warning signal". A single "action signal", or "warning signals" in two successive rounds shall be taken as evidence that an anomaly has occurred that requires investigation.

7.2.3 Example: Antibody concentrations

Table 5 shows the results of applying this method to the data from Table 2, and Figure 4 shows histograms of percentage differences. Comparison of Figures 1 and 3 shows that the percentage differences have the same distribution as the original data (apart from small effects due to rounding), but are centred on zero, just like the estimates of laboratory biases.



NOTE The data are numbers of units (U) in thousands (k) per litre (l) of sample, where a unit is defined by the concentration of an international reference material.

Figure 4 — Concentrations of three allergen specific IgE antibodies — Histograms of percentage differences (data from Table 5)

Table 5 — Concentrations of three allergen specific IgE antibodies (d1, f1 and e3) — Percentage differences

Laboratory	Percentage difference		
	d1 %	f1 %	e3 %
A	2	-8	15
B	-25	-60 W	-42
C	8	22	18
D	41	-4	18
E	21	4	11
F	13	-7	4
G	-6	3	37
H	-15	-38	-20
I	29	-5	3
J	-10	31	9
K	-27	69 W	-15
L	-2	-24	8
M	25	17	29
N	-37	18	-22
O	-2	-2	-36
P	-80 W	38	-57
Q	-24	0	-13
R	-37	5	-19
S	7	-14	12
T	-1	-56	-36
U	48	31	29
V	-12	-34	-23
W	-5	5	23
X	23	22	27
Y	-8	-11	-27
Z	46	47	89 A
a	-23	18	7
300 $\hat{\sigma}/X$	82,7	82,0	86,2
200 $\hat{\sigma}/X$	55,1	54,6	57,5
-200 $\hat{\sigma}/X$	-55,1	-54,6	-57,5
-300 $\hat{\sigma}/X$	-82,7	-82,0	-86,2

NOTE 1 When following concentration data, A = Action signal and W = Warning signal.

NOTE 2 The percentage differences in this table have been derived from the data given in Table 2, using the robust averages in Table 2 as the assigned values for the three levels. The action and warning limits shown at the bottom of the table have been calculated using the robust standard deviations in Table 2 as the standard deviations for proficiency assessment.

**7.3 Ranks and percentage ranks [see ISO/IEC Guide 43-1:1997, A.2.1.4 item c)]**

**7.3.1 General**

With results from  $p$  laboratories in a round of a proficiency test, the ranks are derived by assigning the rank of 1 to the laboratory that reports the lowest result, assigning 2 to the laboratory that reports the next lowest result, and so on, until the laboratory that reports the highest result is assigned the rank of  $p$ . If two or more results are equal, they are assigned the same average rank. For example, in Table 2, laboratories C and X both reported a concentration of 2,23 for f1. They share ranks 21 and 22, so in Table 6 they are both assigned a rank of 21,5. If the trial involves several measurands, ranks are assigned for each measurand separately.

If the ranks are denoted by  $i = 1, 2, \dots, p$ , then the percentage ranks are calculated as  $100 (i - 0,5) / p$  %. An example of the calculation of percentage ranks is shown in Table 6.

**7.3.2 Interpretation of ranks and percentage ranks**

The interpretation of ranks or percentage ranks does not involve an assumption that the data follow a particular probability distribution, and their derivation does not use the assigned value or the standard deviation for proficiency assessment. Hence, ranks and percentage ranks provide a simple method of identifying the laboratories that report the most extreme results. They are of particular use in the early rounds of a proficiency scheme when they can be used to identify the laboratories where improvements in performance are most likely to be achieved. However, the warning given in ISO Guide 43-1:1997, 6.6.5 should be noted: *"Reporting of performance by ranking laboratories in a table according to their performance is not recommended in proficiency testing. Therefore, ranking should only be used with extreme caution as it can be misleading and open to misinterpretation."*

ISO 13528:2005(E)



Table 6 — Concentrations of three allergen specific IgE antibodies (d1, f1 and e3) — Ranks and percentage ranks

Laboratory	Rank			Percentage rank		
	d1	f1	e3	d1 %	f1 %	e3 %
A	16	10	19	57	35	69
B	5	1	2	17	2	6
C	18	21,5	20,5	65	78	74
D	25	13	20,5	91	46	74
E	21	17	17	76	61	61
F	20	11	13	72	39	46
G	11	16	26	39	57	94
H	8	3	8	28	9	28
I	24	12	12	87	43	43
J	19	23,5	16	69	85	57
K	4	27	10	13	98	35
L	13	5	15	46	17	54
M	23	7	24	83	24	87
N	3	6	7	9	20	24
O	14	14	3,5	50	50	11
P	1	25	1	2	91	2
Q	16	15	11	20	54	39
R	2	18	9	6	65	31
S	17	8	18	61	28	65
T	15	2	3,5	54	6	11
U	27	23,5	25	98	85	91
V	9	4	6	31	13	20
W	12	19	22	43	69	80
X	22	21,5	23	80	78	83
Y	10	9	5	35	31	17
Z	26	26	27	94	94	98
a	7	20	14	24	72	50

#### 7.4 $z$ -scores [See ISO/IEC Guide 43-1:1997, A.2.1.4 item d)]

##### 7.4.1 General

With notation as in 7.2, the  $z$ -score is calculated as:

$$z = (x - \bar{X}) / \hat{\sigma} \quad (19)$$

where  $\hat{\sigma}$  is the standard deviation for proficiency assessment.

system is in operation,  $\zeta$ -scores shall be used only in conjunction with  $z$ -scores, as an aid for performance of laboratories, as follows. If a laboratory obtains  $z$ -scores that repeatedly exceed of 3,0, they may find it of value to examine their test procedure step by step and derive an jet for that procedure. The uncertainty budget will identify the steps in the procedure where ertainties arise, so that the laboratory can see where to expend effort to achieve an their  $\zeta$ -scores also repeatedly exceed the critical value of 3,0, it implies that their uncertainty include all significant sources of uncertainty (i.e. they are missing something important).

as a large bias and its uncertainty interval  $X \pm U_x$  does not include the assigned value, then it large  $\zeta$ -score or  $E_n$  number.

ay be defined as:  $E_{z-} = \frac{x - (X - U_x)}{U_x}$  and  $E_{z+} = \frac{x - (X + U_x)}{U_x}$

assigned value derived using 7.4 or 7.6, or a reference value derived using 7.5, and  $U_x$  is the tainty of  $X$ .

atory's value and  $U_x$  is the expanded uncertainty of  $x$ .  $U_x$  can be  $2\hat{\sigma}$  or  $3\hat{\sigma}$  in 7.4 and 7.6 and

compare values of  $E_z$  with a critical value of 1,0:

$E_{z-}$  and  $E_{z+}$  fall within the range  $-1,0$  to  $1,0$ , the laboratory's performance is satisfactory;

of  $E_{z-}$  and  $E_{z+}$  falls outside the range  $-1,0$  to  $1,0$ , the laboratory's performance is ble;

and  $E_{z+}$  are both lower than  $-1,0$  or both higher than  $1,0$ , the laboratory's performance is .ory.

### ample of the analysis of data when uncertainties are reported

il

is those shown in Figures 6 and 7 provide helpful summaries of laboratories' results and their .laboratories that have the largest laboratory biases will also have the largest  $z$ -scores. (They he either end of the graph.)

authentic data obtained in an exercise in which 181 laboratories reported results with or lead in water. The data were reported as mol/l. They have been multiplied by  $10^{10}$  to give anaged values. This means that results do not have either the same uncertainty or the same ult has its own bias  $D_i$  resulting from the sum of *Method bias* and *Laboratory bias*; moreover, has of course its own Reproducibility standard deviation  $\sigma_R$ . The laboratories used various re determinations, and calculated the uncertainties themselves. The uncertainties are treated ded uncertainties. The data are given in Table 8 as reported, except that they have been sorted ults are in ascending order, and the laboratories are numbered in this order.  $U = 0$  may indicate ort uncertainty. There are several values in the table of doubtful validity that would merit further n practice, and it is not possible to show the most extreme results on the graphs. In particular, values are not included in the graphs although they are used in the calculations. These data ample of where negative results have been reported as required by 4.6, even though negative are not logically possible.

### 7.9.2 The assigned value and its uncertainty

The assigned value is calculated as described in 5.6 as the robust average of the results, using Algorithm A in Annex C. This gives an assigned value

$$X = x^* = 605 \times 10^{-10} \text{ mol/l} \quad (24)$$

and a robust standard deviation

$$s^* = 142 \times 10^{-10} \text{ mol/l} \quad (25)$$

According to 5.6, the standard uncertainty of this assigned value is

$$u_X = 1,23 \times s^* / \sqrt{181} = 13 \times 10^{-10} \text{ mol/l} \quad (26)$$

### 7.9.3 The standard deviation for proficiency assessment

The standard deviation for proficiency assessment is obtained as described in 6.6 as the robust standard deviation

$$\hat{\sigma} = s^* = 142 \times 10^{-10} \text{ mol/l} \quad (27)$$

### 7.9.4 Guidelines for interpreting the uncertainty of the assigned value

According to the guidelines in 4.2, the uncertainty of the assigned value is negligible if

$$u_X \leq 0,3\hat{\sigma} \quad (28)$$

With  $u_X = 1,25 \times s^* / \sqrt{p}$  and  $\hat{\sigma} = s^*$  as in this example, it can be calculated that Equation (28) is satisfied with  $p > 16$ . With  $p = 181$  laboratories participating, the criterion is easily satisfied. There is therefore no point with these data in considering z'-scores as described in 7.6.

### 7.9.5 Analysing data for a large number of laboratories using Normal probability plots

Figure 6 shows the results for the 181 laboratories plotted against their percentage ranks (calculated as described in 7.3), using a Normal probability scale for the percentage ranks. Results below 0 mol/l or above  $1600 \times 10^{-10}$  mol/l are not included in this figure.

The z-scores may be calculated as  $z = (x - 605)/142$ . When a laboratory achieves a z-score above 3,0 or below -3,0, the value of the z-score is shown on this figure alongside the corresponding value.

The cumulative distribution function for the Normal distribution with mean  $605 \times 10^{-10}$  mol/l and standard deviation  $142 \times 10^{-10}$  mol/l is also shown on the figure as the dashed straight line.

The final "cut-off" values used in the robust algorithm are:

$$x^* - 1,5s^* = 605 - 1,5 \times 142 = 392 \times 10^{-10} \text{ mol/l} \quad (29)$$

and

$$x^* + 1,5s^* = 605 + 1,5 \times 142 = 818 \times 10^{-10} \text{ mol/l} \quad (30)$$

It can be seen on Figure 6 that the points veer away from the dashed line outside this range. The implication of this is that all the results cannot be considered to be drawn from the same normal population. Points away from the dashed line are drawn from a population with a larger variance than points close to the line.

It can also be seen in the figure that results that give  $z$ -scores above 3,0, or below  $-3,0$ , are well away from the dashed line. This supports a conclusion to treat these  $z$ -scores as giving rise to "action" signals. (It is possible with a large number of results that  $z$ -scores above 3,0 or below  $-3,0$  are obtained, but when plotted as in Figure 6 all the points lie close to the dashed line. In such a case, the figure would not support a conclusion to treat the  $z$ -scores as giving rise to action signals.)

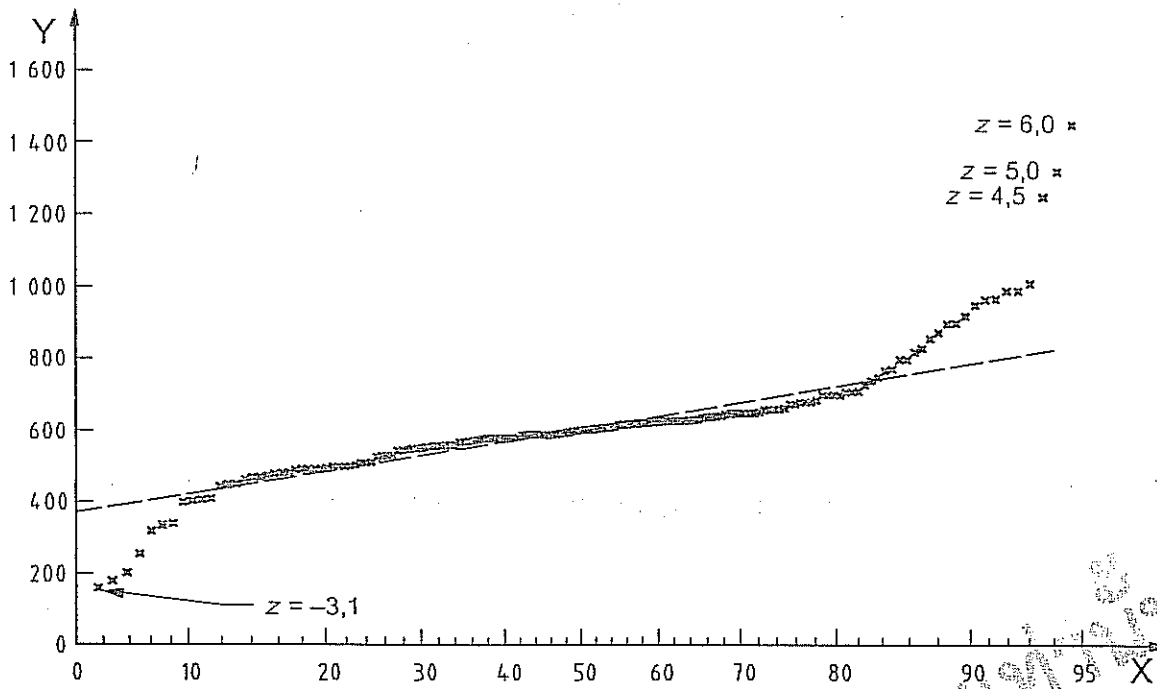
Figure 7 shows the results for just those laboratories that give  $z$ -scores within the range  $\pm 3,0$ , together with vertical lines that indicate the uncertainties that they reported. (Results for two laboratories, numbers 24 and 112, that reported very high uncertainties, have not been included in the figure.) The continuous, horizontal line on this figure represents the assigned value, and the dashed lines on either side of it represent the expanded uncertainty of the assigned value.

If the laboratories carried out valid calculations of the expanded uncertainties of their results, then nearly all of the vertical lines in Figure 7 would cut the region defined by the assigned value  $\pm$  its expanded uncertainty. However, it can be seen in the figure that there are many laboratories, on both sides of the assigned value, whose vertical lines do not reach this region. This implies that many laboratories have not carried out valid uncertainty calculations; very probably they have not included all the important sources of uncertainty in their calculations.

สำนักงานมาตรฐาน  
การบริการลูกค้า

Table 3 — Determinations of the lead content of water (mol/l  $\times 10^{10}$ ) by 181 laboratories, together with values for the expanded uncertainties ( $U$ ) of the results as reported by the laboratories

Lab	Result	$U$	Lab	Result	$U$	Lab	Result	$U$	Lab	Result	$U$
1	-960 000	0	51	545	43	101	618	224	151	740	20
2	-12 100	0	52	545	123	102	618	170	152	748	3
3	-4 800	0	53	550	8	103	620	25	153	767	113
4	-3 860	0	54	550	55	104	620	40	154	772	213
5	-1 500	0	55	550	5	105	621	6	155	800	60
6	-1 010	0	56	555	79	106	622	9	156	800	150
7	-1 000	0	57	556	30	107	622	6	157	821	203
8	-1 000	0	58	557	28	108	623	18	158	830	10
9	-965	0	59	557	28	109	625	15	159	857	27
10	-483	0	60	559	26	110	626	5	160	874	200
11	160	20	61	560	7	111	627	0	161	898	59
12	180	20	62	560	60	112	627	1 010	162	900	100
13	203	0	63	569	116	113	627	15	163	920	140
14	256	13	64	570	86	114	628	3	164	950	110
15	319	0	65	571	16	115	629	26	165	965	0
16	335	18	66	572	40	116	630	40	166	968	0
17	340	180	67	574	35	117	630	580	167	990	0
18	400	20	68	578	0	118	632	50	168	990	80
19	404	36	69	579	52	119	637	96	169	1 010	0
20	407	0	70	579	35	120	639	83	170	1 250	140
21	410	97	71	579	8	121	640	130	171	1 320	410
22	444	58	72	579	10	122	640	77	172	1 450	460
23	450	20	73	579	17	123	642	20	173	1 640	241
24	450	3 400 000	74	579	87	124	647	63	174	1 900	46
25	463	19	75	580	150	125	647	0	175	2 413	20
26	470	10	76	582	122	126	650	160	176	2 460	0
27	470	30	77	589	57	127	650	30	177	2 900	900
28	474	0	78	589	10	128	650	80	178	10 000	0
29	480	100	79	590	0	129	650	48	179	386 000	31 000
30	482	122	80	590	0	130	650	30	180	670 000	60 000
31	483	241	81	590	45	131	653	5	181	630 $\times 10^6$	60 $\times 10^6$
32	490	60	82	590	60	132	658	27			
33	492	25	83	590	0	133	660	20			
34	492	1	84	591	112	134	660	120			
35	493	24	85	591	9	135	660	34			
36	493	5	86	594	4	136	663	32			
37	495	0	87	594	119	137	675	280			
38	500	70	88	597	9	138	675	0			
39	500	10	89	600	20	139	680	8			
40	500	10	90	600	300	140	680	50			
41	501	75	91	603	60	141	680	70			
42	504	0	92	603	24	142	685	0			
43	510	130	93	603	13	143	700	0			
44	510	110	94	604	18	144	700	100			
45	512	6	95	608	30	145	700	110			
46	526	26	96	608	8	146	700	300			
47	530	9	97	609	9	147	708	44			
48	530	40	98	610	61	148	709	48			
49	530	60	99	613	22	149	710	100			
50	545	30	100	618	7	150	729	41			

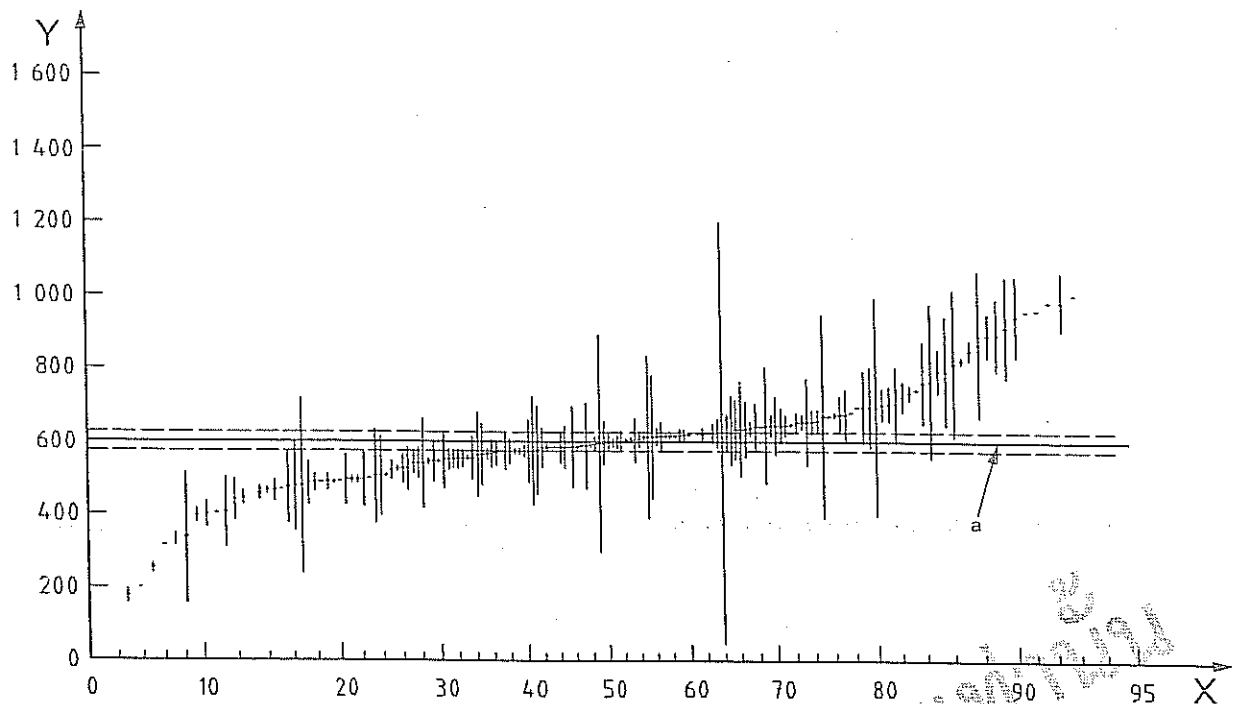


**Key**

- X laboratory-percentage rank, %
- Y lead content, mol/l × 10<sup>10</sup>

NOTE The results for 19 laboratories are not included.

**Figure 6 — Normal probability plot of results of determinations of the lead content of water by 162 laboratories**

**Key**

X laboratory-percentage rank, %  
 Y lead content, mol/l  $\times 10^{10}$

<sup>a</sup> Assigned value  $\pm$  its expanded uncertainty.

NOTE The results for 25 laboratories are not included.

**Figure 7 — Normal probability plot of expanded uncertainties for determinations of the lead content of water by 156 laboratories**

### 7.10 Combined performance scores

It is common, within a single round of a proficiency testing scheme, for results to be obtained for more than one test item or for several measurands. In this situation, the results for each test item and for each measurand shall be interpreted as described in 7.2 to 7.9, i.e. the results for each test item and each measurand shall be analysed separately. There are applications when two or more materials with specially designed levels are included in a scheme to measure other aspects of performance, such as repeatability or linearity. In such instances, the coordinator shall provide participants with complete descriptions of the statistical design and procedures that are used. If two similar materials have been incorporated in the scheme with the intention of treating them as a Youden Pair, the special methods described in 8.5 shall be applied.

Further, it is recommended that the graphical methods described in Clause 8 should also be used when results are obtained for more than one test item or for several measurands. They combine scores in ways that do not conceal high values of individual scores, and they may reveal additional information on the performance of laboratories, such as correlation between results for different measurands, not apparent in tables of the individual scores.

In schemes that involve a large number of measurands, a count of the numbers of action and warning signals may be used to allow the laboratories that obtain one or more such signals to be identified. They can be provided with a report containing detailed results using the methods described in 7.2 to 7.9. Laboratories that obtain no signal can be provided with only a brief report.

**NOTE** The use of composite scores when there is more than one test item for the same measurand (the average or summed  $z$ -score, or the average or summed absolute difference or squared difference), or the use of composite scores when there are results for more than one measurand (the average absolute  $z$ -score, or the average absolute difference relative to the evaluation limits) is not recommended. The average (or summed)  $z$ -score has the serious failing that a high score on one test item can be concealed if other scores are low or if another score is also high but has the opposite sign. The average (or summed) absolute difference and the sum of squared differences also have the serious failing that a high score on one test item can be concealed if other scores are low. The average absolute  $z$ -score and the average absolute difference relative to the evaluation limits also have this same failing.

## 8 Graphical methods for combining performance scores for several measurands from one round of a proficiency test (see ISO/IEC Guide 43-1:1997, A.2.2.1)

### 8.1 Application

The coordinator shall consider using the performance scores obtained in each round of a proficiency testing scheme to prepare graphs, as described in 8.2 and 8.3. The use of  $z$ -scores in these graphs has the advantage that they can be drawn using standardized axes, thereby simplifying their presentation and interpretation. Graphs shall be made available to the participants, enabling each participant to see where his own results fall in relation to those obtained by the other participants. Letter codes or number codes shall be used to represent the participants so that each participant is able to identify his own results but not able to determine which participant obtained any other result. The graphs shall also be made available to the coordinator, enabling them to judge the overall effectiveness of the scheme, and to see if there is a need for reviewing the criterion used to assess proficiency.

### 8.2 Histograms of performance scores

#### 8.2.1 General

To prepare a histogram of  $z$ -scores, collect the  $z$ -scores for the measurement of a characteristic from one round of a proficiency testing scheme into a histogram as shown in Figure 8. Use an interval in the histogram of about 0,3 to 0,5 so that the histogram gives a good visual impression. Draw lines at  $\pm 2,0$  and  $\pm 3,0$  to represent the proficiency assessment criterion. Use a range for the histogram of about  $\pm 6,0$ . If the results do not span this range, or if the central points are tightly clustered, the analyst may consider an alternative range.

When histograms of laboratory biases or percentage differences are preferred, the equivalent histogram intervals and warning and action limits are as below. In these cases, it may be simpler to derive the action limits directly from the prescribed or perceived requirements instead of calculating them from the standard deviation for proficiency assessment.

Table 9 — Warning and action limits for performance scores

Performance statistic	Histogram interval	Warning limits	Action limits	Histogram range
Laboratory bias	$0,3 \hat{\sigma}$ to $0,5 \hat{\sigma}$	$\pm 2,0 \hat{\sigma}$	$\pm 3,0 \hat{\sigma}$	$\pm 6,0 \hat{\sigma}$
Percentage difference	$30 \hat{\sigma} / X$ to $50 \hat{\sigma} / X$	$\pm 200 \hat{\sigma} / X$	$\pm 300 \hat{\sigma} / X$	$\pm 600 \hat{\sigma} / X$
$z$ -score	0,3 to 0,5	$\pm 2,0$	$\pm 3,0$	$\pm 6,0$

Histograms are a suitable method of graphical presentation when the number of characteristics measured is small, or when a number of dissimilar characteristics are measured. Individual participants can identify the position of their own scores and assess their performance and the need to investigate their methods. A participant who has achieved a high  $z$ -score is able to use a histogram to see how exceptional his score is, in comparison with the scores achieved by the other participants.



The coordinator is able to use a histogram to see how frequently participants fail to satisfy the proficiency assessment criterion. If the tails of the histogram extend outside the  $\pm 3,0$  limits, then the fault would appear to lie with the measurement method (or methods) being used rather than with individual participants. The measurement method (or methods) should be improved, or the proficiency assessment criterion should be relaxed (by increasing  $\hat{\sigma}$ ). If the histogram lies within the  $\pm 2,0$  limits, with perhaps one or two isolated  $z$ -scores outside these limits, then it suggests that the proficiency assessment criterion could be made more strict (by reducing  $\hat{\sigma}$ ).

### 8.2.2 Example: Antibody concentrations

The  $z$ -scores for d1 are shown in Figure 8 in the form of a histogram.

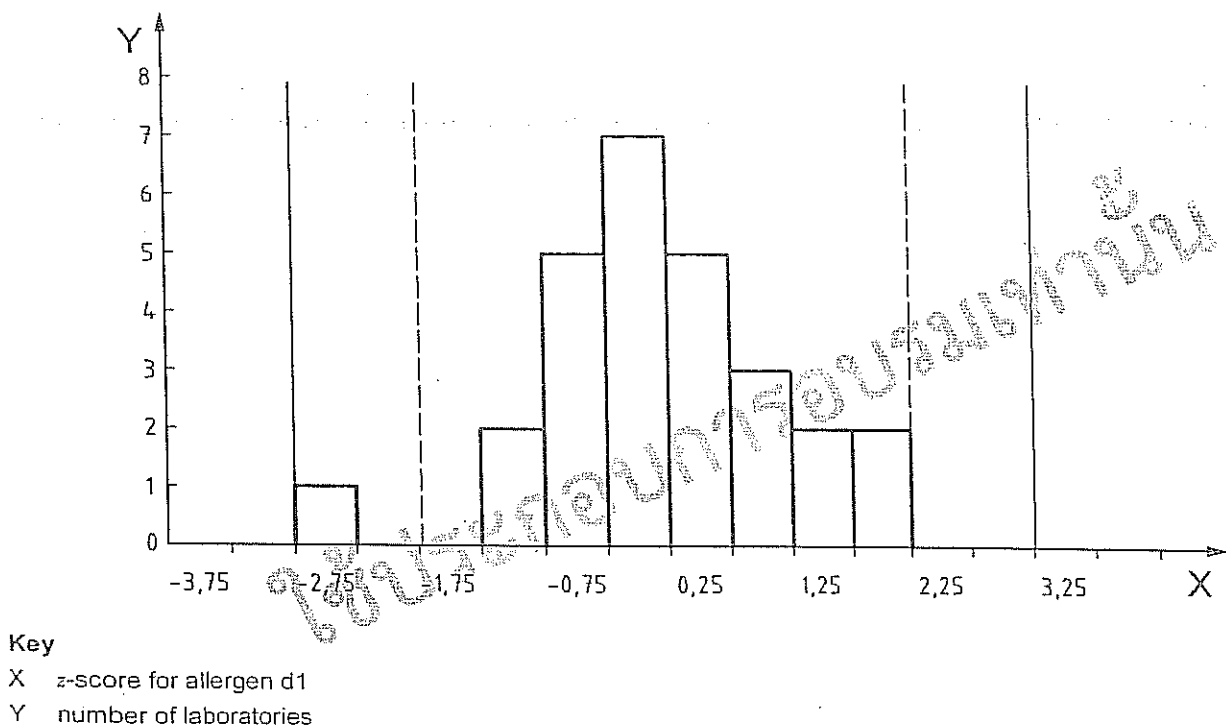


Figure 8 — Histogram of  $z$ -scores for one round of a proficiency test (data for allergen d1 from Table 7)

## 8.3 Bar-plots of standardized laboratory biases

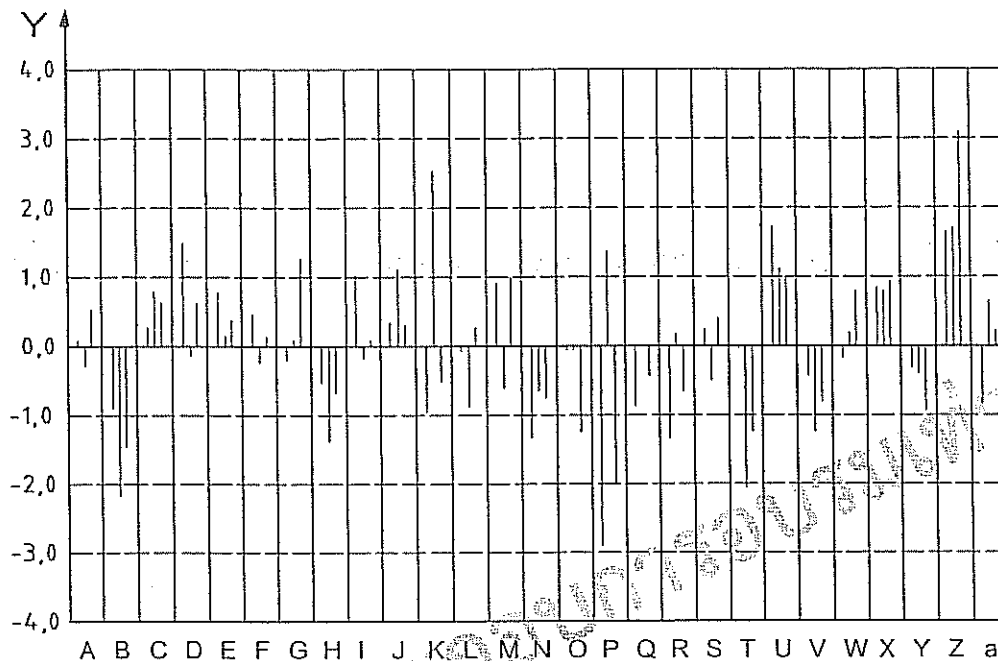
### 8.3.1 General

To prepare a bar-plot, collect the  $z$ -scores into a bar-plot as shown in Figure 9, in which the  $z$ -scores for each participant are grouped together. (The  $z$ -score is the same statistic as that referred to as the  $h$ -statistic in ISO 5725-2, and the bar-plots described here are the same graphs as the plots of  $h$ -statistics described in ISO 5725-2.)

Bar-plots are a suitable method of presenting the  $z$ -scores for a number of similar characteristics in one graph. They will reveal if there is any common feature in the  $z$ -scores for a participant, for example if he achieves several high  $z$ -scores indicating generally poor performance.

### 8.3.2 Example: Antibody concentrations

The  $z$ -scores from Table 7 are shown plotted as a bar-chart in Figure 9. From this graph, laboratories B and Z can see that they should look for a cause of bias that affects all three levels by approximately the same amount, whereas laboratories K and P can see that in their case the sign of the  $z$ -score depends on the type of antibody.



Key  
Y  $z$ -score

NOTE "A" to "a" indicates the laboratory number.

Figure 9 — Bar-chart of  $z$ -scores (4,0 to -4,0) for one round of a proficiency test in which the participants determined the concentrations of three allergen specific IgE antibodies (data from Table 7)

### 8.4 Bar-plots of standardized repeatability measurements

When replicate determinations are made in a round of a proficiency test, the results may be used to calculate a graph of  $k$ -statistics, as described in ISO 5725-2.

### 8.5 Youden Plot

#### 8.5.1 General

When samples of two similar materials have been tested in a round of a proficiency scheme, the Youden Plot provides a very informative graphical method of studying the results. It is constructed by plotting the  $z$ -scores obtained on one of the materials against the  $z$ -scores obtained on the other material. A confidence ellipse, calculated as described in 8.5.2, is used as an aid to interpretation of the plot. A Youden Plot for the original data, laboratory biases, or percentage biases may be derived from the  $z$ -scores as explained in Note 1 below.

When a Youden Plot is constructed, interpret it as follows.

- a) Inspect the plot for points that are well-separated from the rest of the data. If a laboratory is not following the test method correctly, so that its results are subject to bias, a point will be given far out along the major axis of the ellipse. Such a point can also occur if a laboratory suffers a large variation from time to time in the level of its results. Points far away from the major axis represent participants whose repeatability is poor.
- b) Inspect the plot to see if there is evidence of a general relationship between the results for the two materials. If there is, then it shows that there is a cause of between-laboratory variation that is common to many of the laboratories, and provides evidence that the measurement method has not been adequately specified. Investigation of the test method may then allow the reproducibility of the method to be generally improved. The rank correlation test described below may be used to test if the relationship between the two materials is statistically significant. The rank correlation coefficient is preferred here to the correlation coefficient as the latter would be more sensitive to non-normality in the data.

### 8.5.2 Confidence ellipse (based on the method of Jackson [2])

#### 8.5.2.1 General

Call the two materials A and B, and denote the results obtained on A by:

$$x_{A,1}, x_{A,2}, \dots, x_{A,p}$$

and those obtained on B by:

$$x_{B,1}, x_{B,2}, \dots, x_{B,p}$$

where  $p$  is the number of laboratories.

Calculate the averages and standard deviations of the two sets of data:

$$\bar{x}_A, \bar{x}_B, s_A, s_B$$

and the correlation coefficient  $\rho$ . Calculate the  $z$ -scores for the two materials as:

$$z_{A,i} = (x_{A,i} - \bar{x}_A) / s_A \quad \text{where } i = 1, 2, \dots, p \quad (31)$$

$$z_{B,i} = (x_{B,i} - \bar{x}_B) / s_B \quad \text{where } i = 1, 2, \dots, p \quad (32)$$

and calculate the combined scores for the two materials:

$$z_{A,B,i} = \sqrt{z_{A,i}^2 - 2\hat{\rho} z_{A,i} z_{B,i} + z_{B,i}^2} \quad (33)$$

Define standardized variables as:

$$z_A = (x_A - \bar{x}_A) / s_A \quad (34)$$

$$z_B = (x_B - \bar{x}_B) / s_B \quad (35)$$

In terms of the standardized variables, the confidence ellipse may be written in terms of Hotelling's  $T^2$ :

$$z_A^2 - 2\hat{\rho} z_A z_B + z_B^2 = (1 - \hat{\rho}^2) T^2 \quad (36)$$

where

$$T^2 = 2\left\{ \frac{(p-1)}{(p-2)} \right\} F_{(1-\alpha)}(2, p-1) \quad (37)$$

Here,  $F_{(1-\alpha)}(2, p-1)$  is the tabulated  $(1-\alpha)$ -fractile of the  $F$ -distribution with 2 and  $(p-1)$  degrees of freedom. The ellipse may be drawn on a graph that has the  $z$ -scores  $z_A$  and  $z_B$  as the axes by plotting a series of points for  $-T \leq z_A \leq T$  with:

$$z_B = \hat{\rho} z_A \pm \sqrt{(1 - \hat{\rho}^2)(T^2 - z_A^2)} \quad (38)$$

NOTE 1 To plot the confidence ellipse on a graph with axes that show the original units of measurement, transform the above series of points back to the original units using:

$$x_A = \bar{x}_{A..} + s_A \times z_A$$

$$x_B = \bar{x}_{B..} + s_B \times z_B$$

To plot the confidence ellipse on a graph with axes that show laboratory biases  $D_A$  and  $D_B$ , transform the above series of points using:

$$D_A = s_A \times z_A$$

$$D_B = s_B \times z_B$$

To plot the confidence ellipse on a graph with axes that show percentage differences  $D_{A\%}$  and  $D_{B\%}$ , transform the above series of points using:

$$D_{A\%} = 100s_A \times z_A / x_A$$

$$D_{B\%} = 100s_B \times z_B / x_B$$

The combined  $z$ -scores may be used as an aid to interpreting the Youden Plot. The highest combined  $z$ -scores correspond to the highest significance levels,  $100\alpha\%$  in the calculation of the confidence ellipse, so the combined  $z$ -scores may be used to identify the most extreme points on the Youden Plot. On occasion, it may be necessary to exclude one or more outlying points and re-calculate the ellipse: the combined  $z$ -scores may then be used as an aid to identifying the points to exclude.

NOTE 2 There is a need for a robust method of calculating the ellipse, but the details of such a method have not yet been worked out. The cut-off value may be calculated by noting that  $(z_{A,B,i})^2 / (1 - \hat{\rho}^2)$  has approximately the chi-squared distribution with 2 degrees of freedom, but the correction factor may have to be derived by simulation.

#### 8.5.2.2 Example: Antibody concentrations

Table 10 shows data obtained by testing two similar samples for antibody concentrations, and the calculations required to derive the confidence ellipse. With  $p = 29$  laboratories, and using a significance level of  $100\alpha\% = 5\%$ ,  $F_{(1-\alpha)}(2, p-1) = 3,34$ . Hence  $T = 2,632$  and, in terms of the standardized variables, the 95% confidence ellipse may be written:

$$z_A^2 - 1,412 z_A z_B + z_B^2 = 3,48 \quad (39)$$

The ellipse is shown, together with the points representing the  $z$ -scores, in Figure 10, together with the ellipses for probability levels of  $100\alpha\% = 1\%$  and  $0,1\%$ . The combined scores are shown in Table 10.

Inspection of Figure 10 reveals two laboratories (numbers 5 and 23) in the top right-hand quadrant. They have combined  $z$ -scores of 1,641 and 2,099. Laboratory 26 has a high  $z$ -score on Material B and a combined  $z$ -score of 2,059. After laboratories 5, 23 and 26, the laboratory that gives the next highest combined  $z$ -score is number 8 (combined score of 1,501).

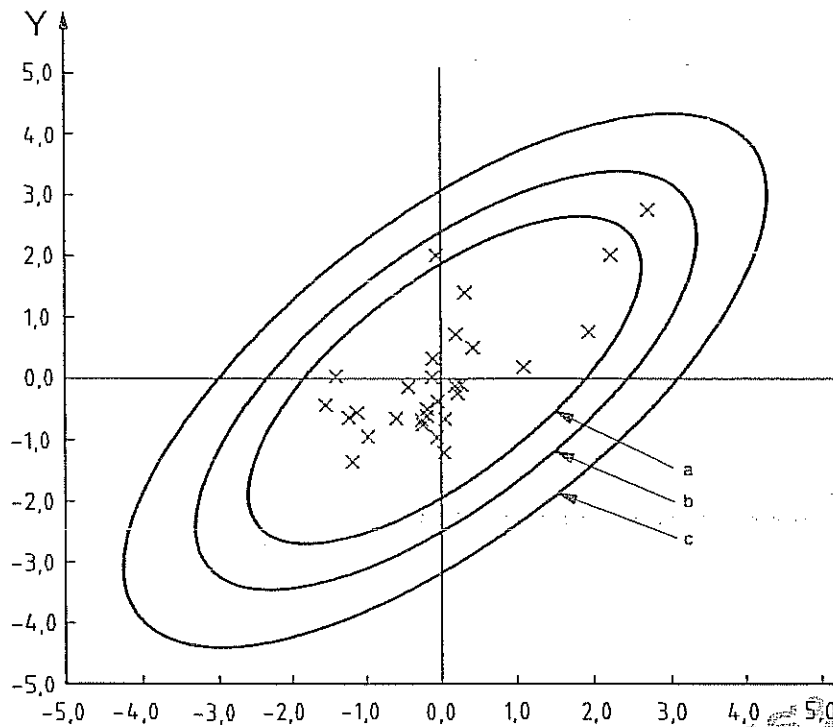
The points for laboratories 23 and 26 fall between the ellipses for the 5 % and 1 % probability levels, so it would be appropriate to treat their results as giving rise to "warning" signals, and to check where their results fall in the next round of the scheme.

Table 10 — Data and calculations on concentrations of antibodies for two similar allergens

Laboratory <i>i</i>	Data		z-score		Combined score $z_{A,B,i}$
	Allergen A $x_{A,i}$	Allergen B $x_{B,i}$	Allergen A $z_{A,i}$	Allergen B $z_{B,i}$	
1	12,95	9,15	0,427	0,515	0,370
2	6,47	6,42	-1,540	-0,428	1,275
3	11,40	6,60	-0,043	-0,366	0,336
4	8,32	4,93	-0,978	-0,942	0,737
5	18,88	13,52	2,228	2,023	1,641
6	15,14	8,22	1,092	0,194	0,965
7	10,12	7,26	-0,432	-0,138	0,349
8	17,94	9,89	1,942	0,770	1,501
9	11,68	4,17	0,042	-1,204	1,234
10	12,44	7,39	0,272	0,093	0,344
11	6,93	7,78	-1,400	0,042	1,430
12	9,57	5,80	-0,599	-0,642	0,477
13	11,73	5,77	0,057	-0,652	0,693
14	12,29	6,97	0,227	-0,238	0,429
15	10,95	6,23	-0,180	-0,493	0,388
16	10,95	5,90	-0,180	-0,607	0,497
17	11,17	7,74	-0,113	0,028	0,134
18	11,20	8,63	-0,104	0,335	0,415
19	7,64	3,74	-1,185	-1,353	0,986
20	12,17	7,33	0,190	-0,114	0,282
21	10,71	5,70	-0,253	-0,676	0,529
22	7,84	6,07	-1,124	-0,549	0,833
23	20,47	15,66	2,710	2,762	2,099
24	12,60	11,76	0,321	1,415	1,210
25	11,37	4,91	-0,052	-0,949	0,913
26	11,36	13,51	-0,055	2,019	2,059
27	10,75	5,48	-0,241	-0,752	0,607
28	12,21	9,77	0,203	0,729	0,603
29	7,49	5,82	-1,230	-0,635	0,902
Average	11,54	7,66	0,00	0,00	
Standard deviation	3,29	2,90	1,00	1,00	
Correlation coefficient	0,706		0,706		

NOTE 1 The data are numbers of units (U) in thousands (k) per litre (l) of sample, where a unit is defined by the concentration of an international reference material.

NOTE 2 The z-scores in this table have been calculated using non-rounded values of the averages and standard deviations, not using the rounded values shown at the bottom of the table.



**Key**

X z-score for allergen A  
 Y z-score for allergen B

- a 5 % level.
- b 1 % level.
- c 0,1 % level.

**Figure 10 — Youden Plot of z-scores from Table 10**

**8.5.3 Rank correlation test**

**8.5.3.1 General**

The rank correlation test is carried out using Spearman's correlation coefficient as follows. Replace the observed results for Material A by their ranks (i.e. replace the lowest value by 1, the next lowest by 2, and so on, until the highest is replaced by  $p$ ). Treat the observed results for Material B in the same way. Ties in the results are replaced by the average value of the ranks for the set of values that are tied. Calculate the correlation coefficient between the two series of ranks and compare the result with the tabulated value given in Table 11. If the calculated value is greater than the tabulated value, the relationship between the two variables is significant. The rank correlation coefficient may be calculated in the following simplified manner. Let

$$k_{A,1}, k_{A,2}, \dots, k_{A,p}$$

represent the ranks of the laboratories for Material A and

$$k_{B,1}, k_{B,2}, \dots, k_{B,p}$$

those for Material B. The rank correlation coefficient may be calculated as:

$$\rho_k = 1 - 6 \sum (k_{A,i} - k_{B,i})^2 / \{p(p^2 - 1)\} \tag{40}$$

where the summation is over the  $p$  laboratories.

Table 11 — Critical values for the rank correlation coefficient

Number of data points	Critical values	
	5 % level	1 % level
8	0,738	0,881
9	0,683	0,833
10	0,648	0,794
11	0,623	0,818
12	0,591	0,780
13	0,566	0,745
14	0,545	0,716
15	0,525	0,689
16	0,507	0,666
17	0,490	0,645
18	0,476	0,625
19	0,462	0,608
20	0,450	0,591
21	0,438	0,576
22	0,428	0,562
23	0,418	0,549
24	0,409	0,537
25	0,400	0,526
26	0,392	0,515
27	0,385	0,505
28	0,377	0,496
29	0,370	0,487
30	0,364	0,478

### 8.5.3.2 Example: Antibody concentrations

An example of the calculation of the rank correlation coefficient is given in Table 12. The calculated rank correlation coefficient is 0,605 which exceeds the tabulated value in Table 11 of 0,487 for the 1 % level for 29 data points, so it may be concluded that the relationship apparent in Figure 10 is statistically significant.

Table 12 — Calculation of the rank correlation coefficient for the data from Table 10

Laboratory <i>i</i>	Data		Rank		Difference $ k_{A,i} - k_{B,i} $
	Allergen A $z_{A,i}$	Allergen B $z_{B,i}$	Allergen A $k_{A,i}$	Allergen B $k_{B,i}$	
1	12,95	9,15	25	23	2
2	6,47	6,42	1	13	12
3	11,40	6,60	17	14	3
4	8,32	4,93	6	4	2
5	18,88	13,52	28	28	0
6	15,14	8,22	26	21	5
7	10,12	7,26	8	16	8
8	17,94	9,89	27	25	2
9	11,68	4,17	18	2	16
10	12,44	7,39	23	18	5
11	6,93	7,78	2	20	18
12	9,57	5,80	7	8	1
13	11,73	5,77	19	7	12
14	12,29	6,97	22	15	7
15	10,95	6,23	11,5	12	0,5
16	10,95	5,90	11,5	10	1,5
17	11,17	7,74	13	19	6
18	11,20	8,63	14	22	8
19	7,64	3,74	4	1	3
20	12,17	7,33	20	17	3
21	10,71	5,70	9	6	3
22	7,84	6,07	5	11	6
23	20,47	15,66	29	29	0
24	12,60	11,76	24	26	2
25	11,37	4,91	16	3	13
26	11,36	13,51	15	27	12
27	10,75	5,48	10	5	5
28	12,21	9,77	21	24	3
29	7,49	5,82	3	9	6
Sum of squared differences					1 605,50
$p(p^2-1)$					24 360
Calculated rank correlation coefficient					0,605
Tabulated critical value for the 1 % significance level					0,487
NOTE The data are numbers of units (U) in thousands (k) per litre (l) of sample, where a unit is defined by the concentration of an international reference material.					



## 8.6 Plots of repeatability standard deviations

### 8.6.1 General

When  $n$  replicate measurements are made by the participants in a round of a proficiency testing scheme, the results may be used to produce a plot to identify any laboratories whose average and standard deviation are unusual. The graph is constructed by plotting the within-laboratory standard deviation  $s_i$  for each laboratory against the corresponding average  $x_i$  for the laboratory. Let

$\bar{X} = x^*$  the robust average of  $x_1, x_2, \dots, x_p$ , as calculated by Algorithm A

$\bar{S} = s^*$  the robust pooled value of  $s_1, s_2, \dots, s_p$ , as calculated by Algorithm S

and assume that the data are Normally distributed. Under the null hypothesis that there is no difference between laboratories in the population values of either the laboratory means or the within-laboratory standard deviations, the statistic

$$\left( \sqrt{n} \frac{x_i - \bar{X}}{\bar{S}} \right)^2 + \left[ \sqrt{2(n-1)} \ln \left( \frac{s_i}{\bar{S}} \right) \right]^2$$

has approximately the  $\chi^2$  distribution with 2 degrees of freedom. Hence a critical region with a significance level of 1 % may be drawn on the graph by plotting

$$s = \bar{S} \exp \left[ \pm \frac{1}{\sqrt{2(n-1)}} \sqrt{\chi_{2,0.99}^2 - \left( \sqrt{n} \frac{x - \bar{X}}{\bar{S}} \right)^2} \right] \quad (41)$$

on the standard deviation axis against  $x$  on the average axis for

$$x = \bar{X} - \bar{S} \sqrt{\frac{\chi_{2,0.99}^2}{n}} \quad \text{to} \quad \bar{X} + \bar{S} \sqrt{\frac{\chi_{2,0.99}^2}{n}} \quad (42)$$

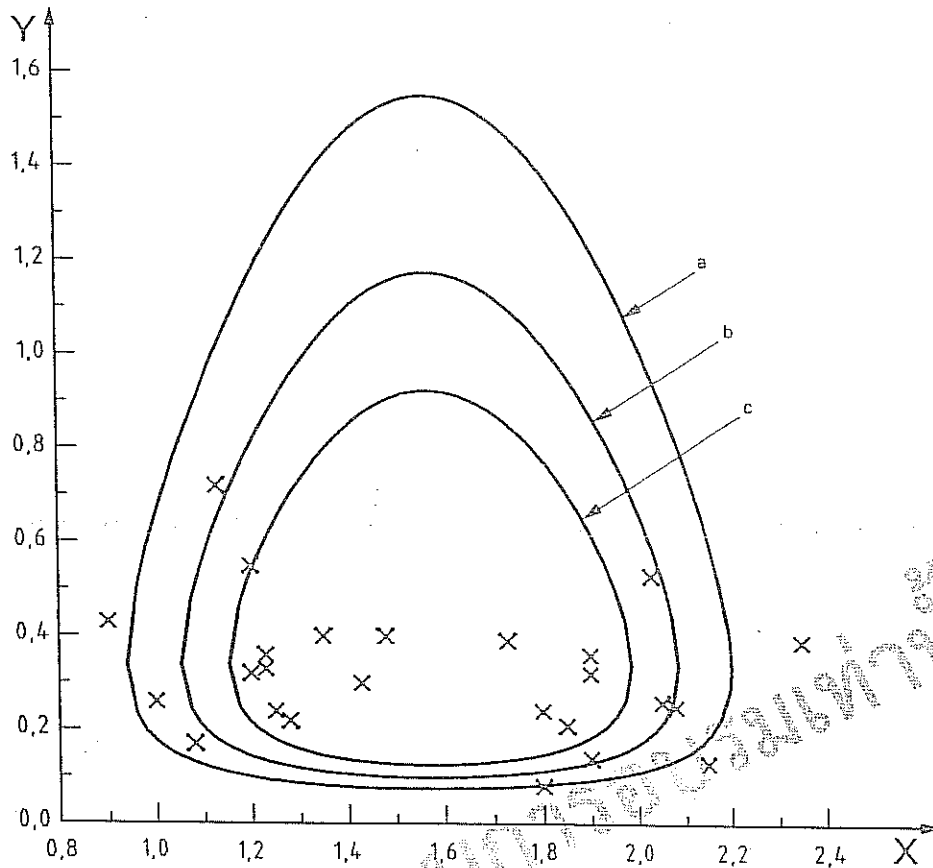
NOTE The Circle Technique was introduced by van Nuland<sup>[6]</sup>. The method described used a simple Normal approximation for the distribution of the standard deviation that could give a critical region containing negative standard deviations. The method given here uses an approximation for the distribution of the standard deviation that avoids this problem, but the critical region is no longer a circle as in the original. Further, robust values are used for the central point in place of simple averages as in the original method.

### 8.6.2 Example: Antibody concentrations

Table 13 shows the results of determining concentrations of a certain antibody in serum samples. Each laboratory made four replicate determinations, under repeatability conditions. The formulae given above are used to obtain the plot shown as Figure 11. The plot shows that several of the laboratories receive action or warning signals.

Table 13 — Concentrations of certain antibodies in serum samples  
(four replicate determinations on one sample in each laboratory)

Laboratory	Average kU/l	Standard deviation kU/l
1	2,15	0,13
2	1,85	0,21
3	1,80	0,08
4	1,80	0,24
5	1,90	0,36
6	1,90	0,32
7	1,90	0,14
8	2,05	0,26
9	2,35	0,39
10	2,03	0,53
11	2,08	0,25
12	1,25	0,24
13	1,13	0,72
14	1,00	0,26
15	1,08	0,17
16	1,20	0,32
17	1,35	0,4
18	1,23	0,36
19	1,23	0,33
20	0,90	0,43
21	1,48	0,40
22	1,20	0,55
23	1,73	0,39
24	1,43	0,30
25	1,28	0,22
Robust average	1,57	
Robust standard deviation		0,34
NOTE The data are numbers of units (U) in thousands (k) per litre (l) of sample, where a unit is defined by the concentration of an international reference material.		



#### Key

- X average  
Y standard deviation

NOTE The data are numbers of units (U) in thousands (k) per litre (l) of sample, where a unit is defined by the concentration of an international reference material.

- a 0,1 % level.  
b 1 % level.  
c 5 % level.

Figure 11 — Plot of standard deviations against averages for 25 laboratories  
(data from Table 13)

### 8.7 Split samples (see ISO/IEC Guide 43-1:1997, A.3.1.2)

#### 8.7.1 General

Split samples are used when it is necessary to carry out a detailed comparison of two laboratories. For example, if one laboratory is operated by a supplier, and the other by a customer, and the two organizations wish to ensure that the two laboratories are in agreement. Samples of several materials are obtained, representing a wide range of the property of interest, each sample is split into two parts, and each laboratory obtains some number (at least two) of replicate determinations on part of each sample.

On occasion, more than two laboratories may be involved, in which case one should be treated as a reference laboratory, and the others should be compared with it using the techniques described here.

The data from a split-sample experiment shall be used to produce graphs that display the variation between replicate measurements for the two laboratories and the differences between their average results for each sample. Further analysis will be dependent on deductions made from these graphs.

### 8.7.2 Example: Antibody concentrations

The concentration of certain antibodies in 21 serum samples were measured with radioimmunoassay methods in two laboratories denoted X and Y. In each laboratory all measurements were performed in duplicate in the same run. The concentrations obtained (in U/l) are presented in Table 14. As the measurement range is large, relative differences are relevant, so the data are transformed by taking logarithms to base e before the calculations are performed. The transformed data are shown in Table 15, and graphs showing the statistics from Table 15 are shown in Figures 12, 13 and 14.

From the graphs of ranges of replicate determinations, it appears that the variation between replicates for laboratory X is higher than for laboratory Y. Pooled values of these statistics are shown in Table 15, and could be compared using an  $F$ -test if it was of interest. Looking at the third graph, it can be seen that there is no obvious pattern or trend in the points. However, whereas the ranges of replicate determinations in Figures 12 and 13 are nearly all less than 0,2, many of the differences between laboratories in Figure 14 are much larger than this. This aspect requires investigation because it implies that the difference between the laboratories depends on the sample. The average difference between the laboratories may be calculated and is shown in Table 15. It may be used to give an indication of the importance of the difference between the laboratories, but it may not be used to predict the difference between the laboratories that might be obtained when analysing some subsequent sample. Thus with the transformed data, on average  $\ln(Y) - \ln(X) = 0,443$ , so  $Y/X = 1,6$ , indicating that laboratory Y obtains results, on average, higher than laboratory X by a factor of 1,6. However, with some samples the difference is much larger, and on others laboratory X obtains the higher results.

สงวนลิขสิทธิ์

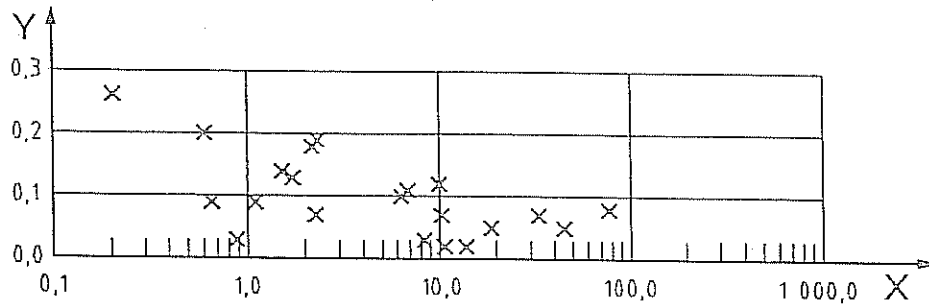
Table 14 — Concentrations of certain antibodies in 21 serum samples

Sample <i>i</i>	Laboratory X		Laboratory Y		Laboratory X	Laboratory Y	Laboratories X and Y
	Replicate 1 U/l	Replicate 2 U/l	Replicate 1 U/l	Replicate 2 U/l	Average U/l	Average U/l	Average U/l
1	19,106	18,174	11,473	11,705	18,640	11,589	15,115
2	6,424	7,171	5,812	5,812	6,798	5,812	6,305
3	6,619	5,989	11,705	11,473	6,304	11,589	8,947
4	0,543	0,664	0,861	0,905	0,604	0,863	0,743
5	43,816	46,063	49,899	55,147	44,940	52,523	48,731
6	2,096	2,535	24,047	26,843	2,316	25,445	13,880
7	10,591	9,875	9,116	8,671	10,233	8,894	9,563
8	13,874	13,599	12,554	12,807	13,737	12,681	13,209
9	1,974	2,363	1,094	1,020	2,169	1,057	1,613
10	9,393	10,591	13,736	14,585	9,992	14,161	12,076
11	1,840	1,616	2,484	2,460	1,728	2,472	2,100
12	31,817	34,124	48,424	55,147	32,971	51,786	42,378
13	1,150	1,051	2,014	2,270	1,101	2,142	1,621
14	0,625	0,684	1,051	1,174	0,655	1,113	0,884
15	73,700	79,838	119,104	127,740	76,769	123,422	100,096
16	2,181	2,340	2,560	3,065	2,261	2,813	2,537
17	8,415	8,166	5,755	5,585	8,291	5,670	6,980
18	1,419	1,632	8,846	8,846	1,526	8,846	5,186
19	0,861	0,887	2,612	3,065	0,874	2,839	1,856
20	10,697	10,486	15,029	14,880	10,592	14,955	12,773
21	0,230	0,177	0,795	0,795	0,204	0,795	0,499

NOTE The data are numbers of units (U) per litre (l) of sample, where a unit is defined by the concentration of an international reference material.

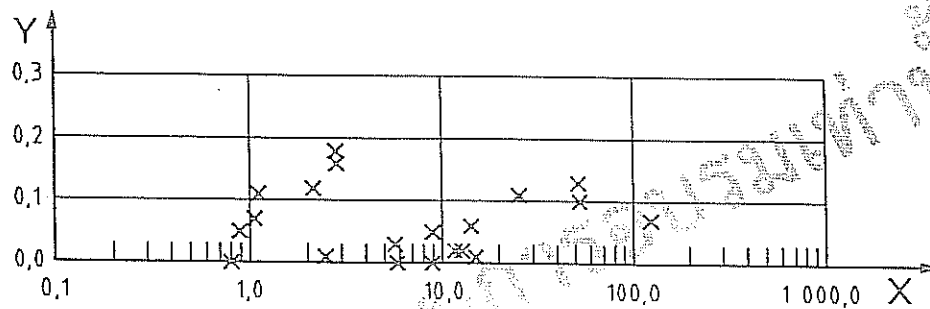
Table 15 — In(concentrations) and statistics for the data in Table 14

Sample <i>i</i>	Laboratory X		Laboratory Y		Laboratory X	Laboratory Y	Difference Y – X  In U/l
	Replicate 1 In U/l	Replicate 2 In U/l	Replicate 1 In U/l	Replicate 2 In U/l	Range In U/l	Range In U/l	
1	2,95	2,90	2,44	2,46	0,05	0,02	-0,475
2	1,86	1,97	1,76	1,76	0,11	0,00	-0,155
3	1,89	1,79	2,46	2,44	0,10	0,02	0,610
4	-0,61	-0,41	-0,15	-0,10	0,20	0,05	0,385
5	3,78	3,83	3,91	4,01	0,05	0,10	0,155
6	0,74	0,93	3,18	3,29	0,19	0,11	2,400
7	2,36	2,29	2,21	2,16	0,07	0,05	-0,140
8	2,63	2,61	2,53	2,55	0,02	0,02	-0,080
9	0,68	0,86	0,09	0,02	0,18	0,07	-0,715
10	2,24	2,36	2,62	2,68	0,12	0,06	0,350
11	0,61	0,48	0,91	0,90	0,13	0,01	0,360
12	3,46	3,53	3,88	4,01	0,07	0,13	0,450
13	0,14	0,05	0,70	0,82	0,09	0,12	0,665
14	-0,47	-0,38	0,05	0,16	0,09	0,11	0,530
15	4,30	4,38	4,78	4,85	0,08	0,07	0,475
16	0,78	0,85	0,94	1,12	0,07	0,18	0,215
17	2,13	2,10	1,75	1,72	0,03	0,03	-0,380
18	0,35	0,49	2,18	2,18	0,14	0,00	1,760
19	-0,15	-0,12	0,96	1,12	0,03	0,16	1,175
20	2,37	2,35	2,71	2,70	0,02	0,01	0,345
21	-1,47	-1,73	-0,23	-0,23	0,26	0,00	1,371
Pooled range					0,119	0,083	
Average difference between the laboratories							0,443
NOTE The data are numbers of units (U) per litre (l) of sample, where a unit is defined by the concentration of an international reference material. The pooled range is calculated according to Algorithm S in Annex C.							



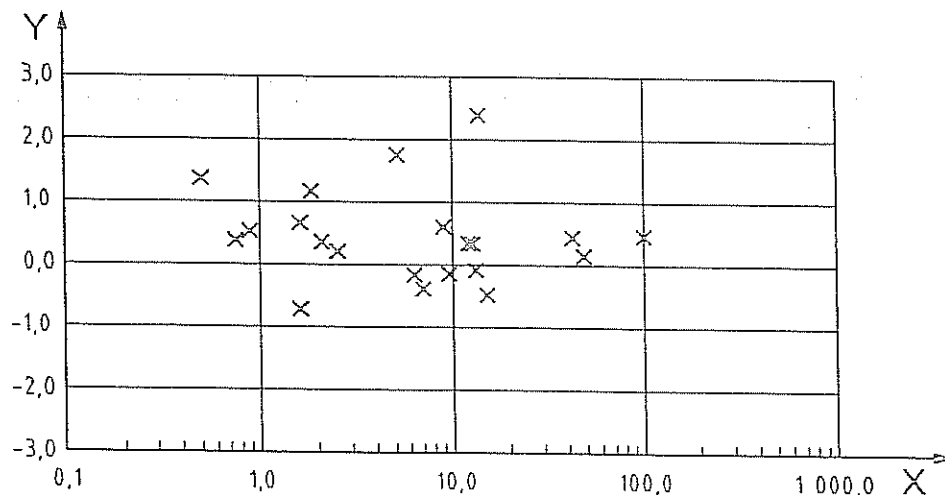
Key  
 X average concentration for laboratory X, %  
 Y range of ln(concentrations) for laboratory X

Figure 12 — Ranges of replicate determinations for laboratory X (with the ranges calculated from the ln concentrations)



Key  
 X average concentration for laboratory Y, %  
 Y range of ln(concentrations) for laboratory Y

Figure 13 — Ranges of replicate determinations for laboratory Y (with the ranges calculated from the ln concentrations)



Key  
 X average concentration for laboratories X and Y, %  
 Y difference in ln(concentrations) laboratory Y - laboratory X

Figure 14 — Differences between laboratory averages Y - X (with the differences calculated from the ln concentrations)

## 9 Graphical methods for combining performance scores over several rounds of a proficiency testing scheme (see ISO/IEC Guide 43-1:1997, A.3.2)

### 9.1 Applications

When standardized scores are to be combined over several rounds, the coordinator shall consider preparing graphs, as described in 9.2 or 9.3. The use of these graphs, in which the scores for several rounds are combined, may allow trends, and other features of the results, to be identified that are not apparent when scores for each round are examined separately.

**NOTE** The use of "running scores", in which the scores obtained by a laboratory are combined over several rounds but not displayed graphically, is not recommended. The laboratory may have a fault that shows up with the test material used in one round but not in the others. A running score may hide this fault. The use of a running scores in the form of counts of the numbers of action and warning signals is described in 7.9.

### 9.2 Shewhart control chart for $z$ -scores

#### 9.2.1 General

To prepare this chart, the  $z$ -scores for a laboratory are plotted as individual points, with action and warning limits set at  $\pm 2,0$  and  $\pm 3,0$  in the style illustrated by Table 16 and Figure 15. When several characteristics are measured in each round, the  $z$ -scores for different characteristics may be plotted on the same graph, but the points for the different characteristics should be plotted using different plotting symbols and/or different colours. See ISO 8258 [4] for advice on plotting Shewhart charts.

The Shewhart control chart is an effective method of identifying problems that cause large erratic values of  $z$ -scores.

The rules for interpreting the Shewhart control chart are that an out-of-control signal is given when

- a) a single point falls outside the action limits ( $\pm 3,0$ ),
- b) two out of three successive points outside the same warning limit ( $\pm 2,0$ ).

When such a Shewhart control chart gives an out-of-control signal, the actions set out in 4.1 shall be initiated. Note that the standard deviation for proficiency assessment  $\hat{\sigma}$  is not necessarily the standard deviation of the laboratory biases  $x - X$ , so the probability levels that are usually associated with the action and warning limits of a Shewhart control chart may not apply.

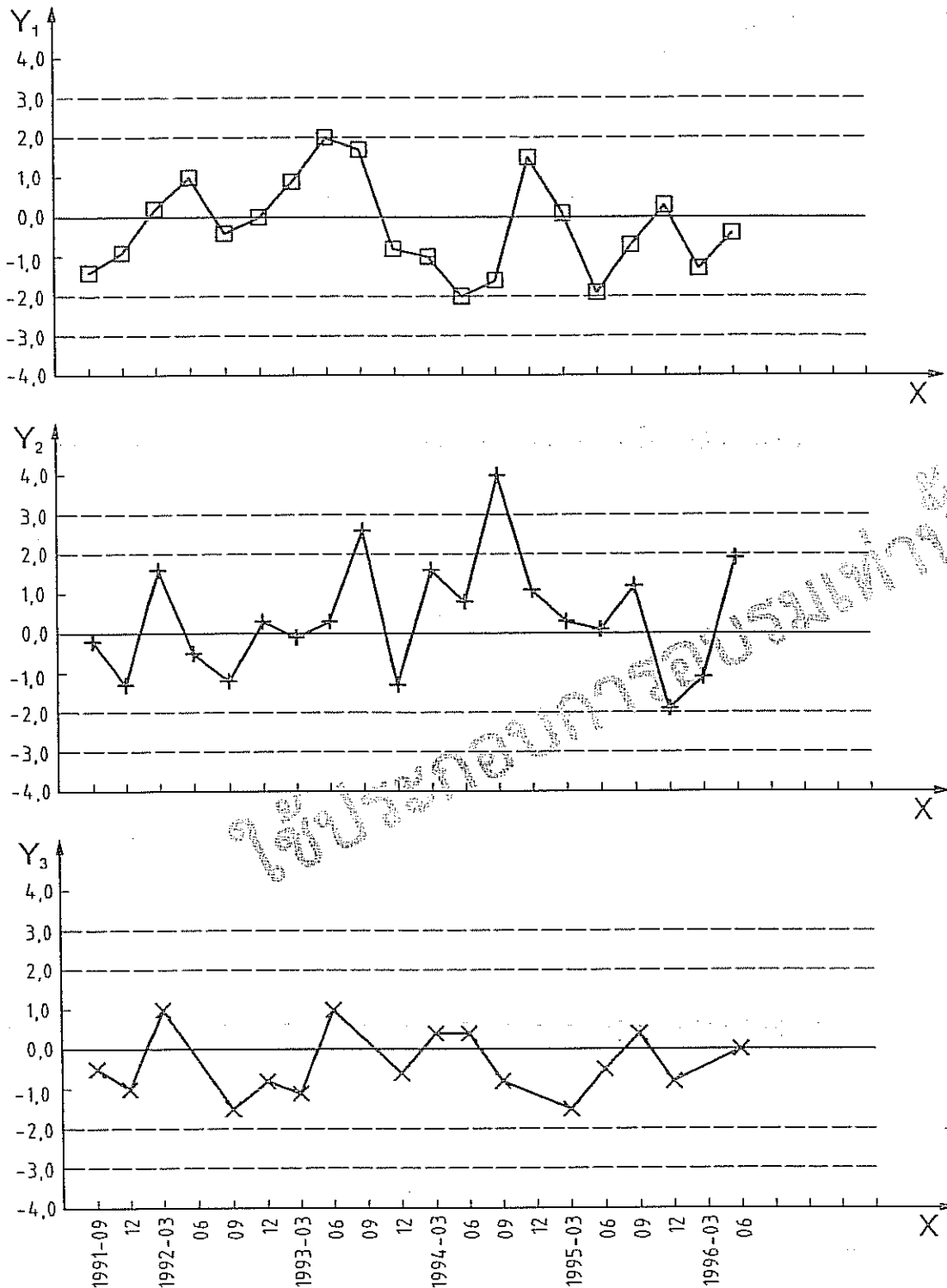
#### 9.2.2 Example: Allergen concentrations

The data are shown in Table 16, and are plotted in Figure 15.



Table 16 — Data for 20 rounds of a proficiency testing scheme involving the determination of the concentration of three allergens (d1, f1 and e3)

Date of round	z-score for allergen			Cusum of z-score		
	d1	f1	e3	d1	f1	e3
1991 September	-1,4	-0,2	-0,5	-1,4	-0,2	-0,5
1991 December	-0,9	-1,3	-1,0	-2,3	-1,5	-1,5
1992 March	0,2	1,6	1,0	-2,1	0,1	-0,5
1992 June	1,0	-0,5		-1,1	-0,4	
1992 September	-0,4	-1,2	-1,5	-1,5	-1,6	-2,0
1992 December	0,0	0,3	-0,8	-1,5	-1,3	-2,8
1993 March	0,9	-0,1	-1,1	-0,6	-1,4	-3,9
1993 June	2,0	0,3	1,0	1,4	-1,1	-2,9
1993 September	1,7	2,6		3,1	1,5	
1993 December	-0,8	-1,3	-0,6	2,3	0,2	-3,5
1994 March	-1,0	1,6	0,4	1,3	1,8	-3,1
1994 June	-2,0	0,8	0,4	-0,7	2,6	-2,7
1994 September	-1,6	4,0	-0,8	-2,3	6,6	-3,5
1994 December	1,5	1,1		-0,8	7,7	
1995 March	0,1	0,3	-1,5	-0,7	8,0	-5,0
1995 June	-1,9	0,1	-0,5	-2,6	8,1	-5,5
1995 September	-0,7	-1,2	0,4	-3,3	9,3	-5,1
1995 December	0,3	-1,9	-0,8	-3,0	7,4	-5,9
1996 March	-1,3	-1,1		-4,3	6,3	
1996 June	-0,4	1,9	0,0	-4,7	8,2	-5,9



Key

- X date of round
- Y<sub>1</sub> z-score for allergen d1
- Y<sub>2</sub> z-score for allergen f1
- Y<sub>3</sub> z-score for allergen e3

Figure 15 — Shewhart control charts for the data from one laboratory for 20 rounds of a proficiency testing scheme (data from Table 16)

### 9.3 Cusum control chart for $z$ -scores

#### 9.3.1 General

To prepare this chart, the cumulative sum (cusum) of the  $z$ -scores achieved by a laboratory over several rounds of a proficiency test is calculated and plotted as illustrated in Figure 16. Because the aim of the cusum chart is to detect cases where the  $z$ -scores deviate from zero, the "target value" for the cusum is zero, and the cusum may be calculated simply as the cumulative sum of the  $z$ -scores. See ISO/TR 7871 [5] for advice on the plotting of cusum charts.

When several characteristics are measured in each round, a separate cumulative sum should be calculated for each characteristic.

The cusum control chart is an effective method of identifying problems that cause a bias in the determination of a characteristic that persists over several rounds. When such a cusum control chart gives an out-of-control signal, the actions set out in 4.1 shall be initiated.

#### 9.3.2 Example: Allergen concentrations

Table 16 contains the  $z$ -scores and cusums derived from them for determinations of three allergens. The cusums are shown in Figure 16. The graphs suggest that a change took place in the laboratory's results from about March 1994 onwards.

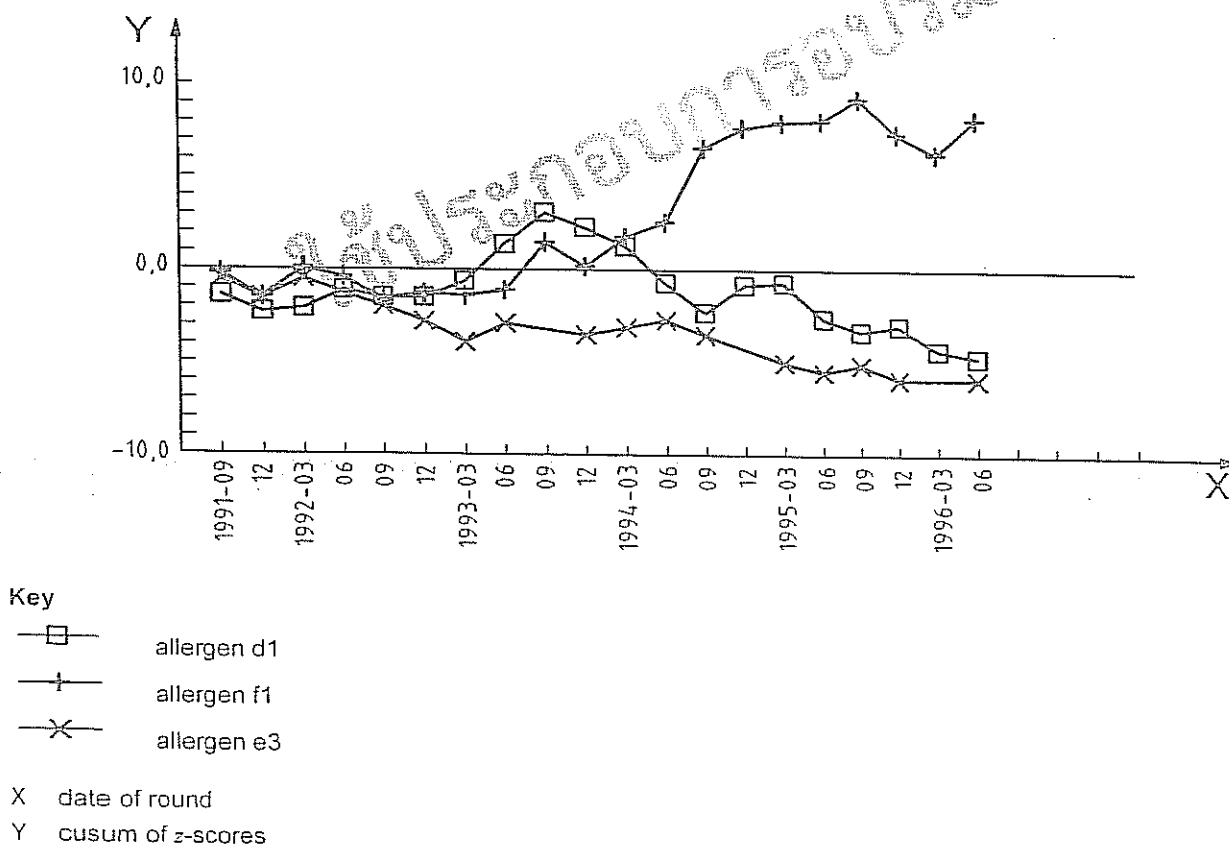


Figure 16 — Cusum control charts for the data from one laboratory for 20 rounds of a proficiency testing scheme (data from Table 16)

## 9.4 Plots of standardized laboratory biases against laboratory averages

### 9.4.1 General

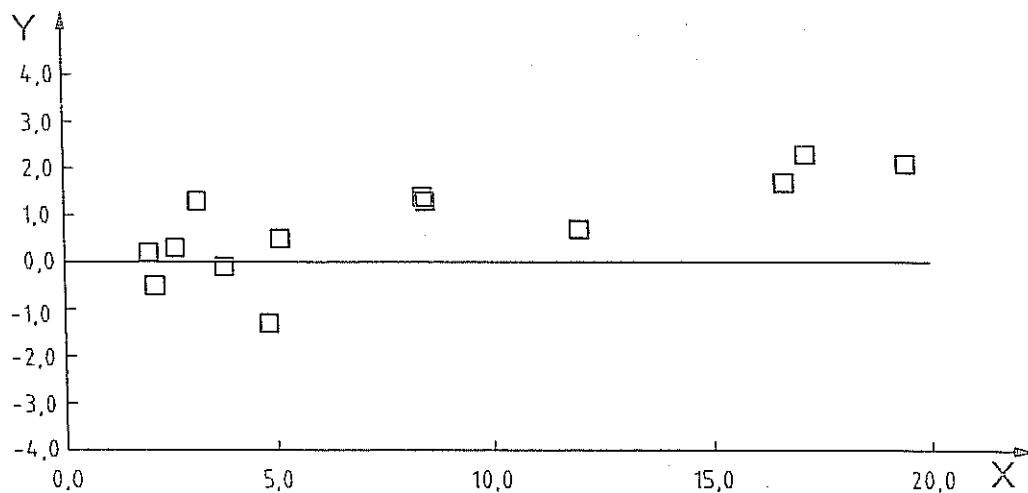
When the level of a characteristic varies over a wide range from one round of a proficiency test to another, plots of  $z$ -scores or other standardized scores against level will show if the laboratory bias changes with level.

### 9.4.2 Example: Allergen concentrations

Table 17 shows the determinations of the concentrations of an allergen reported by one laboratory over 13 rounds of a proficiency test, and Figure 17 shows the  $z$ -scores plotted against the level of concentration. The graph suggests that the laboratory tends to obtain positive  $z$ -scores at higher levels of concentration. With these data, regression analysis may be used to demonstrate that the trend is statistically significant.

Table 17 — Determinations of the concentration of allergen d1 in one laboratory

Date	Average concentration kU/l	$z$ -score
1992 June	12,00	0,7
1992 September	4,82	1,3
1992 December	17,20	2,3
1993 March	8,53	1,3
1993 June	8,47	1,4
1993 September	2,57	0,3
1993 December	1,94	0,2
1994 March	3,78	-0,1
1994 June	2,07	-0,5
1994 September	19,50	2,1
1994 December	3,13	1,3
1995 March	16,70	1,7
1995 June	5,12	0,5
NOTE The data are numbers of units (U) in thousands (k) per litre (l) of sample, where a unit is defined by the concentration of an international reference material.		



#### Key

X concentration, kU/l

Y z-score for allergen d1

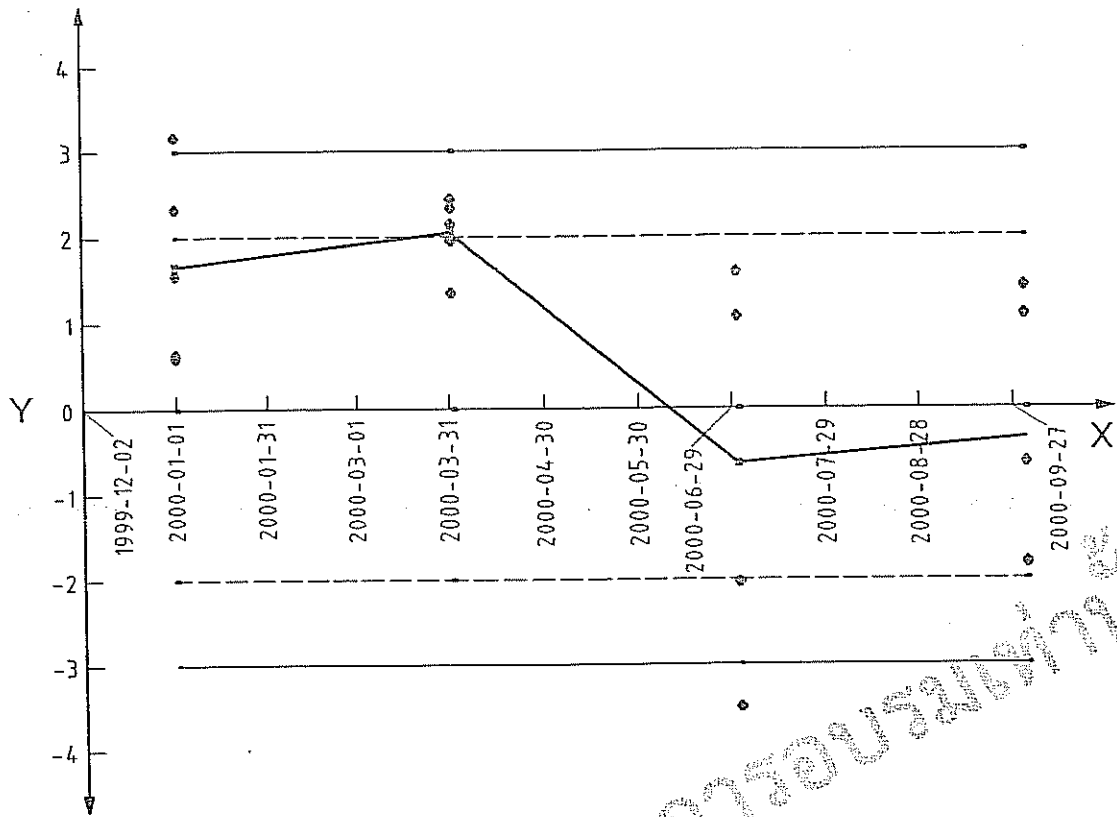
Figure 17 — Relation between z-score and level of concentration for determinations of allergen d1 in one laboratory

### 9.5 Dot plot

When more than one test item is tested for the same measurand in a round of a proficiency test, plots such as the Shewhart control chart can still be useful. A dot plot is similar to a Shewhart plot, but with multiple points at each time period, and where the lines connect the mean of the performance measures at each time period.

The following is an example of a situation where glucose is tested in five different samples every three months. The criterion for an acceptable interpretation is  $\pm 10\%$ , for assigned values of 60 mg/dl or higher, and 6,0 mg/dl for assigned values below 60 mg/dl. This situation and the broad range of values are suitable for evaluation with z-scores, using a fixed standard deviation for proficiency testing of 3,33 % or 2,0 mg/dl, depending on the assigned value.

The graph in Figure 18 shows that there were single results outside the  $\pm 3$  limits in the first and third rounds, and a few questionable values in the series. In this dataset, the laboratory had poor repeatability in the first round and a slightly high bias. In the second round, the precision problem was fixed but bias was not adjusted. The third round shows corrected bias, but with the original precision problem returned. The fourth round shows similar bias, but better precision.



Key  
 X test date  
 Y z-score

Figure 18 — Glucose z-scores in four proficiency testing events

## Annex A (normative)

### Symbols

$g$	Number of samples tested in a homogeneity check
$n$	Number of repeat measurements made per sample
$p$	Number of laboratories taking part in a round of a proficiency test
$s_s$	Between-samples standard deviation
$s_x$	Standard deviation of sample averages
$s^*$	Robust standard deviation (a standard deviation calculated by a robust algorithm)
$u_X$	Standard uncertainty of the assigned value
$w_t$	Between-test-portion range
$x$	Measurement result
$x^*$	Robust average (an average calculated by a robust algorithm)
$z$	Score used for proficiency assessment
$D$	Laboratory bias
$X$	Assigned value for proficiency assessment
$\sigma_L$	Between-laboratory standard deviation
$\hat{\sigma}$	Standard deviation for proficiency assessment
$\sigma_r$	Repeatability standard deviation
$\sigma_R$	Reproducibility standard deviation
$\phi$	Factor used to check the value of $\hat{\sigma}$ obtained by perception

## Annex B (normative)

### Homogeneity and stability checks of samples

#### B.1 Procedure for a homogeneity check

Follow the procedure given below.

- a) Choose a laboratory to carry out the homogeneity check. When it is acceptable not to carry out the homogeneity check on every measurand, choose a measurement method to use in the homogeneity check, and the test material characteristic to measure, which will be sensitive to heterogeneity between the samples. Thus, with a granular material, choose fully liberated characteristic (such as percentage passing a sieve of a certain aperture size). When the measurement is a proportion, a characteristic that is a small proportion can be more difficult to homogenize and so be more revealing in a homogeneity check. (A fully liberated characteristic is one such that the individual particles exhibit the extreme values of the characteristic. A characteristic is not fully liberated if individual particles possess the characteristic to varying degrees.)
- b) Prepare and package the samples for a round of the proficiency testing scheme, ensuring that there are sufficient samples for the participants in the scheme and for the homogeneity check.
- c) Select a number  $g$  of the samples in their final packaged form at random, where  $g \geq 10$ . The number of samples included in the homogeneity check may be reduced if suitable data are available from previous homogeneity checks on similar samples prepared by the same procedures.
- d) Prepare two test portions from each sample using techniques appropriate to the test material to minimize between-test-portion differences.
- e) Taking the  $2g$  test portions in a random order, obtain a measurement result on each, completing the whole series of measurements under repeatability conditions.
- f) Calculate the general average  $\bar{x}$ , within-samples standard deviation  $s_w$ , and between-samples standard deviation  $s_s$ , as shown in B.3.

#### B.2 Assessment criterion for a homogeneity check

Compare the between-samples standard deviation  $s_s$  with the standard deviation for proficiency assessment  $\hat{\sigma}$ . The samples may be considered to be adequately homogeneous if:

$$s_s \leq 0,3 \hat{\sigma} \quad (\text{B.1})$$

The justification for the factor of 0,3 is that when this criterion is met the between-samples standard deviation contributes no more than about 10 % of the standard deviation for proficiency testing. If this criterion is not met, the coordinator shall consider the following possibilities.

- a) Examine the sample preparation procedure to see if improvements are possible.
- b) Distribute a number of samples to each participant in the proficiency scheme, and require a measurement result to be obtained on each sample. The heterogeneity of the samples will increase the within-samples standard deviation to:

$$\sigma_{r1} = \sqrt{\sigma_r^2 + s_s^2} \quad (\text{B.2})$$



Use  $\sigma_{r,1}$  in place of  $\sigma_r$  in the guideline for choosing the number of replicate measurements, Equation (2).

- c) Include the between-samples standard deviation in the standard deviation for proficiency testing, by calculating  $\hat{\sigma}$  as

$$\hat{\sigma} = \sqrt{\hat{\sigma}_1^2 + s_s^2} \quad (\text{B.3})$$

where  $\hat{\sigma}_1$  is the standard deviation for proficiency testing that does not include any allowance for the heterogeneity of the samples.

### B.3 Formulae for homogeneity check

The data from a homogeneity check are represented by:

$x_{t,k}$

where

$t$  represents the sample ( $t = 1, 2, \dots, g$ )

$k$  represents the test portion ( $k = 1, 2$ )

Define the sample averages as:

$$x_{t..} = (x_{t,1} + x_{t,2}) / 2 \quad (\text{B.4})$$

and the between-test-portion ranges as:

$$w_t = |x_{t,1} - x_{t,2}| \quad (\text{B.5})$$

Calculate the general average:

$$\bar{x}_{..} = \sum \bar{x}_{t..} / g \quad (\text{B.6})$$

the standard deviation of sample averages:

$$s_x = \sqrt{\sum (x_{t..} - \bar{x}_{..})^2 / (g - 1)} \quad (\text{B.7})$$

and the within-samples standard deviation:

$$s_w = \sqrt{\sum w_t^2 / (2g)} \quad (\text{B.8})$$

where the summations are over samples ( $t = 1, 2, \dots, g$ ).

Finally, calculate the between-samples standard deviation as:

$$s_s = \sqrt{s_x^2 - (s_w^2 / 2)} \quad (\text{B.9})$$

NOTE Instead of using ranges, one could use between test portion standard deviations such as  $s_t = w_t / \sqrt{2}$ .

#### B.4 Procedure for a stability check

- a) Use the same laboratory for the stability check as for the homogeneity check. Use the same measurement method, and measure the same test material characteristic.
- b) Allow a time delay between the homogeneity tests and the stability tests similar to the time delay that will be experienced by the samples tested by the participants in the proficiency test.
- c) Select a number  $g$  of the samples at random, where  $g \geq 3$ .
- d) Prepare two test portions from each sample using the same techniques as for the homogeneity check.
- e) Taking the  $2g$  test portions in a random order, obtain a measurement result  $y_{i,k}$  on each, completing the whole series of measurements under repeatability conditions.
- f) Calculate the general average  $\bar{y}_{..}$  of the measurements obtained in the stability test.

#### B.5 Assessment criterion for a stability check

Compare the general average of the measurements obtained in the homogeneity check with the general average of the results obtained in the stability check. The samples may be considered to be adequately stable if:

$$|\bar{x}_{..} - \bar{y}_{..}| \leq 0,3\hat{\sigma} \quad (\text{B.10})$$

If this criterion is not met, examine the sample preparation and storage procedures to see if improvements are possible.

#### B.6 Example: Copper in soya flour ( $\mu\text{g/g}$ )

In a proficiency testing scheme in which soya flour is to be used as the test material the standard deviation for proficiency assessment has been set at  $\hat{\sigma} = 1,1 \mu\text{g/g}$  using the results of a precision experiment. According to the criterion (B.1), the between-samples standard deviation should be no more than

$$0,3 \times 1,1 = 0,330 \mu\text{g/g}$$

For a homogeneity check, 12 of the samples of soya flour prepared for the proficiency testing scheme were selected at random, and the copper contents of two test portions from each sample were determined. The data are shown in Table B.1, together with sample averages and between-test-portion ranges.

The formulae given in B.3 may be used to calculate the following values:

general average	$\bar{x}_{..} = 10,02 \mu\text{g/g}$
-----------------	--------------------------------------

standard deviation of sample averages	$s_x = 0,340 \mu\text{g/g}$
---------------------------------------	-----------------------------

within-samples standard deviation	$s_w = 0,246 \mu\text{g/g}$
-----------------------------------	-----------------------------

and hence the

between-samples standard deviation	$s_b = 0,292 \mu\text{g/g}$
------------------------------------	-----------------------------

This value is less than  $0,330 \mu\text{g/g}$ , so it may be concluded that the samples are adequately homogeneous for use in the proficiency testing scheme.

For the stability check, three of the samples were tested by the same laboratory one month later, and gave an average result of  $\bar{y}_{...} = 10,78 \mu\text{g/g}$ . The difference  $\bar{y}_{...} - \bar{x}_{...} = 0,76 \mu\text{g/g}$  exceeds  $0,3 \sigma = 0,33 \mu\text{g/g}$ , so it has to be concluded that the samples are not adequately stable.

Table B.1 — Measurement results for homogeneity check, copper in soya flour (mg/g)

Sample number <i>i</i>	Test portion 1	Test portion 2	Sample average $x_{i..}$	Between-test-portion range $w_i$
1	10,5	10,4	10,45	0,1
2	9,6	9,5	9,55	0,1
3	10,4	9,9	10,15	0,5
4	9,5	9,9	9,70	0,4
5	10,0	9,7	9,85	0,3
6	9,6	10,1	9,85	0,5
7	9,8	10,4	10,10	0,6
8	9,8	10,2	10,00	0,4
9	10,8	10,7	10,75	0,1
10	10,2	10,0	10,10	0,2
11	9,8	9,5	9,65	0,3
12	10,2	10,0	10,10	0,2

## Annex C (normative)

### Robust analysis

#### C.1 Robust analysis: Algorithm A

This algorithm yields robust values of the average and standard deviation of the data to which it is applied.

NOTE 1 Algorithms A and S given in this annex are reproduced from ISO 5725-5.

NOTE 2 Robustness is a property of the estimation algorithm, not of the estimates it produces, so it is not strictly correct to call the averages and standard deviations calculated by such an algorithm robust. However, to avoid the use of excessively cumbersome terminology, the terms "robust average" and "robust standard deviation" should be understood in this International Standard to mean estimates of the population mean or of the population standard deviation calculated using a robust algorithm.

Denote the  $p$  items of data, sorted into increasing order, by:

$$x_1, x_2, \dots, x_i, \dots, x_p$$

Denote the robust average and robust standard deviation of these data by  $x^*$  and  $s^*$ .

Calculate initial values for  $x^*$  and  $s^*$  as:

$$x^* = \text{median of } x_i \quad (i = 1, 2, \dots, p) \quad (\text{C.1})$$

$$s^* = 1,483 \text{ median of } |x_i - x^*| \quad (i = 1, 2, \dots, p) \quad (\text{C.2})$$

Update the values of  $x^*$  and  $s^*$  as follows. Calculate:

$$\delta = 1,5s^* \quad (\text{C.3})$$

For each  $x_i$  ( $i = 1, 2, \dots, p$ ), calculate:

$$x_i^* = \begin{cases} x^* - \delta, & \text{if } x_i < x^* - \delta \\ x^* + \delta, & \text{if } x_i > x^* + \delta \\ x_i, & \text{otherwise} \end{cases} \quad (\text{C.4})$$

Calculate the new values of  $x^*$  and  $s^*$  from:

$$x^* = \sum x_i^* / p \quad (\text{C.5})$$

$$s^* = 1,134 \sqrt{\sum (x_i^* - x^*)^2 / (p - 1)} \quad (\text{C.6})$$

where the summation is over  $i$ .

The robust estimates  $x^*$  and  $s^*$  may be derived by an iterative calculation, i.e. by updating the values of  $x^*$  and  $s^*$  several times using the modified data, until the process converges. Convergence may be assumed when there is no change from one iteration to the next in the third significant figure of the robust standard deviation and of the equivalent figure in the robust average. This is a simple method to program on a computer.

## C.2 Algorithm S

This algorithm is applied to standard deviations (or ranges). It yields a robust pooled value of the standard deviations or ranges to which it is applied.

Denote the  $p$  items of data, sorted into increasing order, by:

$$w_1, w_2, \dots, w_i, \dots, w_p$$

(These may be ranges or standard deviations.)

Denote the robust pooled value by  $w^*$ , and the degrees of freedom associated with each  $w_i$  by  $\nu$ . (When  $w_i$  is a range,  $\nu = 1$ . When  $w_i$  is the standard deviation of  $n$  test results,  $\nu = n - 1$ .) Obtain the values of  $\xi$  and  $\eta$  required by the algorithm from Table C.1.

Calculate an initial value for  $w^*$  as:

$$w^* = \text{median of } w_i \quad (i = 1, 2, \dots, p) \quad (\text{C.7})$$

Update the value of  $w^*$  as follows. Calculate:

$$\psi = \eta \times w^* \quad (\text{C.8})$$

For each  $w_i$  ( $i = 1, 2, \dots, p$ ), calculate:

$$w_i^* = \begin{cases} \psi, & \text{if } w_i > \psi \\ w_i, & \text{otherwise} \end{cases} \quad (\text{C.9})$$

Calculate the new value of  $w^*$  from:

$$w^* = \xi \sqrt{\sum (w_i^*)^2 / p} \quad (\text{C.10})$$

The robust estimate  $w^*$  may be derived by an iterative calculation by updating the value of  $w^*$  several times, until the process converges. Convergence may be assumed when there is no change from one iteration to the next in the third significant figure of the robust range. This is a simple method to program on a computer.

Table C.1 — Factors required for robust analysis: Algorithm S

Degrees of freedom $\nu$	Limit factor $\eta$	Adjustment factor $\xi$
1	1,645	1,097
2	1,517	1,054
3	1,444	1,039
4	1,395	1,032
5	1,359	1,027
6	1,332	1,024
7	1,310	1,021
8	1,292	1,019
9	1,277	1,018
10	1,264	1,017

NOTE The values of  $\xi$  and  $\eta$  are derived in Annex B of ISO 5725-5:1998.

## Bibliography

- [1] THOMPSON, M and WOOD, R. *The international harmonized protocol for the proficiency testing of (chemical) analytical laboratories*, AOAC, September 1993
- [2] JACKSON, J.E. Quality control methods for two related variables. *Industrial Quality Control*, 7, 1956, pp. 2-6
- [3] HORWITZ, W. Evaluation of analytical methods used for regulations of food and drugs. *Anal. Chem.*, 54, 1982, pp. 67A-76A
- [4] ISO 8258, *Shewhart control charts*
- [5] ISO/TR 7871, *Cumulative sum charts — Guidance on quality control and data analysis using CUSUM techniques*
- [6] VAN NULAND, Y. *ISO 9002 and the circle technique*, *Qual. Eng.*, 5, 1992, pp. 269-291
- [7] GUM:1993, *Guide to the expression of uncertainty in measurement* BIPM; IEC, IFCC; ISO; IUPAC; IUPAP, OIML<sup>2)</sup>
- [8] ISO 5725-2, *Accuracy (trueness and precision) of measurement methods and results — Part 2: Basic method for the determination of repeatability and reproducibility of a standard measurement method*
- [9] ISO 5725-3, *Accuracy (trueness and precision) of measurement methods and results — Part 3: Intermediate measures of the precision of a standard measurement method*
- [10] ISO 5725-4, *Accuracy (trueness and precision) of measurement methods and results — Part 4: Basic methods for the determination of the trueness of a standard measurement method*
- [11] ISO 5725-5, *Accuracy (trueness and precision) of measurement methods and results — Part 5: Alternative methods for the determination of the precision of a standard measurement method*
- [12] ISO 5725-6, *Accuracy (trueness and precision) of measurement methods and results — Part 6: Use in practice of accuracy values*
- [13] ISO/IEC 17025, *General requirements for the competence of testing and calibration laboratories*
- [14] ISO/IEC Guide 43-2:1997, *Proficiency testing by interlaboratory comparisons — Part 2: Selection and use of proficiency testing schemes by laboratory accreditation bodies*

---

2) Corrected and reprinted in 1995.



ลิขสิทธิ์ของสถาบันวิจัยและพัฒนา  
เทคโนโลยีการเกษตรและอาหาร



การอบรมเชิงปฏิบัติการ  
Statistical approach for PT  
(Testing and Calibration)

วันที่ ๒๘ -๓๐ พฤษภาคม ๒๕๕๕  
ณ ห้อง ๓๒๐ อาคารสถานศึกษาเคมีปฏิบัติ  
กรมวิทยาศาสตร์บริการ

จัดโดย สำนักบริหารและรับรองห้องปฏิบัติการ

## อบรมเชิงปฏิบัติการ

### Statistical approach for PT (Testing and Calibration)

วันที่ ๒๘ - ๓๐ พฤษภาคม ๒๕๕๕

ณ ห้อง ๓๒๐ อาคารสถานศึกษาเคมีปฏิบัติ กรมวิทยาศาสตร์บริการ

จัดโดย สำนักบริหารและรับรองห้องปฏิบัติการ

กรมวิทยาศาสตร์บริการ

#### ๑. หลักการและเหตุผล

ตาม พรบ. การมาตรฐานแห่งชาติ พ.ศ.๒๕๕๑ สำนักบริหารและรับรองห้องปฏิบัติการ กรมวิทยาศาสตร์บริการ เป็นหน่วยงานเดียวที่รับผิดชอบการรับรองความสามารถผู้จัดโปรแกรมการทดสอบความชำนาญห้องปฏิบัติการตามมาตรฐาน ISO/IEC 17043 : 2010 Conformity assessment - General requirements for proficiency testing ซึ่งเป็นมาตรฐานสากลว่าด้วยข้อกำหนดสำหรับการทดสอบความชำนาญห้องปฏิบัติการ ที่ห้องปฏิบัติการใช้เป็นเทคนิคหนึ่งในการประกันคุณภาพผลการทดสอบของห้องปฏิบัติการและใช้ในการเฝ้าระวังสมรรถนะในการดำเนินงานอย่างต่อเนื่อง สถิติเป็นเครื่องมือสำคัญที่ผู้จัดโปรแกรมการทดสอบความชำนาญได้นำมาใช้ในการออกแบบโปรแกรมฯ การสุ่มตัวอย่าง การทดสอบความเป็นเนื้อเดียวกัน การทดสอบความเสถียร การหาค่ากำหนดของตัวอย่าง PT การประมาณค่าความไม่แน่นอน การประเมินสมรรถนะของห้องปฏิบัติการ สำนักฯ จึงเห็นความจำเป็นในการจัดอบรมหลักสูตรการอบรมสถิติสำหรับการทดสอบความชำนาญห้องปฏิบัติการจึงได้เสนอแผนการดำเนินงาน (Work plan ๒๐๑๑-๒๐๑๒) ด้านการทดสอบความชำนาญห้องปฏิบัติการต่อผู้จัดชุดโครงการอบรม APLAC-PTB Training course "Enhanced Training for PT Experts" นี้ เพื่อส่งเสริมความรู้ความเข้าใจแก่เจ้าหน้าที่ ผู้ประเมินของสำนักฯ และเจ้าหน้าที่ของหน่วยงานผู้จัดโปรแกรมการทดสอบความชำนาญห้องปฏิบัติการ ใช้สถิติที่เกี่ยวข้องกับการทดสอบความชำนาญได้อย่างถูกต้องและเป็นไปในแนวทางเดียวกัน อันจะทำให้การรับรองผู้จัดโปรแกรมการทดสอบความชำนาญห้องปฏิบัติการ ดำเนินไปอย่างมีประสิทธิภาพ และเป็นไปตามมาตรฐานสากล

#### ๒. วัตถุประสงค์

-เพื่อส่งเสริมความรู้ความเข้าใจแก่เจ้าหน้าที่ ผู้ประเมินของสำนักฯ และเจ้าหน้าที่ของหน่วยงานผู้จัดโปรแกรมการทดสอบความชำนาญห้องปฏิบัติการ ใช้สถิติที่เกี่ยวข้องกับการทดสอบความชำนาญได้อย่างถูกต้องและเป็นไปในแนวทางเดียวกัน

-เพื่อให้การรับรองผู้จัดโปรแกรมการทดสอบความชำนาญห้องปฏิบัติการ ดำเนินไปอย่างมีประสิทธิภาพ และเป็นไปตามมาตรฐานสากล

๓. กลุ่มเป้าหมาย เจ้าหน้าที่ / ผู้ประเมินของสำนักฯ และเจ้าหน้าที่ของหน่วยงานผู้จัดโปรแกรมการทดสอบความชำนาญห้องปฏิบัติการ

๔. วิทยากร ๑) ผศ.ดร.ครรชิต จุดประสงค์ จากสถาบันโภชนาการ

๒) ดร.พจมาน ท่าจิ้น จาก บท.วศ.

๓) นางจันทรัตน์ วรสรรพวิทย์ จาก บร.วศ.

๔) นางสุดา นันทวิทยา จาก บร.วศ.

๕) นางภัทรภร ธนะภวาริศ จาก บร.วศ.

๖) นางวรรณิ อุไพบูรณ์ จาก บท.วศ.

๕. วิธีดำเนินการ บรรยายภาคทฤษฎีและฝึกปฏิบัติ

๖. ระยะเวลา ๓ วัน (๒๘-๓๐ พฤษภาคม ๒๕๕๕)

๗. จำนวนผู้เข้าร่วม ๕๕ คน

๘. สถานที่ ห้อง ๓๒๐ อาคารสถานศึกษาเคมีปฏิบัติ

๙. งบประมาณ โครงการพัฒนาการรับรองระบบงานห้องปฏิบัติการ

๑๐. ที่ปรึกษาโครงการ รองอธิบดีกรมวิทยาศาสตร์บริการ (นางสาวจันทร์เพ็ญ ใจธีรภาพกุล)

๑๑. ผู้รับผิดชอบโครงการ ผู้อำนวยการสำนักบริหารและรับรองห้องปฏิบัติการ ( นางครุณี วัชรเวียงวิทย์ )

๑๒. ผู้รับผิดชอบการจัด กลุ่มส่งเสริมพัฒนาห้องปฏิบัติการ และกลุ่มกำกับดูแลมาตรฐานห้องปฏิบัติการ

กำหนดการอบรมเชิงปฏิบัติการ  
Statistical approach for PT (Testing and Calibration)

วันที่ ๒๘ -๓๐ พฤษภาคม ๒๕๕๕

ณ ห้อง ๓๒๐ อาคารสถานศึกษาเคมีปฏิบัติ กรมวิทยาศาสตร์บริการ  
จัดโดย สำนักบริหารและรับรองห้องปฏิบัติการ กรมวิทยาศาสตร์บริการ

วันที่	หัวข้อเรื่อง	วิทยากร
๒๘ พฤษภาคม ๒๕๕๕		
๐๘.๓๐ น. - ๐๙.๐๐ น.	ลงทะเบียนและพิธีเปิดการอบรม (อวศ.)	
๐๙.๐๐ น. - ๑๑.๓๐ น.	การออกแบบและการสุ่ม PT sample	ดร.พจมาน ทำจิ้น
๑๑.๓๐ น. - ๑๒.๐๐ น.	แบบฝึกหัด-ฝึกคำนวณ	วิทยากรภาคปฏิบัติ
๑๒.๐๐ น. - ๑๓.๐๐ น.	พักรับประทานอาหารกลางวัน	
๑๓.๐๐ น. - ๑๕.๓๐ น.	การทดสอบความเป็นเนื้อเดียวกันของ PT sample	ผศ.ดร.ครรชิต จุดประสงค์/ นางจันทรัตน์ วรสรรพวิทย์
๑๕.๓๐ น. - ๑๖.๐๐ น.	แบบฝึกหัด-ฝึกคำนวณ	วิทยากรภาคปฏิบัติ
๒๙ พฤษภาคม ๒๕๕๕		
๐๘.๓๐ น. - ๐๙.๐๐ น.	ลงทะเบียน	
๐๙.๐๐ น. - ๑๐.๐๐ น.	การทดสอบความเสถียร ของ PT sample	ผศ.ดร.ครรชิต จุดประสงค์
๑๐.๐๐ น. - ๑๐.๓๐ น.	แบบฝึกหัด - ฝึกคำนวณ	วิทยากรภาคปฏิบัติ
๑๐.๓๐ น. - ๑๑.๓๐ น.	การหาค่ากำหนดของ PT sample การคำนวณค่า target standard deviation	ผศ.ดร.ครรชิต จุดประสงค์/ นางจันทรัตน์ วรสรรพวิทย์
๑๑.๓๐ น. - ๑๒.๐๐ น.	แบบฝึกหัด - ฝึกคำนวณ	วิทยากรภาคปฏิบัติ
๑๒.๐๐ น. - ๑๓.๐๐ น.	พักรับประทานอาหารกลางวัน	
๑๓.๐๐ น. - ๑๕.๓๐ น.	การคำนวณค่าความไม่แน่นอนของ PT sample ความสอกลับได้ของค่ากำหนด	ดร.พจมาน ทำจิ้น
๑๕.๓๐ น. - ๑๖.๐๐ น.	แบบฝึกหัด - ฝึกคำนวณ	วิทยากรภาคปฏิบัติ
๓๐ พฤษภาคม ๒๕๕๕		
๐๘.๓๐ น. - ๐๙.๐๐ น.	ลงทะเบียน	
๐๙.๐๐ น. - ๑๑.๐๐ น.	สรุปการออกแบบทางสถิติสำหรับการทดสอบความชำนาญ	ดร.พจมาน ทำจิ้น
๑๑.๐๐ น. - ๑๒.๐๐ น.	การประเมินสมรรถนะของห้องปฏิบัติการ	ผศ.ดร.ครรชิต จุดประสงค์/ นางจันทรัตน์ วรสรรพวิทย์
๑๒.๐๐ น. - ๑๓.๐๐ น.	พักรับประทานอาหารกลางวัน	
๑๓.๐๐ น. - ๑๕.๓๐ น.	การประเมินสมรรถนะของห้องปฏิบัติการ (ต่อ) การคำนวณค่า Z-score, En และอื่นๆ	ผศ.ดร.ครรชิต จุดประสงค์/ นางจันทรัตน์ วรสรรพวิทย์
๑๕.๓๐ น. - ๑๖.๐๐ น.	แบบฝึกหัด - ฝึกคำนวณ	วิทยากรภาคปฏิบัติ

- หมายเหตุ
- วิทยากรบรรยาย : ดร.พจมาน ทำจิ้น ผศ.ดร.ครรชิต จุดประสงค์ นางจันทรัตน์ วรสรรพวิทย์
  - วิทยากรภาคปฏิบัติ : ดร.พจมาน ทำจิ้น ผศ.ดร.ครรชิต จุดประสงค์ นางจันทรัตน์ วรสรรพวิทย์  
นางสุคนธ์ นันทวิทยา นางภัทรรณ ณะภาววิศ นางวรรณิ อุไพบูรณ์
  - พักรับประทานอาหารว่าง เวลา ๑๐.๓๐ น. - ๑๐.๔๕ น. และ ๑๔.๓๐ น. - ๑๔.๔๕ น.
  - กำหนดการอาจเปลี่ยนแปลงตามความเหมาะสม

## ประวัติวิทยากร “ผศ.ดร. ครรชิต จุดประสงค์”

ข้อมูลเพิ่มเติม: <http://www.inmu.mahidol.ac.th/th/about-us/faculties/kunchit.pdf>

### ตำแหน่งปัจจุบัน

ผู้ช่วยศาสตราจารย์ ณ สถาบันโภชนาการ มหาวิทยาลัยมหิดล

ผู้ช่วยผู้จัดการด้านคุณภาพของห้องปฏิบัติการสถาบันโภชนาการ ตาม ISO/IEC 17025

### วุฒิการศึกษา

วทม. อาหารและโภชนาการ (Food and Nutrition for Development)

Ph.D เคมีวิเคราะห์ (Analytical Chemistry)

### ประวัติการทำงาน

พ.ศ. 2531-2539 พนักงานวิทยาศาสตร์การแพทย์: วิเคราะห์คุณค่าทางโภชนาการ สถาบันโภชนาการ

พ.ศ. 2539-2550 นักวิจัย: วิจัยและพัฒนาทางด้านอาหารและโภชนาการ สถาบันโภชนาการ

พ.ศ. 2550-ปัจจุบัน อาจารย์: สอนในด้านอาหารและโภชนาการ เครื่องมือและวิธีวิเคราะห์ สถาบันโภชนาการ

### การฝึกอบรม ฐาน ประชุมและสัมมนาที่สำคัญ

#### International (ต่างประเทศ):

- Food Analysis Queensland Health Scientific Services Laboratory, Brisbane, Australia
- Production and Use of Food Composition Data in Nutrition, FoodComp'98 the Netherlands. (UNU Fellow) Interregional Training Course on Organizational โดย IAEA, Austria (IAEA fellow)
- Certificate of Preparation of Food Proficiency Testing and Reference Material โดย IAEA, Austria (IAEA fellow)
- Laboratory Quality Standards Towards Global Competitiveness: Management System, Reference Materials, and Proficiency Testing สถาบันโภชนาการ มหาวิทยาลัยมหิดล (AusAID)
- Laboratory Quality Standards Towards Global Competitiveness: Management System, Method Validation, Uncertainty of Measurement, Reference Materials, and Proficiency Testing, Food and Nutrition Research Institute (FNRI), Manila, Philippines (AusAID)
- Fructan (Inulin and Oligofructose) analysis Orafiti Analytical Department, Orafiti Active Food Ingredients, Belgium
- ISO 13528:2005 Statistical methods for use in proficiency testing interlaboratory comparisons, Bangkok, Thailand
- Assessment of micronutrients in nutrition by stable isotope mass spectrometry, Scottish Universities Environmental Research Centre (SUERC), University of Glasgow, Scotland, United Kingdom (IAEA fellow)

### รางวัล

- 5 April – 28 May, 2010 Recipient of the International Atomic Energy Agency Fellowship Awards: Training at University of Glasgow, Scotland, UK
- 2007 Graduate Thesis Publication Awards, Doctoral degree level, 2007
- 2006 Outstanding poster in the title of "Use of dialysis profiles from a continuous-flow dialysis system hyphenated with inductively coupled plasma optical emission", presented at 2006 Asia-Pacific Winter Conference on Plasma Spectrochemistry, Bangkok, Thailand.
- 2005 Outstanding poster in the title of "Continuous - Flow Dialysis System with Inductively Coupled Plasma Optical Emission Spectrometric Detector for Vitro Estimation of Bioavailability", presented at 2005 Asia-Pacific Winter Conference on Plasma Spectrochemistry, April 25-30, 2005, Chiang Mai, Thailand.
- 16 January – 15 April, 2006 Recipient of the International Atomic Energy Agency Fellowship Awards: Training at Seibersdorf Austria
- 8-24 March, 2004 Recipient of the International Atomic Energy Agency Fellowship Awards: Training at Seibersdorf, Austria
- 5-23 October, 1998 Recipient of the United Nations University Fellowship Awards: Training at Wageningen, the Netherlands
- 1998 Best Civil Officer, Institute of Nutrition, Mahidol University, Thailand.
- 1989 Best Civil Officer, Institute of Nutrition, Mahidol University, Thailand.

ผลงานตีพิมพ์ - ตีพิมพ์ในวารสารระดับนานาชาติ 16 เรื่อง  
- ตีพิมพ์ในวารสารระดับชาติ 7 เรื่อง  
- รายงานฉบับสมบูรณ์ 12 เล่ม

- นำเสนอผลงานวิชาการ แบบ Oral presentation 10 เรื่อง  
- นำเสนอผลงานวิชาการ แบบ Poster presentation 25 เรื่อง  
- หนังสือ 3 เล่ม

# ประวัติวิทยากร

ชื่อ - นามสกุล	ดร.พจมาน ท่าจีน
ตำแหน่งปัจจุบัน	นักวิทยาศาสตร์ชำนาญการพิเศษ
สถานที่ทำงาน	ศูนย์บริหารจัดการทดสอบความชำนาญ ห้องปฏิบัติการ กรมวิทยาศาสตร์บริการ
โทรศัพท์	0-2201-7331
โทรสาร	0-2201-7239
คุณวุฒิ	- วท.บ. (ฟิสิกส์) จุฬาลงกรณ์มหาวิทยาลัย - Ph.D. (Physics) Edinburgh University ประเทศอังกฤษ
ประสบการณ์	- Uncertainty of Measurement, UKAS ประเทศอังกฤษ - Advances of Industrial Metrology, Brunel University ร่วมกับ NSTDA - Assessor Training Course, สผอ. ร่วมกับ UKAS ประเทศอังกฤษ - ISO / IEC 17025-An Introduction to the Laboratory Accreditation Requirements, บร. วศ. ร่วมกับ NATA Laboratory Assessor / Lead Assessor Training Course, บร. วศ. ร่วมกับ NATA - Interlaboratory Comparisons : Their Use for Proficiency Testing, Uncertainty Statements, Traceability and More, NCSL ประเทศสหรัฐอเมริกา - Arrangements of Accreditation Body for The Recognition and Acceptance of Conformity Assessment Results, T & S International Training and Consultancy Partnership - Training in Metrology in Chemistry: Train MIC, European Commission- Joint Research Centre. Institute of Reference Materials and Measurements.
บทความทางวิชาการ	-ผู้แปลหนังสือ "มลภาวะทางเสียงในสิ่งแวดล้อม" จัดทำโดย บริษัท เมเทอโรโทรนิคส์ จำกัด และ Bruel & Kjaer, Denmark. -สุภาพร จาตุรันต์เรืองศรี พจมาน ท่าจีน สุริภณ สมนครพาณิชย์, "Primary Vibration Calibration of Accelerometer Using Laser Interferometry", LADKRABANG ENGINEERING JOURNAL, Vol.2 June 1999 -W.Chinchusak, V.Tipsuwanaporn and P.Tagheen, "A New Calibration Technique of Autocollimator by using Small Angle Generator and Laser Interferometer", KMITL SCIENCE Journal, Vol.1 No.1 Jan.-Dec.2001
ความรู้ความชำนาญพิเศษ	- คณะอนุกรรมการพิจารณารับรองห้องปฏิบัติการสอบเทียบทางกลและทางเคมี ในคณะกรรมการมาตรฐานผลิตภัณฑ์อุตสาหกรรม - คณะอนุกรรมการทดสอบความชำนาญ ในคณะกรรมการมาตรฐานผลิตภัณฑ์อุตสาหกรรม - อาจารย์พิเศษ หลักสูตรวิทยาศาสตร์มหาบัณฑิต สาขาวิชามาตรวิทยา คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

# ประวัติวิทยากร

## จันทร์รัตน์ วรสรรพวิทย์

ตำแหน่งปัจจุบัน : นักวิทยาศาสตร์ชำนาญการพิเศษ  
หัวหน้ากลุ่มทะเบียนและดัชนีความสามารถห้องปฏิบัติการ  
สำนักบริหารและรับรองห้องปฏิบัติการ  
กรมวิทยาศาสตร์บริการ กระทรวงวิทยาศาสตร์และเทคโนโลยี

### วุฒิทางการศึกษา :

- ❖ ประกาศนียบัตรวิชาชีพครู จากมหาวิทยาลัยสุโขทัยธรรมมาธิราช
- ❖ ปริญญาตรี วิทยาศาสตร์บัณฑิต สาขาเคมี เกียรตินิยมอันดับสอง มหาวิทยาลัยสงขลานครินทร์
- ❖ ปริญญาโท วิทยาศาสตร์มหาบัณฑิต สาขาสถิติ (เกียรตินิยม) สถาบันบัณฑิตพัฒนบริหารศาสตร์
- ❖ ได้รับรางวัลเหรียญดี สาขาสถิติ จากสถาบันบัณฑิตพัฒนบริหารศาสตร์

### ประสบการณ์/ความรู้ความชำนาญพิเศษ:

- ❖ เป็นคณะทำงานของ ILAC Proficiency Testing Consultative Group 6 (PTCG) for the revision of ISO 13528
- ❖ เป็นหัวหน้าผู้ประเมิน / ผู้ประเมินด้านวิชาการระบบบริหารงานคุณภาพห้องปฏิบัติการตาม ISO/IEC 17025 ของสำนักบริหารและรับรองห้องปฏิบัติการ กรมวิทยาศาสตร์บริการ
- ❖ เป็นผู้ประเมินด้านวิชาการ (สถิติ) ผู้จัดโปรแกรมการทดสอบความชำนาญห้องปฏิบัติการ ตาม ISO/IEC 17043 ของสำนักบริหารและรับรองห้องปฏิบัติการ กรมวิทยาศาสตร์บริการ
- ❖ เป็นวิทยากรในการอบรมหลักสูตรสถิติสำหรับการทดสอบความชำนาญห้องปฏิบัติการ
- ❖ เป็นวิทยากรในการอบรมหลักสูตรการคำนวณค่าสถิติ สำหรับงานวิเคราะห์ทดสอบ
- ❖ เป็นวิทยากรในการฝึกอบรมหลักสูตร สถิติสำหรับงานวิเคราะห์ทดสอบวิจัย
- ❖ เป็นวิทยากรในการฝึกอบรมหลักสูตรการประเมินผลการทดสอบทางสถิติในระบบคุณภาพ
- ❖ เป็นวิทยากรในการฝึกอบรมหลักสูตรการรับรองผู้ผลิตวัสดุอ้างอิงและการสอบกลับได้ของการวัด
- ❖ เป็นวิทยากรในการฝึกอบรมหลักสูตรความใช้ได้ของการวัด
- ❖ เป็นวิทยากรในการฝึกอบรมหลักสูตรการตรวจติดตามคุณภาพภายใน
- ❖ เป็นวิทยากรในการฝึกอบรมหลักสูตรนักวิเคราะห์หี้อาชีพด้านจุลชีววิทยา
- ❖ เป็นวิทยากรในการฝึกอบรมหลักสูตรการประมาณค่าความไม่แน่นอนของการวัดทางจุลชีววิทยา
- ❖ เป็นวิทยากรในการฝึกอบรมหลักสูตรการตรวจสอบความใช้ได้ของวิธีทางจุลชีววิทยา

สถานที่ติดต่อ : กลุ่มทะเบียนและดัชนีความสามารถห้องปฏิบัติการ  
สำนักบริหารและรับรองห้องปฏิบัติการ กรมวิทยาศาสตร์บริการ  
โทรศัพท์ 0-2201-7137, 0-2201-7165, 08-9123-3321  
โทรสาร 0-2201-7201 e-mail : chantaiat@dss.go.th

การออกแบบกิจกรรม  
การทดสอบความชำนาญ

วิทยากร: ดร.พจมาน ทำจิ้น

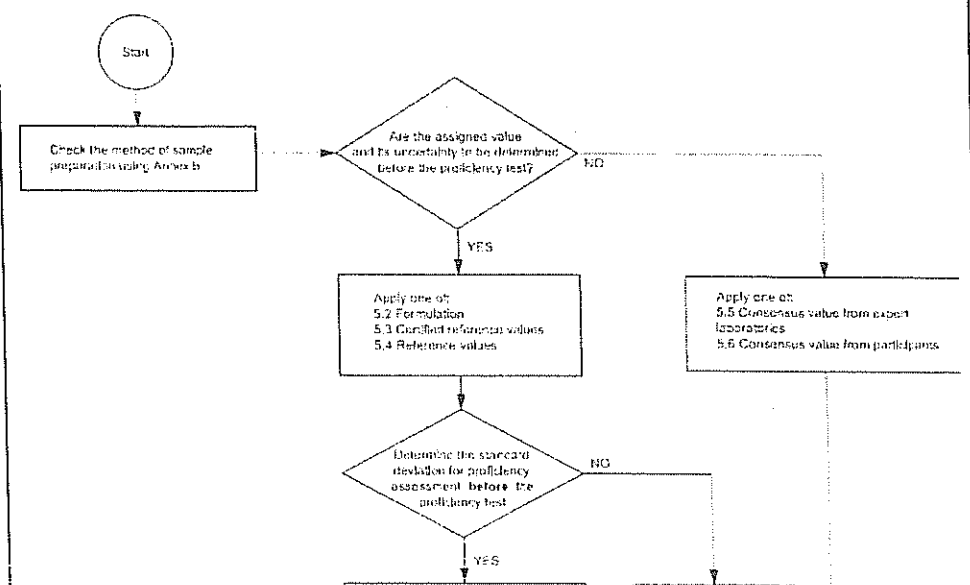




การออกแบบ  
กิจกรรม  
การทดสอบความชำนาญ

ดร.พอมาน ทำจิ้น

ตาม ISO 13528

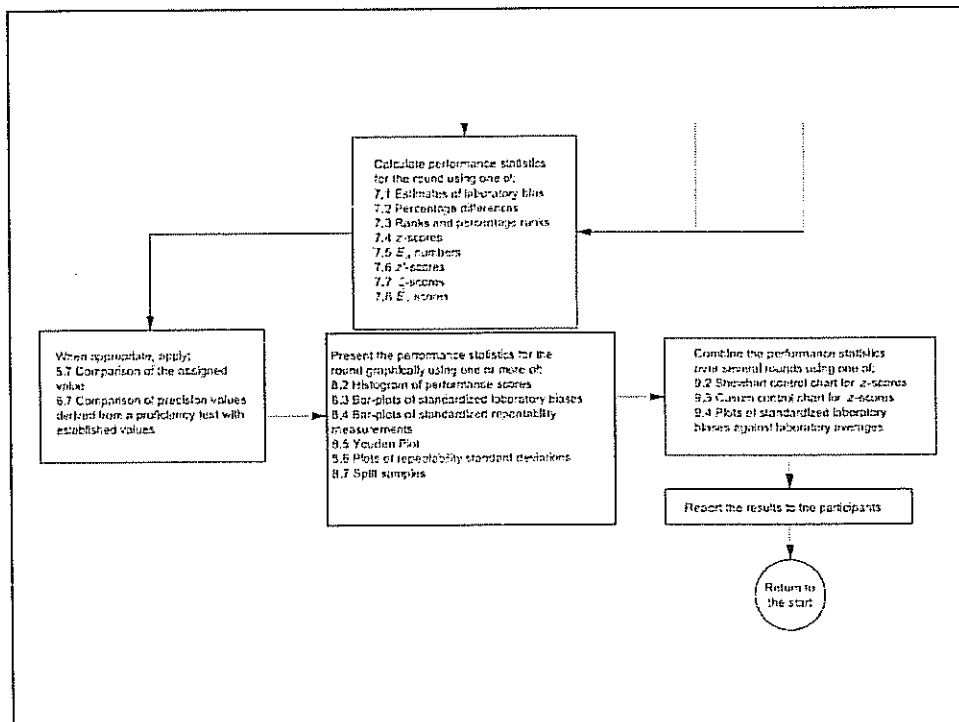
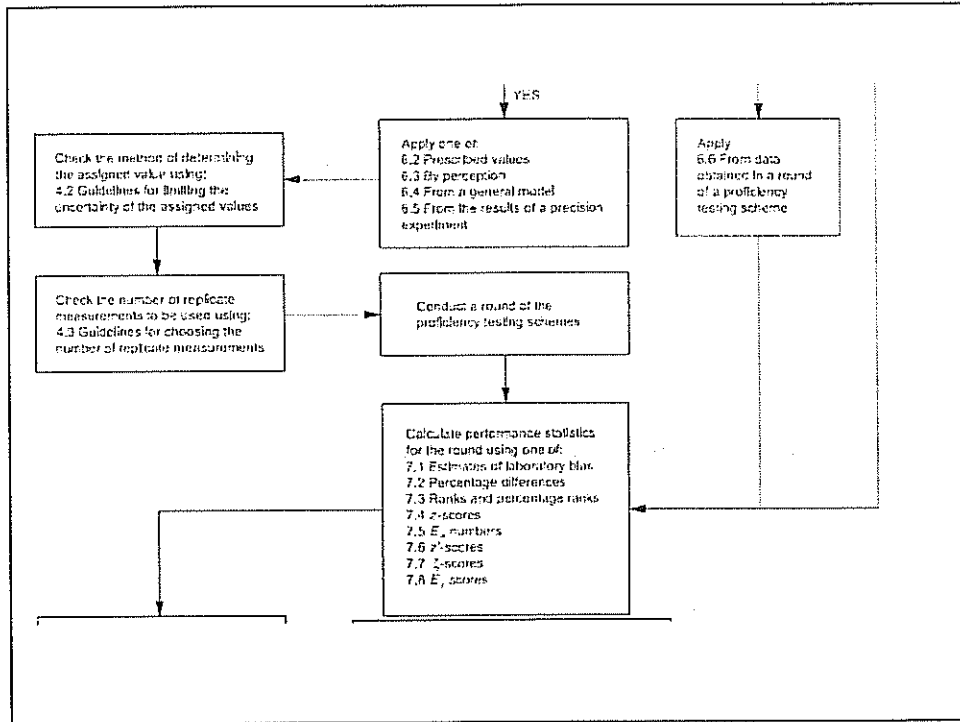


การอบรมเชิงปฏิบัติการ Statistical approach for PT (Testing and Calibration)

วันที่ ๒๘-๓๐ พฤษภาคม ๒๕๕๕

ณ ห้อง ๓๒๐ อาคารสถานศึกษาเคมีปฏิบัติ กรมวิทยาศาสตร์บริการ

จัดโดย สำนักบริหารและรับรองห้องปฏิบัติการ



การอบรมเชิงปฏิบัติการ Statistical approach for PT (Testing and Calibration)

วันที่ ๒๘-๓๐ พฤษภาคม ๒๕๕๕

ณ ห้อง ๓๒๐ อาคารสถานศึกษาเคมีปฏิบัติ กรมวิทยาศาสตร์บริการ

จัดโดย สำนักบริหารและรับรองห้องปฏิบัติการ

### ขั้นตอนการออกแบบทางสถิติ

- การออกแบบโปรแกรม (program design)
- การทดสอบความเป็นเนื้อเดียวกัน และความคงตัว (homogeneity and stability test)
- ขั้นตอนการปฏิบัติงานที่ส่งให้ผู้ร่วมกิจกรรม (instructions to participant)
- ค่ากำหนด (assigned value)
- การประเมินค่า (evaluation of results)
- การรายงานผล และการติดตามผล (reporting and follow up)

### การออกแบบโปรแกรม

Lab.	Level								
	1		2		j		q		
	a	b	a	b	a	b	a	b	
1									
2									
i									
p									

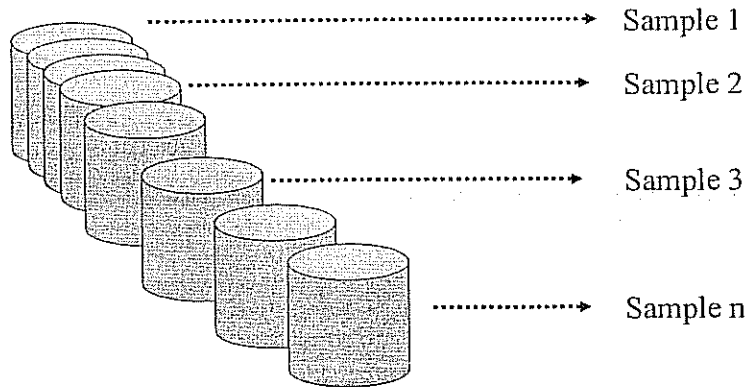
การอบรมเชิงปฏิบัติการ Statistical approach for PT (Testing and Calibration)

วันที่ ๒๔-๓๐ พฤษภาคม ๒๕๕๕

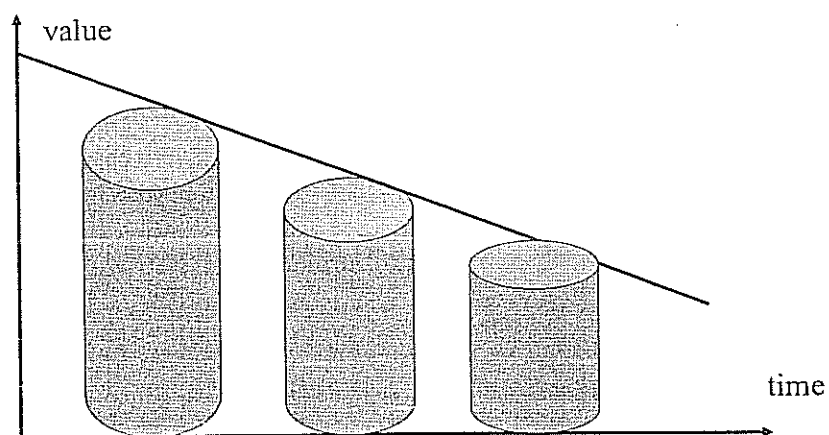
ณ ห้อง ๓๒๐ อาคารสถานศึกษาเคมีปฏิบัติ กรมวิทยาศาสตร์บริการ

จัดโดย สำนักบริหารและรับรองห้องปฏิบัติการ

## การทดสอบความเป็นเนื้อเดียวกัน



## การทดสอบความคงตัว



การอบรมเชิงปฏิบัติการ Statistical approach for PT (Testing and Calibration)

วันที่ ๒๘ -๓๐ พฤษภาคม ๒๕๕๕

ณ ห้อง ๓๒๐ อาคารสถานศึกษาเคมีปฏิบัติ กรมวิทยาศาสตร์บริการ

จัดโดย สำนักบริหารและรับรองห้องปฏิบัติการ

## ขั้นตอนการปฏิบัติงานที่ส่งให้ผู้ร่วมกิจกรรม

LAB	A1	A2	B1	B2	C1	C2
1						
2						
..						
..						
P						

## ค่ากำหนด (assigned value)

- Known value ; *by definition*
- Certified reference values; *use definition method*
- Reference values; *traceable to national standard*
- Consensus values from expert laboratories;
- Consensus values from participants.

การอบรมเชิงปฏิบัติการ Statistical approach for PT (Testing and Calibration)

วันที่ ๒๘ -๓๐ พฤษภาคม ๒๕๕๔

ณ ห้อง ๓๒๐ อาคารสถานศึกษาเคมีปฏิบัติ กรมวิทยาศาสตร์บริการ

จัดโดย สำนักบริหารและรับรองห้องปฏิบัติการ

## Scores

Z-score (zed score):	$z = \frac{x_i - X}{\hat{\sigma}}$
Z'score (adjusted zed score):	$z' = \frac{x_i - X}{\sqrt{\hat{\sigma}^2 + u_x^2}}$
$\zeta$ -score (zeta score):	$\zeta = \frac{x_i - X}{\sqrt{u_i^2 + u_x^2}}$
En-score:	$E_n = \frac{x_i - x_{ref}}{\sqrt{u_i^2 + u_{ref}^2}}$
Ez-score:	$E_z = \frac{x_i - (x_{ref} \pm u_{ref})}{u_i^2} \quad  E_{z-}  < 1 \quad  E_{z+}  < 1$

## การประเมินผล

- z-score	$z = \frac{x_i - X}{\hat{\sigma}}$
$ z  \leq 2$	= satisfactory
$2 <  z  < 3$	= questionable
$ z  \geq 3$	= unsatisfactory

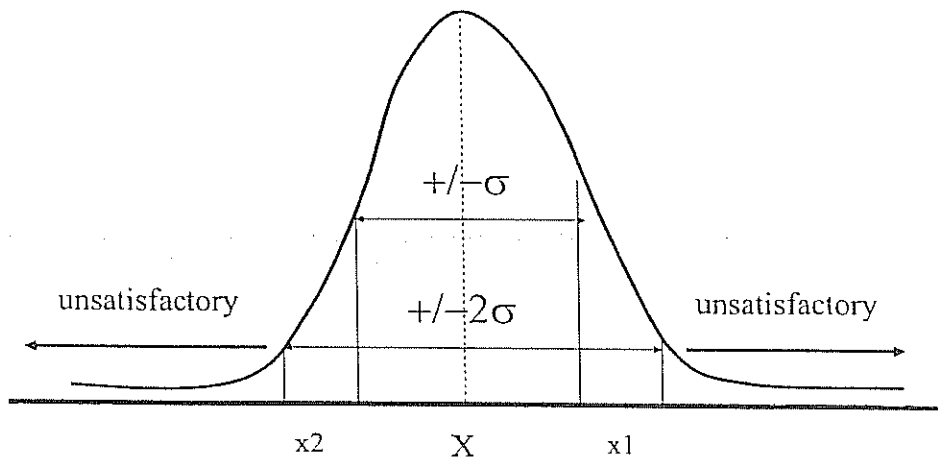
การอบรมเชิงปฏิบัติการ Statistical approach for PT (Testing and Calibration)

วันที่ ๒๘-๓๐ พฤษภาคม ๒๕๕๕

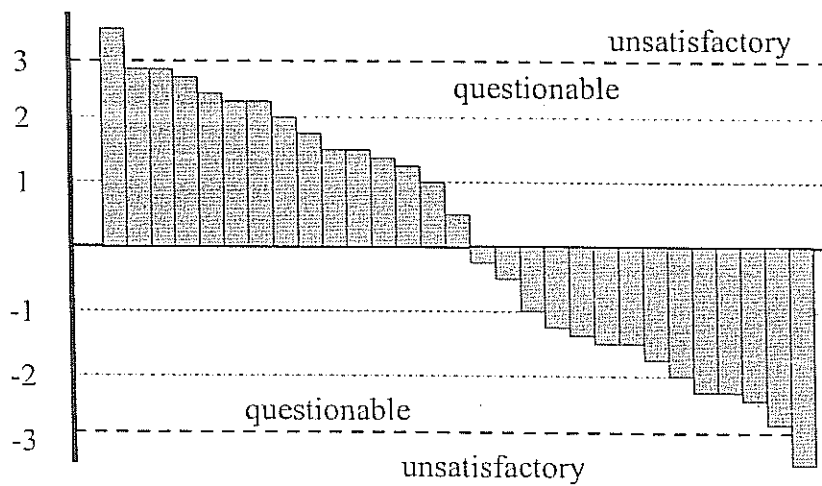
ณ ห้อง ๓๒๐ อาคารสถานศึกษาเคมีปฏิบัติ กรมวิทยาศาสตร์บริการ

จัดโดย สำนักบริหารและรับรองห้องปฏิบัติการ

## Z-score



## Z-score



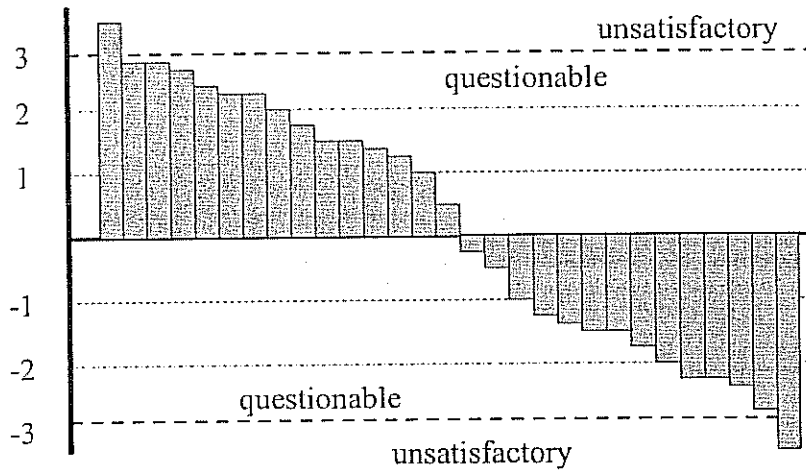
การอบรมเชิงปฏิบัติการ Statistical approach for PT (Testing and Calibration)

วันที่ ๒๔-๓๐ พฤษภาคม ๒๕๕๕

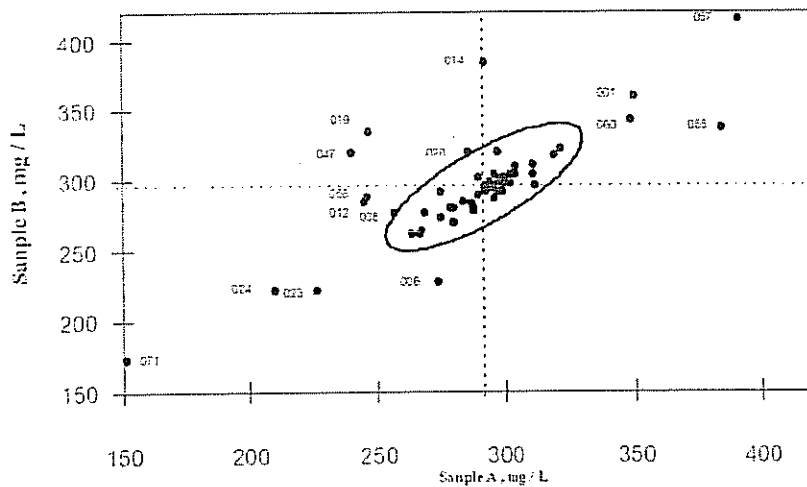
ณ ห้อง ๓๒๐ อาคารสถานศึกษาเคมีปฏิบัติ กรมวิทยาศาสตร์บริการ

จัดโดย สำนักบริหารและรับรองห้องปฏิบัติการ

## Z-score



Youden Plot of COD (mg / L)



รูปที่ 3 แผนภูมิ Youden plot แสดงความผันแปรของผลการทดสอบ Chemical Oxygen Demand (COD) ระหว่าง Sample A กับ Sample B ของห้องปฏิบัติการที่เข้าร่วมโครงการ (ในระดับความเชื่อมั่น 95 %)

การอบรมเชิงปฏิบัติการ Statistical approach for PT (Testing and Calibration)

วันที่ ๒๘ -๓๐ พฤษภาคม ๒๕๕๕

ณ ห้อง ๓๒๐ อาคารสถานศึกษาเคมีปฏิบัติ กรมวิทยาศาสตร์บริการ

จัดโดย สำนักบริหารและรับรองห้องปฏิบัติการ



## การประเมินผล (สอบเทียบ)

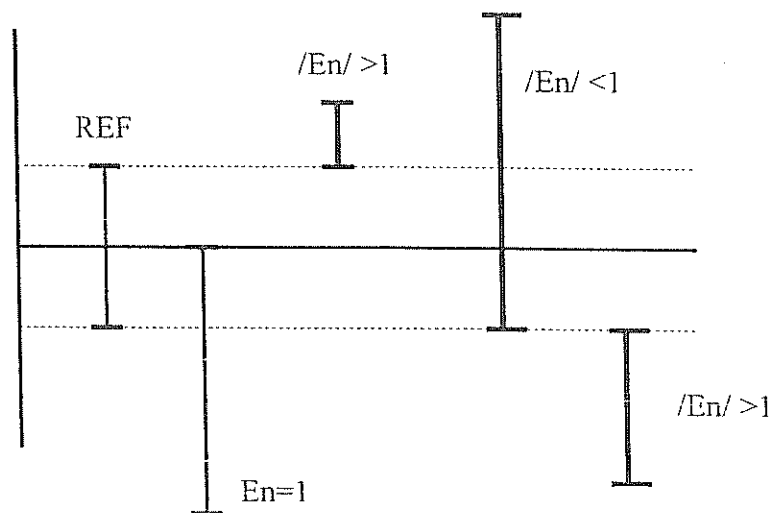
- En number 
$$E_n = \frac{x_i - X_{ref}}{\sqrt{u_i^2 + u_{ref}^2}}$$

$|E_n| \leq 1$  = satisfactory

$E_n > 1$  = unsatisfactory

หมายเหตุ  $u_i$  จะต้องเป็นค่า BMC ของห้องปฏิบัติการ

## En number



การยอมรับเชิงปฏิบัติการ Statistical approach for PT (Testing and Calibration)

วันที่ ๒๘ -๓๐ พฤษภาคม ๒๕๕๕

ณ ห้อง ๓๒๐ อาคารสถานศึกษาเคมีปฏิบัติ กรมวิทยาศาสตร์บริการ

จัดโดย สำนักบริหารและรับรองห้องปฏิบัติการ

## Scores

Z-score (zed score):	$z = \frac{x_i - X}{\hat{\sigma}}$
Z'score (adjusted zed score):	$z' = \frac{x_i - X}{\sqrt{\hat{\sigma}^2 + u_x^2}}$
$\zeta$ -score (zeta score):	$\zeta = \frac{x_i - X}{\sqrt{u_i^2 + u_x^2}}$
En-score:	$E_n = \frac{x_i - x_{ref}}{\sqrt{u_i^2 + u_{ref}^2}}$
Ez-score:	$E_z = \frac{x_i - (x_{ref} \pm u_{ref})}{u_i^2} \quad  E_{z-}  < 1 \quad  E_{z+}  < 1$

การสุ่ม PT sample

วิทยากร: ดร.พจมาน ทำจิ้น



## Sampling

- วัตถุประสงค์การสุ่มตัวอย่าง

วัตถุประสงค์ของการสุ่มตัวอย่าง เพื่อให้ได้ตัวอย่างที่ให้ผลสัมฤทธิ์ที่ดีและให้ผลการคาดคะเนที่มีประสิทธิภาพที่สุด การคาดคะเนจึงเป็นวิธีการของการกำหนดค่าของการกระจายจากตัวอย่าง ซึ่งจะใช้แบบจำลองทางสถิติของประชากร การสุ่มตัวอย่างควรจะใช้ค่าใช้จ่ายที่น้อยที่สุดโดยให้ค่าความเที่ยงตรงเพียงพอกับวัตถุประสงค์ของการวัด

## Sampling

- Target Population
- Matched Samples
- Independent Samples
- Random Sampling
- Simple Random Sampling
- Stratified Sampling
- Cluster Sampling

## Sampling

- Quota Sampling
- Spatial Sampling
- Sampling Variability
- Standard Error
- Bias
- Precision

## Target Population

- The target population คือสมาชิก  
ของกลุ่มประชากรในขอบข่ายของ  
การศึกษาทั้งหมด

## Matched Samples

- a) ตัวอย่าง 2 ตัวอย่างที่เป็นคู่อย่างชัดเจน เช่น การศึกษา IQ ของเด็กคู่แฝด
- b) เป็นการศึกษาให้ทำการวัด 2 ครั้ง ในแต่ละตัวอย่าง โดยทั่วไปเรียกว่าการทวนซ้ำ

## Independent Sampling

Independent samples คือสมาชิกตัวอย่างที่เลือกมาจากกลุ่มประชากรเดียวกัน หรือต่างกลุ่มประชากร โดยแต่ละตัวอย่างไม่มีผลต่อกัน และไม่มีความสัมพันธ์ร่วมกัน

## Random Sampling

- Random sampling คือเทคนิคการสุ่มตัวอย่างให้ได้กลุ่มตัวอย่างเพื่อศึกษาผลจากกลุ่มใหญ่นั้นคือกลุ่มประชากร แต่ละตัวอย่างถูกสุ่มอย่างสุ่ม ขึ้นกับโอกาสที่ถูกสุ่ม โดยแต่ละตัวอย่างมีโอกาสจะถูกสุ่มเท่าๆ กัน

## Stratified Sampling

- เป็นการแบ่งประชากรออกเป็น sub-population (group/strata) และดำเนินการวัดในแต่ละ sub-population



## Cluster Sampling

- Cluster sampling คือเทคนิคการสุ่มตัวอย่างโดยการแบ่งประชากรทั้งหมดเป็นกลุ่ม หรือ cluster และดำเนินการสุ่ม cluster อย่าง random

## Quota Sampling

- Quota sampling คือวิธีการสุ่มตัวอย่างที่ใช้มากในการสอบถามความคิดเห็น และการวิจัยตลาด โดยมีจำนวนและขนาดของตัวอย่างกำหนดไว้ เช่น การสัมภาษณ์การสนใจดูรายการโทรทัศน์ โดยการสัมภาษณ์ผู้ใหญ่ชาย 20 คน ผู้ใหญ่หญิง 20 คน วัยรุ่นชาย 10 คน วัยรุ่นหญิง 10 คน

## Spatial Sampling

- เป็นการพิจารณาพื้นที่การสุ่ม ตัวอย่าง ด้วยการสุ่ม 2 หรือมากกว่า 2 พื้นที่

## Sampling Variability

- คือค่าความแตกต่างของแต่ละ ตัวอย่างที่คำนวณจากตัวอย่างสุ่ม จากกลุ่มประชากร

## Sample size

- การกำหนดขนาดของตัวอย่างสิ่งแรกต้องพิจารณาเปอร์เซ็นต์ความเที่ยงตรงของผลการวัด เช่น 5% (0.05) , 2.5% (0.025) เป็นต้น

การกำหนดขนาดของตัวอย่างอาจพิจารณาได้ จากค่าความแปรปรวนของตัวแปรสุ่ม ดังนี้คือ

## Sample size (ต่อ)

$$n_0 = \frac{p(1-p)}{V}$$

- เมื่อ

$n_0$  คือ ขนาดของตัวอย่าง

$p$  คือ ความน่าจะเป็นของตัวอย่างสุ่ม

$V$  คือ ความแปรปรวนของตัวแปรสุ่ม



# การทดสอบความเป็นเนื้อเดียวกันของ PT sample

วิทยากร: ผศ.ดร.ครรชิต จุดประสงค์  
นางจันทร์รัตน์ วรสรรพวิทย์





# Statistical approach for PT (Testing and Calibration)

Asst. Prof. Kunchit Judprasong\*

Assoc. Prof. Prapasri Puwastien\*

Chantarat Vorasapavit\*\*

\*Institute of Nutrition, Mahidol University, Salaya, Putthamonthon 4,  
Nakhon Pathom 73170, THAILAND  
Tel: 0 2800 2380 ext . 324

\*\* Department of Science and Service

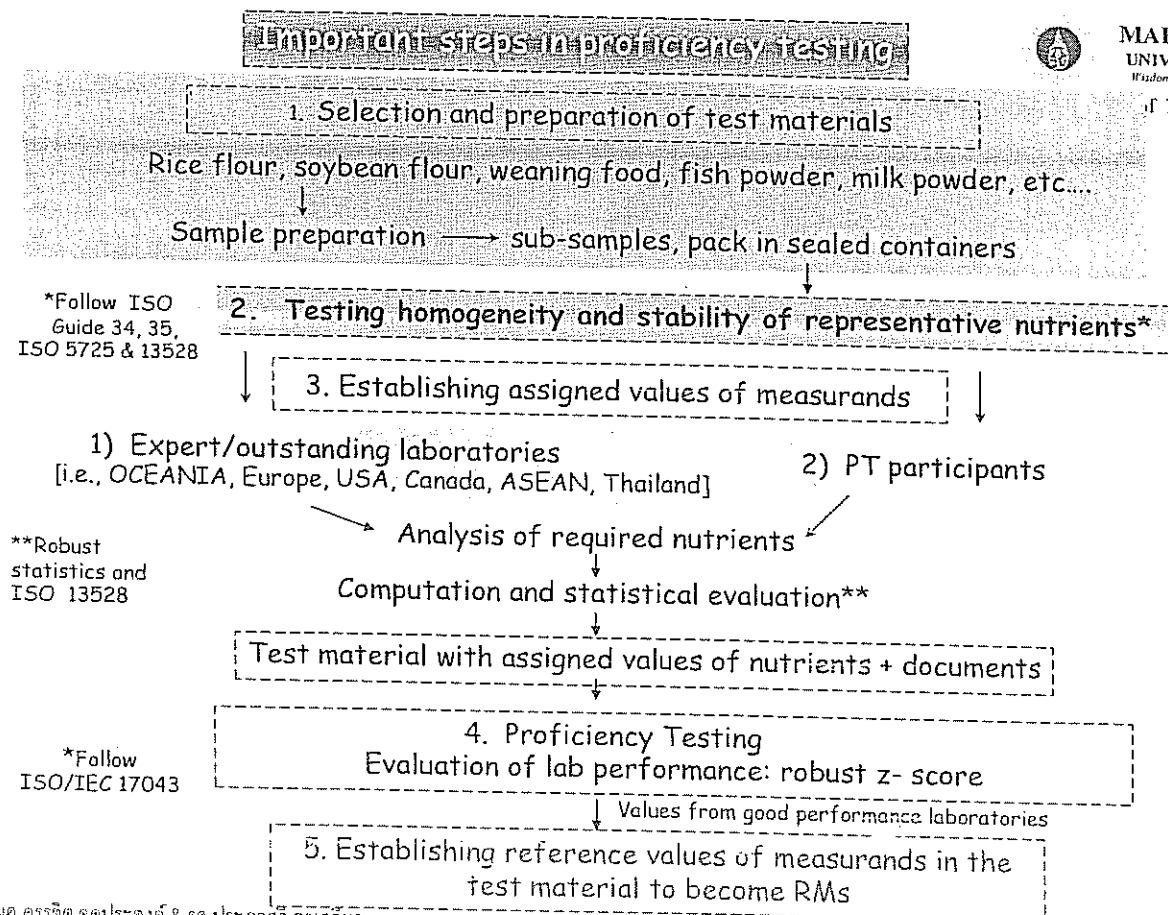
[kunchit.jud@mahidol.ac.th](mailto:kunchit.jud@mahidol.ac.th)

Statistical approach for PT (Testing and Calibration)  
(APLAC-PTB approved work plan)  
28-30 พฤษภาคม 2555

ผศ.ดร.วิชิต จุดประจงค์ & รศ.ประภาศรี กุลเสถียร

1

## Important steps in proficiency testing



ผศ.ดร.วิชิต จุดประจงค์ & รศ.ประภาศรี กุลเสถียร

2



Outline for presentation

## Statistics for Proficiency testing

- Testing homogeneity - statistical evaluation
- Testing stability - statistical evaluation
- Estimating assigned values of measurands
- Evaluating laboratory performance



## References for PT Statistics

- ISO 13528 "Statistical Methods for use in Proficiency Testing by Interlaboratory Comparisons, Geneva, Switzerland 2005.
- ISO/IEC 17043: Conformity Assessment - General requirements for Proficiency Testing, Geneva, Switzerland 2010.
- Thompson M, Ellison SLR, Wood R. The International harmonised protocol for the proficiency testing of analytical chemistry laboratories (IUPAC Technical Report). Pure Appl Chem 78: 145-196, 2006.
- ISO 5725. Accuracy (trueness and precision) of measurement methods and results. 1994.
- Miller JN & Miller CM. Statistics and Chemometrics for Analytical Chemistry. Fourth edition. 2000. Pearson Education Limited., England.
- Other APLAC, NATA Statistical procedures





## Statistical evaluation in Proficiency Testing

### 1. Demonstration of proficiency test item homogeneity and stability



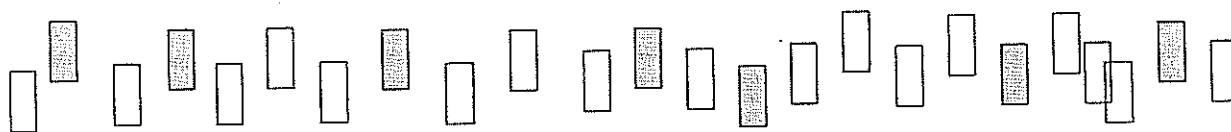
## Part I.

### 1A. Testing for "sufficient homogeneity" of a test material

Objective for statistical evaluation: to show that variation in composition among the distributed units (characterised by sampling SD,  $S_s$  or  $s_{sam}$ ) is negligible in relation to variation introduced by the measurements conducted by the participants in the proficiency testing ( $S_p$  or  $s_p$ )

## Testing for sample homogeneity

Homogenised test materials, packed in laminated aluminum foil bags or in bottles



selected 10 representative samples at random

Analysed *randomly* for representative components\*, *in duplicate*

Statistical analysis without exclusion of any value

\*each, measured on the same day(s), using the same conditions, by the same competent analyst, using same method, instrument and calibrants

ผศ.ดร.ชิต จุตประสงค์ & รศ.ประภาศรี ภูวเดทีย

7

- Select 10 sets of test materials at random for homogeneity checking



MAHIDOL UNIVERSITY  
Wisdom of the Land

Institute of Nutrition

1	2	3	4	5	6	7	8	9	10
11	12	13	14	15	16	17	18	19	20
21	22	23	24	25	26	27	28	29	30
31	32	33	34	35	36	37	38	39	40
41	42	43	44	45	46	47	48	49	50
51	52	53	54	55	56	57	58	59	60
61	62	63	64	65	66	67	68	69	70
71	72	73	74	75	76	77	78	79	80
81	82	83	84	85	86	87	88	89	90
91	92	93	94	95	96	97	98	99	100
101	102	103	104	105	106	107	108	109	110
111	112	113	114	115	116	117	118	119	120
121	122	123	124	125	126	127	128	129	130
131	132	133	134	135	136	137	138	139	140
141	142	143	144	145	146	147	148	149	150
151	152	153	154	155	156	157	158	159	160
161	162	163	164	165	166	167	168	169	170
171	172	173	174	175	176	177	178	179	180
181	182	183	184	185	186	187	188	189	190
191	192	193	194	195	196	197	198	199	200

No.	Random No.
1	25
2	131
3	59
4	117
5	126
6	179
7	68
8	13
9	11
10	130

- Analyse for specific components at random, in duplicate or triplicate, on the same day using the same reagents, instrument, conditions, and the same analysts for each parameter

ผศ.ดร.ชิต จุตประสงค์ & รศ.ประภาศรี ภูวเดทีย

8

# Statistical analysis: homogeneity testing

## 2 Steps:

Step 1. Checking for within sample variation (precision of the analyst): using Cochran's maximum range test

Step 2. Checking for between sample variation using

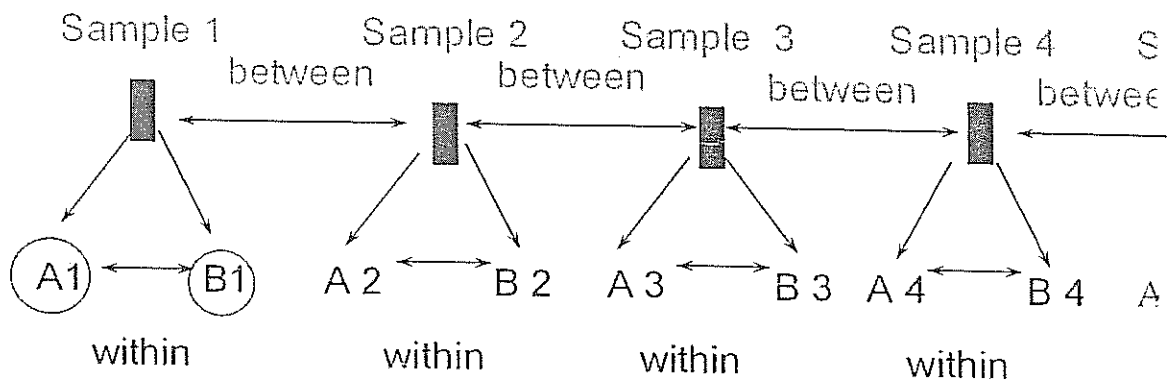
- One Way ANOVA
- ISO 13528:2005
- IUPAC (Thompson et al., 2006)
- %RSD, HORRAT approach

1) Thompson M, Ellison SLR, Wood R. The International harmonised protocol for the proficiency testing of analytical chemistry laboratories (IUPAC Technical Report). Pure Appl Chem 78: 145-196, 2006.

2) ISO 5725-1994: Accuracy (trueness and precision) of measurement methods and results.

3) ISO 13528: 2005 Statistical methods for use in proficiency testing of interlaboratory comparisons.

## Check homogeneity of the test materials



## Check precision of testing laboratory (analyst who does the work)



### 1) Checking for within sample variation: Cochran's test

**Duplicate analysis:**  
Cochran's maximum range test

Calculate the ratio:  $\frac{D_{max}^2}{\sum D_i^2}$

where

$D_{max}$  = the maximum difference of the duplicates  
 $D_i$  = difference of each pair of duplicates

**Triplicate analysis:**  
Cochran's maximum variance test

Calculate the ratio:  $\frac{S_{max}^2}{\sum S_i^2}$

where

$S_{max}^2$  = the maximum variance in the set  
 $S_i^2$  = variance of each set of triplicate results

- Compare the ratio to the critical value from Cochran Table (95% confidence).
- If the ratio is < critical value, there is no evidence of analytical outliers
- If any outlier is detected, results must be inspected closely to see any errors and appropriate action must be taken.
- An outlier set should not be rejected unless it is significant at 99% level or any permanent analytical errors are found. It should be excluded from ANOVA unless there is a reason to the contrary.



### Cochran's maximum range test: within sample variation

No.	Total protein (g/100g)			
	A	B	Range <sup>2</sup> (R)	R/R-total
1	7.99	7.90	0.008	0.25
2	7.92	7.91	0.000	0.00
3	7.96	7.87	0.008	0.25
4	7.97	7.93	0.002	0.05
5	7.93	7.85	0.006	0.19
6	7.89	7.87	0.000	0.01
7	7.95	7.93	0.000	0.01
8	7.90	7.92	0.000	0.01
9	7.92	7.97	0.002	0.08
10	7.89	7.96	0.005	0.15
<b>R Total</b>			$\sum D_i^2$ → 0.033	
<b>N</b>			10	
<b>Cochran critical value<sup>(a)</sup> at 95% confidence level</b>			0.602	

Ratio  $C = \frac{D_{max}^2}{\sum D_i^2}$

$C_{max} < C_{crit}$   
= no evidence of analytical outliers  
= Analysts who performed the tests have good performance

<sup>(a)</sup>Critical value for Cochran's maximum variance test for 10 sets of data, when the number of results per set (n) = 2, given in the ISO 5725



A1.2.3 Tables of critical values for homogeneity testing

Table 1 Critical values for the Cochran test statistic for duplicates.

<i>m</i>	95 %	99 %
7	0.727	0.838
8	0.68	0.794
9	0.638	0.754
10	0.602	0.718
11	0.57	0.684
12	0.541	0.653
13	0.515	0.624
14	0.492	0.599
15	0.471	0.575
16	0.452	0.553
17	0.434	0.532
18	0.418	0.514
19	0.403	0.496
20	0.389	0.480

Thompson M, Ellison SLR, Wood R. The International harmonised protocol for the proficiency testing of analytical chemistry laboratories (IUPAC Technical Report). Pure Appl Chem 78: 145-196, 2006.



Statistical Analysis: sample homogeneity (continued)

2) Checking for between sample variation: sample homogeneity

(1) One Way ANOVA (allowed but not recommend)

(Microsoft Excel or commercial software)

**Null Hypothesis (H<sub>0</sub>):** values of a measured parameter samples are not significantly different

$$F\text{-value} = \frac{MSB}{MSW} < \text{Critical F-value}$$

where MSB: mean square between

MSW: mean square within

"Samples are homogeneous" = the differences between test items will not affect the between laboratory results.

Limitation: when MSW (good precision)  $\ll$  MSB  $\rightarrow$  not pass ???  
Interpretation of results must be careful!!!!



## การวิเคราะห์ความแปรปรวน (Analysis of Variance : ANOVA)

- ใช้สำหรับทดสอบค่าเฉลี่ยของประชากร  $\geq 3$  ประชากร
- การวิเคราะห์ความแปรปรวนทางเดียว ( One-way ANOVA)
- การวิเคราะห์ความแปรปรวนสองทาง (Two-way ANOVA)
- เงื่อนไข
  - ประชากรทั้ง  $a$  ประชากร มีการแจกแจงแบบปกติ
  - ค่าความแปรปรวนของแต่ละประชากรมีค่าเท่ากัน
  - การสุ่มตัวอย่างแต่ละชุดจากแต่ละประชากรจะเป็นอิสระกัน



## การวิเคราะห์ความแปรปรวนทางเดียว (One-way ANOVA)

- หลักการ ค่าความแปรปรวนที่เกิดขึ้นจากการทดลอง แยกตามแหล่งได้ 2 แหล่ง คือ
  - **Between Group** เป็นผลจากการที่หน่วยทดลองได้รับ **treatment** ที่แตกต่างกัน
  - **Within Group** เป็นความแตกต่างของหน่วยทดลองที่ได้รับ **treatment** เดียวกัน

สมมติฐานการทดสอบ

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \dots = \mu_a$$

$$H_1 : \mu_i \neq \mu_j \text{ อย่างน้อย 1 คู่ } i \neq j$$

สถิติทดสอบ

$$F = \frac{MS_{\text{Treatment}}}{MSE}, \quad df = (a - 1, na - a)$$

เขตปฏิเสธ

$$F > F_{\alpha, a-1, na-a}$$

Example: protein (g/100g) in rice powder by  
 One way Anova

No.	Result	
	A	B
1	7.99	7.90
2	7.92	7.91
3	7.96	7.87
4	7.97	7.93
5	7.93	7.85
6	7.89	7.87
7	7.95	7.93
8	7.90	7.92
9	7.92	7.97
10	7.89	7.96
Mean	7.93	7.91

$$F\text{-value} = \frac{MSB}{MSW} < \text{Critical F-value}$$

One way ANOVA:  
 between samples variation

F-value < F critical  
 Sample is  
 homogeneous

Using Excel: Data analysis

ANOVA						
Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	0.023	9	0.0025	0.71	0.69	3.02
Within Groups	0.035	10	0.0035			
Total	0.058	19				

## Criteria for homogeneity testing



### Criteria 1. ANOVA - F value

$$F\text{-value} = \frac{MSB}{MSW} < \text{Critical F-value}$$

where MSB: mean square between  
MSW: mean square within

"Samples are homogeneous" = the differences between test items will not affect the between laboratory results.

If F-value is > critical F-value → sample inhomogeneity

→ determine a "tolerable" level → fallback criteria  
(ISO 13528: 2005)

### Summary results of sample homogeneity testing of salted broad bean powder: based on One-Way ANOVA



Parameter	Mean squares between (MSb)	Mean squares within (MSw)	F-value= (MSb/MSw)	P-value	Conclusion
			Critical F-value (df=9,10) = 3.02		
Moisture (g/100g)	0.0141	0.0005	31.43	0.0000	Not pass
Protein (g/100g)	0.0450	0.0788	0.57	0.7936	Pass
Total fat (g/100g)	1.0562	0.1606	6.58	0.0034	Not pass
Dietary fibre (g/100g)	1.7226	0.0875	19.68	0.0000	Not pass
Ash (g/100g)	0.0059	0.0006	10.13	0.0006	Not pass
Sodium (mg/100g)	409.4293	31.6225	12.95	0.0002	Not pass
Vitamin B <sub>1</sub> (mg/100g)	0.00009	0.00003	2.89	0.0569	Pass



Criteria 2: ISO 13528: 2005<sup>(1)</sup> and  
Fallback criteria: IUPAC guideline<sup>(2)</sup>

Between sample variation = within sample variation + sampling variation  
(analytical variation) (S<sub>s</sub>)

$$s_s \leq 0.3\hat{\sigma}^{(1, 2)}$$

If sampling variation  $\leq$  0.3 times **target SD**  
(Sampling SD or S<sub>s</sub>)

The test materials is sufficiently homogeneous.  
"Sample variation will not significantly affect the between-laboratory results"

(1) ISO 13528. Statistical methods for use in proficiency testing by interlaboratory comparisons, First edition, 2005.

(2) IUPAC (International harmonisation for the PT of analytical lab), 2006

## การคำนวณค่า S<sub>s</sub>

(between- samples standard deviation)

2 วิธี

1) จากตาราง ANOVA

$$S_s = \sqrt{\frac{(MS_{between} - MS_{within})}{2}} \quad \text{ISO Guide 35}$$

2) จากสูตร

$$S_s = \sqrt{S_x^2 - \left(\frac{S_w^2}{2}\right)} \quad \text{ISO 13528}$$



## การคำนวณค่า S<sub>s</sub> (between- samples standard deviation)

เมื่อ S<sub>x</sub> = standard deviation of sample average

$$S_x = \sqrt{\frac{\sum_{i=1}^g (\bar{x}_i - \bar{\bar{x}})^2}{(g-1)}}$$

$\bar{x}_i$  = sample average

$\bar{\bar{x}}$  = general average

$$\bar{x}_i = \frac{x_{i1} + x_{i2}}{2}$$

$$\bar{\bar{x}} = \frac{\sum_{i=1}^g \bar{x}_i}{g}$$



## การคำนวณค่า S<sub>s</sub> (between- samples standard deviation)

เมื่อ S<sub>w</sub> = within-samples standard deviation

$$S_w = \sqrt{\frac{\sum_{i=1}^g w_i^2}{(2g)}}$$

W<sub>i</sub> = between-test portion ranges

$$w_i = |x_{i1} - x_{i2}|$$

## Example 1: calculation of sampling variation ( $S_s$ )

Total fat results in salted broad bean powder from One-way ANOVA:

MSB = 1.0562, MSW = 0.1606, F-value is 6.576,  
Critical F-value (9, 10) = 3.02 ( $p=0.05$ )

$$S_s^2 = \text{Sampling variance} \\ = \frac{(MSB - MSW)}{2}$$

$$\text{Sampling variation } (S_s) = \sqrt{\frac{(MSB - MSW)}{2}}$$

$$= \sqrt{\frac{(1.0562 - 0.1606)}{2}}$$

$$= 0.669 \Rightarrow \text{Compare to "target SD"}$$

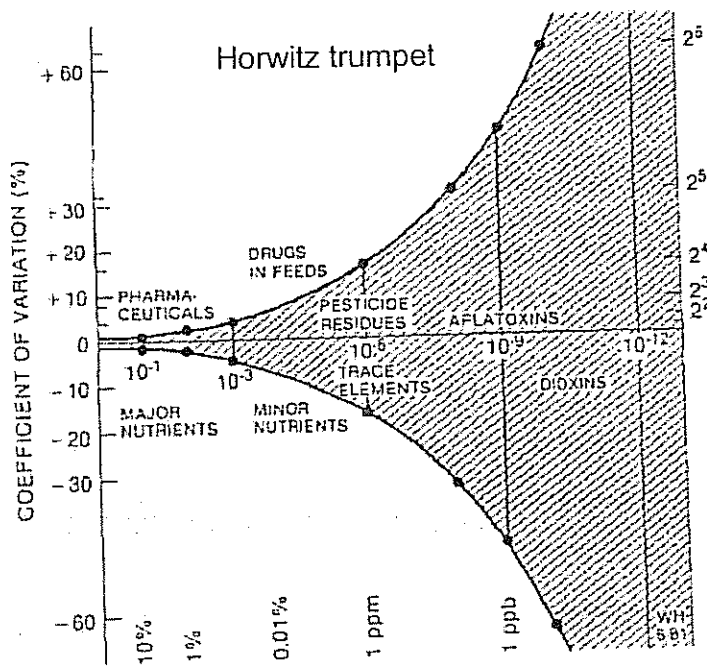
## How to get "target standard deviation:

1. By expert perception: fixing arbitrarily
2. By prescription: i.e., estimate from reproducibility coefficient variation (actual variation) to fit-for-purpose
3. By reference to validated methodology: estimate from RSD obtained from method performance study (collaborative study)
4. By reference to a generalised model:

"Horwitz' s equation"



# Overall interlaboratory relative standard deviation



Between-laboratory variation

Conc	Expected % RSD(R)
100%	2
1%	4
0.01%	8
1 ppm	16
10 ppb	32
1 ppb	45

Ref: William Horwitz, Richard Albert. Precision in analytical measurements: Expected values and consequences in geochemical analyses. *Fresenius J Anal Chem* (1995) 351:507-513

ผศ.ดร.चित จิตประสงค์ & รศ.ประภาสรี ภูวเสถียร



## Target SD: calculated by "Horwitz equation"

$$\text{Horwitz predicted RSD (RSD}_p) = 2^{(1-0.5 \log C)} \text{ or } 2C^{-0.1505}$$

where "C" is expressed as a mass fraction concentration  
(i.e., concentration is 10 mg/100g,  $C = 10/100/1000 = 0.0001$ )

$$\text{RSD}_p \text{ (\%CV)} = \frac{\text{SD} \times 100}{\text{Mean}}$$

Therefore,

$$\text{Target SD (SD}_p) = \frac{\text{RSD}_p \times \text{Mean}}{100}$$

ผศ.ดร.चित จิตประสงค์ & รศ.ประภาสรี ภูวเสถียร

$$\text{Horwitz predicted RSD (RSDp)} = 2^{1-0.5 \log C}$$

Example: i.e. average total fat = 27.82 g/100g

$$"C" \text{ (fractional concentration)} = \frac{27.82}{100}$$

$$C = 0.2782$$



$$\text{Target SD, SDp} = 0.675$$

Fallback criteria:  $\frac{\text{sampling variation}}{\text{target SD}} < 0.3$

Ss of vitamin B<sub>1</sub> = 0.005

$$S_s/SDp = \frac{0.669}{0.675} = 0.99 \Rightarrow \text{X not pass the fallback criteria}$$

The samples seem not to be homogeneous based on ISO 13528 approach

### Summary sample homogeneity testing of broad bean powder: based on ISO 13528 (2005)

Parameter	Mean	SD	%RSD or %CV	Sampling SD (Ss)	Target SD from Horwitz	Ss/SDp	Conclusion Pass if Ss/SDp < 0.3
Moisture (g/100g)	2.79	0.08	3	0.08	0.096	0.84	Not pass
Protein (g/100g)	22.62	0.15	0.7	0.13	0.566	0.23	Pass
Total fat (g/100g)	27.82	0.73	2.6	0.67	0.675	0.99	Not pass
Dietary fibre (g/100g)	12.72	0.93	7.3	0.9	0.347	2.59	Not pass
Ash (g/100g)	3.02	0.05	1.8	0.05	0.102	0.49	Not pass
Sodium (mg/100g)	460	14	3.1	13.74	20.698	0.66	Not pass
Vitamin B1 (mg/100g)	0.24	0.01	2.7	0.01	0.034	0.29	Pass



### Criteria 3: IUPAC (Thompson et al., 2006)

"sufficiently homogeneous"

$$s_{sam}^2 < F_1 \sigma_{all}^2 + F_2 s_{an}^2$$

If  $s_{sam}^2 > c$ , there is evidence (significant at the 95% level of confidence) that the sampling standard deviation in the population of samples exceeds the allowable fraction of the target standard deviation, and the test for homogeneity has failed.

If  $s_{sam}^2 < c$ , there is no such evidence, and the test for homogeneity has been passed.

"Sample variation will not significantly affect the between laboratory results"

### Criteria 3: IUPAC (con't)



Between sample variation = within sample variation + sampling variation  
(total variation) (analytical variation) (S<sub>s</sub>)

$s_{sam}^2$  = sampling variance

$$s_{sam}^2 = (V_s / 2 - s_{an}^2) / 2$$

$$V_s = \sum (s_i - \bar{s})^2 / (m-1)$$

$s_{an}^2$  = analytical variance

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

where:

$V_s$  = variance of sum in a homogeneity test

$F_1, F_2$  = constant factor in the Table

where  $D_i$  = difference of  $i^{th}$  pair of duplicates in a homogeneity test  
 $m$  = number of distribution units in a homogeneity test

$\sigma_{all}^2$  = allowable sampling variance

$$\sigma_{all}^2 = (0.3\sigma_p)^2 \quad \text{where } \sigma_p = \text{the target SD or SD for PT}$$

Table 2 Factors  $F_1$  and  $F_2$  for use in testing for sufficient homogeneity.

$m^*$	20	19	18	17	16	15	14	13	12	11	10	9	8	7
$F_1$	1.59	1.60	1.62	1.64	1.67	1.69	1.72	1.75	1.79	1.83	1.88	1.94	2.01	2.10
$F_2$	0.57	0.59	0.62	0.64	0.68	0.71	0.75	0.80	0.86	0.93	1.01	1.11	1.25	1.43

\* $m$  is the number of samples that have been measured in duplicate.

Example IUPAC: i.e. average total fat = 27.82 g/100g

$$s_{sam}^2 < F_1 \sigma_{all}^2 + F_2 s_{an}^2$$

Sample	Fat (g/100g)		D	S	D <sup>2</sup>
	a	b			
1	28.61	28.00	0.61	56.61	0.3721
2	28.53	28.68	-0.15	57.21	0.0225
3	28.02	28.26	-0.24	56.28	0.0576
4	28.10	28.21	-0.11	56.31	0.0121
5	26.95	28.00	-1.05	54.95	1.1025
6	28.75	27.84	0.91	56.59	0.8281
7	27.77	27.56	0.21	55.33	0.0441
8	27.87	28.19	-0.32	56.06	0.1024
9	27.37	27.62	-0.25	54.99	0.0625
10	25.65	26.43	-0.78	52.08	0.6084

$$s_{sam}^2 = (V_s / 2 - s_{an}) / 2$$

Variance of the sum:

$$V_s = \sum (s_i - \bar{s})^2 / (m - 1) = 2.112$$

Analytical variance

$$s_{an}^2 = \sum D_i^2 / 2m = 0.1606$$

Sampling variance:  $s_{sam}^2$

$$= (2.112/2 - 0.1606) / 2 = 0.448$$

Sum of (a-b)<sup>2</sup> = 3.2123

$$s_{an}^2 = \sum D_i^2 / 2m \text{ Analytical variance } s_{an}^2 = 0.1606$$

$$s_{sam}^2 < F_1 \sigma_{all}^2 + F_2 s_{an}^2$$

$\sigma_{all}^2$  = allowable sampling variance

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

The target SD (from Horwitz) = 0.675 g/100g

$$\sigma_{all}^2 = (0.3 \times 0.675)^2 = 0.041$$

$$F_1 \sigma_{all}^2 + F_2 s_{an}^2$$

$$= (1.88 \times 0.041) + (1.01 \times 0.1606)$$

$$= 0.239$$

Sampling variance = 0.448

Allowable fraction of the target standard deviation = 0.239

$$0.448 > 0.239$$



The test for homogeneity has failed

The samples seem not to be homogeneous based on IUPAC approach



## Summary sample homogeneity testing of broad bean powder: based on IUPAC approach

Parameter	Mean	Standard deviation (SD)	Relative standard deviation (RSD), %CV	Sampling variance ( $S_{stat}^2$ )	Critical value	Conclusion
					$c = F_1 \sigma_L^2 + F_2 S_{pr}^2$	
Moisture (g/100g)	2.79	0.08	3.02	0.0068	0.0020	Not passed
Protein (g/100g)	22.62	0.15	0.66	-0.0169	0.1338	Passed
Total fat (g/100g)	27.82	0.73	2.61	0.4478	0.2392	Not passed
Dietary fibre (g/100g)	12.72	0.93	7.30	0.8175	0.1088	Not passed
Ash (g/100g)	3.02	0.05	1.80	0.0026	0.0024	Not passed
Sodium (mg/100g)	460	14	3.11	188.903	104.422	Not passed
Vitamin B1 (mg/100g)	0.24	0.01	2.70	0.0000	0.0002	Passed



## Criteria 4: RSD (%CV) and HORRAT

$$\text{HORRAT} = \text{RSD}/\text{pRSD}^{(a,b)}$$

For interlaboratory studies: Acceptable HORRAT =  $(\text{RSD}_R/\text{pRSD}_R) < 1$   
(lower limit = 0.5, upper limit = 2.0)

For intralaboratory studies: Acceptable HORRAT  $(\text{RSD}_r/\text{pRSD}_r) = 0.3-1.3$   
(one half of lower limit to two thirds of the upper limit)

The test materials is sufficiently homogeneous.  
"Sample variation will not significantly affect the between laboratory results"

- (a) [http://www.aoac.org/dietsupp6/Dietary-Supplement-web-site/HORRAT\\_SLV.pdf](http://www.aoac.org/dietsupp6/Dietary-Supplement-web-site/HORRAT_SLV.pdf); Definitions and calculations of HORRAT values from intralaboratory data.
- (a) W. Horwitz and R. Albert. The Horwitz ratio (Horrat): a useful index of method performance with respect to precision. Journal of AOAC International, 89, 4, 2006, 1095-1108.



Summary sample homogeneity testing of  
broad bean powder: based on %RSD (%CV) and HORRAT

Parameter	Mean	Standard deviation (SD <sub>p</sub> )	Relative standard deviation (RSD <sub>r</sub> or %CV)	Target RSD from Horwitz RSD <sub>p</sub>	HORRAT RSD <sub>r</sub> /RSD <sub>p</sub>	Acceptable HORRAT (0.3-1.3) (a, b)
Moisture (g/100g)	2.79	0.08	3.0	3.4	0.88	Pass
Protein (g/100g)	22.62	0.15	0.7	2.5	0.27	Pass
Total fat (g/100g)	27.82	0.73	2.6	2.4	1.08	Pass
Dietary fibre (g/100g)	12.72	0.93	7.3	2.7	2.68	Not pass*
Ash (g/100g)	3.02	0.05	1.8	3.4	0.53	Pass
Sodium (mg/100g)	460	14	3.1	4.5	0.69	Pass
Vitamin B <sub>1</sub> (mg/100g)	0.24	0.01	2.7	14.0	0.19	Pass

\* Must be taken into account at further steps - estimation of assigned values, performance evaluation

(a) [http://www.aoac.org/dietsupp6/Dietary-Supplement-web-site/HORRAT\\_SLV.pdf](http://www.aoac.org/dietsupp6/Dietary-Supplement-web-site/HORRAT_SLV.pdf); Definitions and calculations of HORRAT values from intralaboratory data.

(b) W. Horwitz and R. Albert. The Horwitz ratio (Horrat): a useful index of method performance with respect to precision. *Journal of AOAC International*, 89, 4, 2006, 1095-1108.

ผศ.ศรวิชิต รุตประจงค์ & รศ.ประภาศิริ ภูวนเทียน

37

## Practice: testing for homogeneity of test materials



# การทดสอบความเสถียร ของ PT sample

วิทยากร: ผศ.ดร.ครรชิต จุดประสงค์



## Part II.

# Statistical Evaluation: testing stability of a test material

*Asst. Prof. Kunchit Judprasong\**

*Assoc. Prof. Prapasri Puwastien\**

*Chantarat Vorasapavit\*\**

\*Institute of Nutrition, Mahidol University, Salaya, Putthamonthon 4,  
Nakhon Pathom 73170, THAILAND  
Tel: 0 2800 2380 ext . 324

\*\* Department of Science and Service

[kunchit.jud@mahidol.ac.th](mailto:kunchit.jud@mahidol.ac.th)

Statistical approach for PT (Testing and Calibration)  
(APLAC-PTB approved work plan)  
28-30 พฤษภาคม 2555

ผศ.ดร.ชิต จุดประดงค์ & รศ.ดร.ประภาศรี ภูวเลศิย

1

## STABILITY TESTING

### a. Classical stability study: i.e., vitamins

- ❖ Store samples at storage conditions which represent conditions of the entire process of transportation and storage
- ❖ Re-analyse the suspected component in 5 randomly single samples at specified storage period
- ❖ If the obtained values are in the range of mean  $\pm$  2SD derived from the homogeneity study or at 0 day storage; the component is stable

### b. Isochronous stability study: i.e., minerals

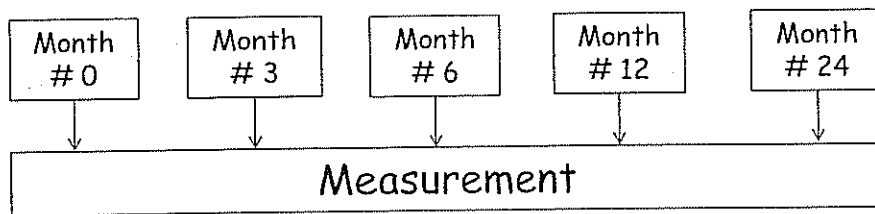
- The sub-samples collected at different periods and kept at  $-40^{\circ}\text{C}$  to  $-80^{\circ}\text{C}$
- The analytes in the sub-samples are measured at the same time  $\rightarrow$  reduce measurement variation  $\rightarrow$  smaller uncertainty

ผศ.ดร.ชิต จุดประดงค์ & รศ.ดร.ประภาศรี ภูวเลศิย

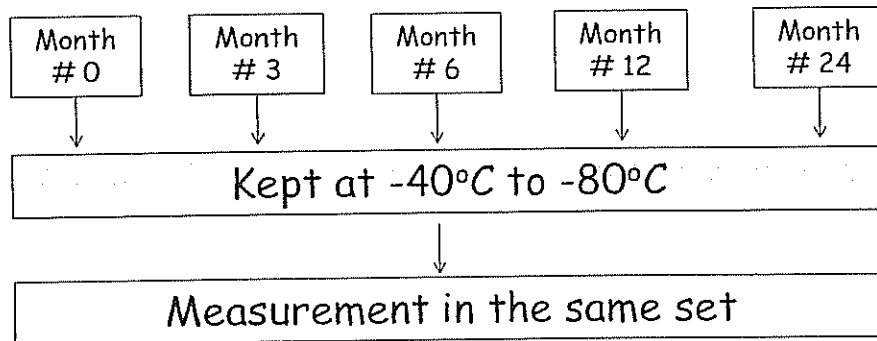
2



a. Classical stability study: i.e., vitamins



b. Isochronous stability study: i.e., minerals



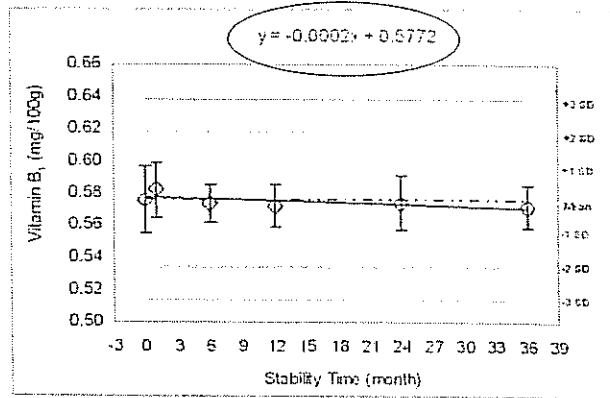
Criteria for evaluation of stability testing

- The stability values are in the range of mean  $\pm$  2SD derived from the homogeneity study or at 0 day storage.
- Slope (x coefficient) is expected to be zero.
- Compare general average value obtained from homogeneity study ( $\bar{y}$ ) and the average value obtained from stability results ( $\bar{x}$ ) [Ref: ISO 13528 (2005)]

$$|\bar{x} - \bar{y}| \leq 0.3\sigma$$

- Trend analysis: as slope/SD(slope), which should be  $\leq t_{cri}$  [Ref: ISO Guide 35 (2006)]

# Various types of stability evaluation



Stability testing: vitamin B<sub>1</sub> in Rice

1. The values were within mean  $\pm$  SD of the levels at 0 day
2. Slope (x coefficient) is expected to be zero and intercept is similar to mean of day 0.



Vitamin B<sub>1</sub> is stable

## Various types of stability evaluation (continued)

3. ISO 13528:  $|\bar{X}_{\text{hom}} - \bar{Y}_{\text{sta}}| < 0.3\sigma$

where  $\bar{X}_{\text{hom}}$  = mean value from homogeneity of test material

$\bar{Y}_{\text{sta}}$  = mean value from stability of each storage time

$\sigma$  = Horwitz's predicted standard deviation

Storage period Month	Mean value of B1 in 5 replicates at each period					Mean (y)	SD	Mean at 0 mo.(y)	ABS Diff. (x-y)	Conclusion**
	1	2	3	4	5					
0	0.56	0.56	0.57	0.61	0.58	0.58	0.02	0.58	-	
1	0.58	0.57	0.58	0.61	0.57	0.58	0.02	0.58	0.006	Pass
6	0.59	0.57	0.56	0.57	0.58	0.57	0.01	0.58	0.002	Pass
12	0.58	0.56	0.59	0.57	0.56	0.57	0.01	0.58	0.004	Pass
24	0.56	0.56	0.60	0.57	0.58	0.57	0.02	0.58	0.002	Pass
36	0.58	0.56	0.59	0.57	0.56	0.57	0.01	0.58	0.004	Pass

\*data from 5 replicates in homogeneity testing

\*\*Criteria based on ISO 13528:  $|\bar{X}_{\text{hom}} - \bar{Y}_{\text{sta}}| < 0.3\sigma$

Target SD =	0.071	(calculate based on Horwitz equation)
0.3 SDp	0.021	



Vitamin B<sub>1</sub> is stable throughout the storage period of three years.

4. ISO Guide 35 (2006): Trend analysis calculated as slope/SD slope

$$\frac{|b_1|}{s(b_1)} < t_{0.95, n-2}$$

where  $b_1$  = slope of stability data

$s(b_1)$  = standard deviation of slope

$t_{0.95}$  = critical t at 95% confidence limit

$n$  = number of stability data

Example:

stability of vitamin B <sub>1</sub> in rice	
slope	-0.00016
SD of slope	0.00021
df (n-2)	28
slope/SD(slope)	0.801
$t_{crit}$	2.048

slope/SD slope is  $< t_{crit}$  (2.085,  $p=0.05$ ),  $0.801 < 2.048$

= no statistically significant change of vitamin B<sub>1</sub> in the RM.



Vitamin B<sub>1</sub> is stable throughout the storage period of three years.

Trend analysis: ISO Guide 35 (2006) - formula

Linear least square

$$y = b_1x + b_0$$

Slope ( $b_1$ )

$$b_1 = \frac{\sum_j \{(x_j - \bar{x})(y_j - \bar{y})\}}{\sum_j (x_j - \bar{x})^2}$$

SD of slope ( $s_{b_1}$ )

$$s_{b_1} = \frac{s}{\sqrt{\sum_j (x_j - \bar{x})^2}}$$

$$s^2 = \frac{\sum_{i=1}^n (y_i - b_0 - b_1x_i)^2}{n - 2}$$

Trend analysis:  $ABS(\text{slope})/SD(\text{slope}) < t_{crit}$

$$\frac{|b_1|}{s(b_1)} < t_{0.95, n-2}$$





## Exercise 2.

# Testing stability of a measurand in a test material



✓

การหาค่ากำหนดของ PT sample  
การคำนวณค่า target standard deviation

วิทยากร: ผศ.ดร.ครรชิต จุดประสงค์  
นางจันทร์รัตน์ วรรณพวิทย์



## Part III

# Establishing assigned values of a test material

*Asst. Prof. Kunchit Judprasong\**

*Assoc. Prof. Prapasri Puwastien\**

*Chantarat Vorasapavit\*\**

\*Institute of Nutrition, Mahidol University, Salaya, Putthamonthon 4,  
Nakhon Pathom 73170, THAILAND  
Tel: 0 2800 2380 ext . 324

\*\* Department of Science and Service

[kunchit.jud@mahidol.ac.th](mailto:kunchit.jud@mahidol.ac.th)

Statistical approach for PT (Testing and Calibration)  
(APLAC-PTB approved work plan)  
28-30 พฤษภาคม 2555

ผศ.ดร.คชรชิต จุฑประสงค์ & รศ.ดร.ประภาศิริ ภูวเลณียร

1

## References

❖ ISO 13528: 2005. Statistical methods for use in proficiency testing by interlaboratory comparisons. Geneva, Switzerland.

❖ ISO/IEC 17043: 2010. Conformity assessment- General requirements for proficiency testing. Geneva, Switzerland.

ผศ.ดร.คชรชิต จุฑประสงค์ & รศ.ดร.ประภาศิริ ภูวเลณียร

2



## Establishing assigned value of test materials

The assigned value:

- shall be the responsibility of the coordinator
- shall not be disclosed to the participants until they have reported their results to the coordinator
- shall giving details in a report



## Establishing assigned values

- 1) Formulation: known value
- 2) Certified reference values
- 3) Reference values
- 4) Consensus from expert laboratories
- 5) Consensus from participants in laboratory performance study (laboratory proficiency study, PT)

## Establishing assigned values

- Consensus of "expert laboratories"
  - Reference laboratories
  - Recognized expert laboratories
  - Previous successful participating results
- Consensus of participants
  - Commonly used
  - Not recommended by IUPAC
  - Additional requirement in ISO/IEC 17043

### Establishing assigned values

#### 5) Consensus from participants in a round of PT scheme

Estimation of assigned values:

Algorithm A in Annex C of ISO 13528

Assigned value: Robust mean ( $x^*$ )  $\pm$  robust SD ( $s^*$ ),

Robust mean ( $x^*$ )  $\pm$  uncertainty ( $u_x$ )

Standard uncertainty 
$$u_x = \frac{1.25 \times s^*}{\sqrt{p}}$$

$s^*$  = the robust standard deviation of the results calculated using Algorithm A in Annex C.  
 $p$  = number of participants report a measurement

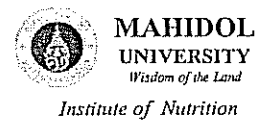
Limitations:

- there may be no real consensus amongst the participants;
- the consensus may be biased by the general use of faulty methodology and this bias will not be reflected in the standard uncertainty of the assigned value

**Protein in Rice-1: Calculation of robust average and standard deviation (ISO 13528)**

Iteration	0	$x_i - x^*$	1	$(x_i - x^*)^2$	2	$(x_i - x^*)^2$	3	$(x_i - x^*)^2$	4	$(x_i - x^*)^2$	5	$(x_i - x^*)^2$	6	$(x_i - x^*)^2$
$\delta = 1.5 s^1$	---		0.34		0.38		0.38		0.38		0.38		0.38	
$x^* - \delta$	---		7.70		7.70		7.70		7.70		7.70		7.70	
$x^* + \delta$	---		8.39		8.45		8.45		8.45		8.45		8.45	
21	6.85	1.20	7.70	0.14	7.70	0.14	7.70	0.14	7.70	0.14	7.70	0.14	7.70	0.14
34	7.55	0.50	7.70	0.14	7.70	0.14	7.70	0.14	7.70	0.14	7.70	0.14	7.70	0.14
53	7.62	0.43	7.70	0.14	7.70	0.14	7.70	0.14	7.70	0.14	7.70	0.14	7.70	0.14
50	7.69	0.36	7.70	0.14	7.70	0.14	7.70	0.14	7.70	0.14	7.70	0.14	7.70	0.14
5	7.77	0.28	7.77	0.09	7.77	0.09	7.77	0.09	7.77	0.09	7.77	0.09	7.77	0.09
47	7.81	0.24	7.81	0.07	7.81	0.07	7.81	0.07	7.81	0.07	7.81	0.07	7.81	0.07
15	7.85	0.20	7.85	0.05	7.85	0.05	7.85	0.05	7.85	0.05	7.85	0.05	7.85	0.05
6	7.86	0.19	7.86	0.05	7.86	0.05	7.86	0.05	7.86	0.05	7.86	0.05	7.86	0.05
8	7.86	0.19	7.86	0.05	7.86	0.05	7.86	0.05	7.86	0.05	7.86	0.05	7.86	0.05
30	7.89	0.16	7.89	0.03	7.89	0.03	7.89	0.03	7.89	0.03	7.89	0.03	7.89	0.03
14	7.90	0.15	7.90	0.03	7.90	0.03	7.90	0.03	7.90	0.03	7.90	0.03	7.90	0.03
23	7.91	0.14	7.91	0.03	7.91	0.03	7.91	0.03	7.91	0.03	7.91	0.03	7.91	0.03
42	7.91	0.14	7.91	0.03	7.91	0.03	7.91	0.03	7.91	0.03	7.91	0.03	7.91	0.03
26	7.92	0.13	7.92	0.03	7.92	0.03	7.92	0.03	7.92	0.03	7.92	0.03	7.92	0.03
51	7.92	0.13	7.92	0.02	7.92	0.02	7.92	0.02	7.92	0.02	7.92	0.02	7.92	0.02
27	7.94	0.11	7.94	0.02	7.94	0.02	7.94	0.02	7.94	0.02	7.94	0.02	7.94	0.02
24	8.14	0.10	8.14	0.00	8.14	0.00	8.14	0.00	8.14	0.00	8.14	0.00	8.14	0.00
46	8.16	0.12	8.16	0.01	8.16	0.01	8.16	0.01	8.16	0.01	8.16	0.01	8.16	0.01
55	8.20	0.15	8.20	0.01	8.20	0.01	8.20	0.01	8.20	0.01	8.20	0.01	8.20	0.01
61	8.22	0.18	8.22	0.02	8.22	0.02	8.22	0.02	8.22	0.02	8.22	0.02	8.22	0.02
56	8.24	0.20	8.24	0.03	8.24	0.03	8.24	0.03	8.24	0.03	8.24	0.03	8.24	0.03
25	8.28	0.23	8.28	0.04	8.28	0.04	8.28	0.04	8.28	0.04	8.28	0.04	8.28	0.04
16	8.28	0.23	8.28	0.04	8.28	0.04	8.28	0.04	8.28	0.04	8.28	0.04	8.28	0.04
36	8.34	0.29	8.34	0.07	8.34	0.07	8.34	0.07	8.34	0.07	8.34	0.07	8.34	0.07
45	8.55	0.50	8.39	0.10	8.39	0.10	8.39	0.10	8.39	0.10	8.39	0.10	8.39	0.10
32	8.57	0.53	8.39	0.10	8.39	0.10	8.39	0.10	8.39	0.10	8.39	0.10	8.39	0.10
9	8.58	0.54	8.39	0.10	8.39	0.10	8.39	0.10	8.39	0.10	8.39	0.10	8.39	0.10
52	8.59	0.54	8.39	0.10	8.39	0.10	8.39	0.10	8.39	0.10	8.39	0.10	8.39	0.10
39	8.15	1.11	8.39	0.10	8.39	0.10	8.39	0.10	8.39	0.10	8.39	0.10	8.39	0.10
18	9.27	1.23	8.39	0.10	8.39	0.10	8.39	0.10	8.39	0.10	8.39	0.10	8.39	0.10
Average	8.11		8.07	2.34	8.07	2.34	8.07	2.34	8.07	2.34	8.07	2.34	8.07	2.34
SD	0.39		0.22	0.05	0.22	0.05	0.22	0.05	0.22	0.05	0.22	0.05	0.22	0.05
New $x^*$	8.05	0.16	8.074	0.22	8.074	0.22	8.074	0.22	8.074	0.22	8.074	0.22	8.074	0.22
New $s^*$	0.23		0.250		0.250		0.250		0.250		0.250		0.250	

N = 49      Assigned value of protein in Rice-1 =  $8.07 \pm 0.25$  (robust mean and SD) g/100g  
 Original %RSD = 4.8      Final %RSD = 3.1 %



**Example of assigned values of test materials obtained from participants**

- Robust mean  $\pm$  robust SD, estimated according to ISO 13528
- If variation of data is too high (indicated by %RSD):  
 Robust mean\*  $\pm$  target SD (SDp, Horwitz's predicted SD)
- Median  $\pm$  Normalised IQR\*\*

\*estimated from ISO 13528 or use median value (not affected by extreme values) of the data obtained from participants

\*\* Normalised IQR = Inter Quartile Range x 0.7143



# Example of assigned values of test materials obtained from participants

P. Puwastien *et al* / *Journal of Food Composition and Analysis* 22 (2009) 453–462

457

**Table 2A**  
Assigned values of nutrients in rice, estimated by ISO 11525.

Measurand	Unit per 100 g	Robust mean	Robust SD	Expanded uncertainty	%RSD
Moisture: estimated from all submitted values	%	11.54	0.73	0.25	5.3
Moisture: estimated from data with empirical method	g	12.15	0.51	0.33	4.2
Protein	g	5.97	0.25	0.09	3.1
Dietary fibre	g	3.48	0.68	0.34	19.6*
Ash	g	1.41	0.05	0.02	3.6
Iron	mg	1.19	0.22	0.06	16.5*
Zinc	mg	1.72	0.21	0.09	12.2
Vitamin B <sub>1</sub>	mg	0.39	0.12	0.07	30.1*

\* Too high %RSD, robust mean = Horwitz predicted SD (SD<sub>p</sub>) was applied as shown in Table 2B.

**Table 2B**  
Assigned values of dietary fibre, iron and vitamin B<sub>1</sub> in rice as robust mean ± predicted SD of Horwitz.

Measurand	Unit per 100 g	Robust mean	Horwitz's Predicted SD	%RSD
Dietary fibre	g	3.48	0.12 (3SD <sub>p</sub> = 0.35)	10.0
Iron	mg	1.19	0.13	11.0
Vitamin B <sub>1</sub>	mg	0.39	0.05	13.0

Ref: Prapasri Puwastien, Kunchit Judprasong, Naruemon Pinprapai. Development of rice reference material and its use for evaluation of analytical performance of food analysis laboratories. *Journal of Food Composition and Analysis* 22 (2009) 453–462.

ผศ.ดร.นครนิต จุดประสงค์ & รศ.ดร.ประภาศรี กุลเตียร

9

## Exercise

Establishing assigned values  
of a test material

ผศ.ดร.นครนิต จุดประสงค์ & รศ.ดร.ประภาศรี กุลเตียร

10





# การหาค่ากำหนด (assigned value) ของ PT sample

จันทรัตน์ วรรณพวิทย์

กรมวิทยาศาสตร์บริการ

โทร. 022017137, 022017165, 0891233321

E-mail : chantarat@dss.go.th

จันทรัตน์ วรรณพวิทย์

1



## ขั้นตอนการหาค่ากำหนด

- เรียงข้อมูลจากน้อยไปหามาก
- คำนวณค่าเฉลี่ย และค่าส่วนเบี่ยงเบนมาตรฐาน
- คำนวณค่ามัธยฐานของชุดข้อมูล เป็นค่า **robust average (New  $x^*$ )**
- คำนวณค่าความแตกต่างระหว่างค่าข้อมูลแต่ละค่า กับค่า **robust average** โดยไม่ติดเครื่องหมาย

$$|x_i - \text{New } x^*|$$

จันทรัตน์ วรรณพวิทย์

2



## ขั้นตอนการหาค่ากำหนด

- คำนวณค่ามัธยฐานของความแตกต่าง ลงใน **column 2**
- คำนวณ **robust standard deviation (New  $s^*$ )**

$$New\ s^* = 1.483 \times \text{median} |x_i - New\ x^*|$$

- คำนวณค่า  $\text{delta}(\delta) = 1.5 \times New\ s^*$

- คำนวณค่า **cut-off**

$$\square\ x^* - \delta = x^* - 1.5s^*$$

$$\square\ x^* + \delta = x^* + 1.5s^*$$

จันทร์ฉวี วรสรรพวิทย์

3



## ขั้นตอนการหาค่ากำหนด

- ตัดออกข้อมูลเริ่มต้นจาก **column 1** ลงใน **column 3**
- พิจารณาข้อมูล  $X_i$  จากค่าต่ำสุดถึงค่าสูงสุด โดยเปลี่ยนตามค่า **cut-off** ดังนี้
  - ถ้า  $X_i < x^* - \delta$  ให้แทน  $X_i$  ด้วย  $x^* - \delta$
  - ถ้า  $X_i > x^* + \delta$  ให้แทน  $X_i$  ด้วย  $x^* + \delta$
- คำนวณค่าเฉลี่ยและค่าส่วนเบี่ยงเบนมาตรฐานใน **column 3**
- คำนวณค่า **robust average**

$$New\ x^* = \frac{\sum x_i}{p}$$

จันทร์ฉวี วรสรรพวิทย์

4



## ขั้นตอนการหาค่ากำหนด

- คำนวณค่า **robust standard deviation (New  $s^*$ )**

$$New\ s^* = 1.134 \sqrt{\frac{\sum (x_i - x^*)^2}{p-1}}$$

- คำนวณค่า  $delta(\delta) = 1.5 \times New\ s^*$

- คำนวณค่า **cut-off** ลงใน **column 4**

- $x^* - \delta = x^* - 1.5s^*$

- $x^* + \delta = x^* + 1.5s^*$

จันทรัตน์ วรรณวิทย์

5



## ขั้นตอนการหาค่ากำหนด

- ตัดลอกข้อมูลที่ปรับแก้แล้วจาก **column 3** ลงใน **column 4**

- พิจารณาข้อมูล  $x_i$  จากค่าต่ำสุดถึงค่าสูงสุด โดยเปลี่ยนตามค่า **cut-off** ดังนี้

- ถ้า  $x_i < x^* - \delta$  ให้แทน  $x_i$  ด้วย  $x^* - \delta$

- ถ้า  $x_i > x^* + \delta$  ให้แทน  $x_i$  ด้วย  $x^* + \delta$

- คำนวณค่าเฉลี่ย และค่าส่วนเบี่ยงเบนมาตรฐาน ใน **column 4**

- คำนวณค่า **robust average**

$$New\ x^* = \frac{\sum x_i}{p}$$

จันทรัตน์ วรรณวิทย์

6



## ขั้นตอนการหาค่ากำหนด

- คำนวณค่า **robust standard deviation (New  $s^*$ )**

$$New\ s^* = 1.134 \sqrt{\frac{\sum (x_i - x^*)^2}{p-1}}$$

- คำนวณค่า

$$\delta(\delta) = 1.5 \times New\ s^*$$

- คำนวณค่า **cut-off**

$$\square\ x^* - \delta = x^* - 1.5s^*$$

$$\square\ x^* + \delta = x^* + 1.5s^*$$



## ขั้นตอนการหาค่ากำหนด

- ทำซ้ำจนกระทั่งค่า **robust average ( $x^*$ )** และ **robust standard deviation ( $s^*$ )** ไม่เปลี่ยนแปลง

- จะได้ค่ากำหนดของ **PT sample = robust average ( $x^*$ )** และค่าส่วนเบี่ยงเบนมาตรฐานของ **PT sample** คือค่า **robust standard deviation ( $s^*$ )**

# Example 6 : การคำนวณค่าสถิติโรบัสต์

PTP Template

Iteration	0	$x_i - x^*$	1	$(x_i - x^*)^2$	2	$(x_i - x^*)^2$	
$\hat{\sigma} = 1.5 s^*$	---		0.29		0.33		
$x^* - \hat{\sigma}$	---		10.69		10.65		
$x^* + \hat{\sigma}$	---		11.27		11.32		
							Z Score
10	10.60	0.38	10.69	0.09	10.65	0.09	-1.74
8	10.69	0.29	10.69	0.09	10.69	0.09	-1.33
11	10.79	0.19	10.79	0.04	10.79	0.04	-0.89
7	10.60	0.18	10.80	0.04	10.80	0.04	-0.84
3	10.85	0.13	10.85	0.02	10.85	0.02	-0.62
15	10.95	0.03	10.95	0.00	10.95	0.00	-0.17
4	10.96	0.02	10.96	0.00	10.96	0.00	-0.12
9	10.98	0.00	10.98	0.00	10.98	0.00	-0.03
13	11.04	0.05	11.04	0.00	11.04	0.00	0.24
6	11.06	0.08	11.06	0.01	11.06	0.01	0.33
14	11.08	0.10	11.08	0.01	11.08	0.01	0.42
5	11.11	0.13	11.11	0.02	11.11	0.02	0.55
2	11.38	0.40	11.27	0.08	11.27	0.08	1.76
1	11.43	0.45	11.27	0.08	11.27	0.08	1.99
12	12.10	1.12	11.27	0.08	11.27	0.08	4.99
Average	11.05		10.99	0.54	10.99	0.54	
SD	0.37		0.20	0.04	0.20	0.04	
New $x^*$	10.98	0.13	10.99	0.20	10.99	0.20	
New $s^*$	0.19		0.22		0.22		
N	15						

$$z \text{ score} = \frac{x - x^*}{s^*}$$

Robust average = 10.99

Robust standard deviation = 0.22

Original %RSD	3.3
Final %RSD	2.0

Target value	10.99
Low Acceptable	10.32
High Acceptable	11.66

$x^*$   
 $= x^* - 3s^*$   
 $= x^* + 3s^*$



## การคำนวณค่าความไม่แน่นอนของค่ากำหนดของ PT sample

$$u(x) = 1.25 \times \frac{s^*}{\sqrt{p}} \implies u(x) \leq 0.3\hat{\sigma}$$

$$U(X) = 2 \times u(x), \quad k=2$$

เมื่อ  $s^*$  = robust standard deviation

$p$  = จำนวนชุดข้อมูล



## Limiting the uncertainty of the Assigned value (X) : 13528 section 4.2

If this cannot be met then  $u(x) > 0.3\hat{\sigma}$

Look for a better way to determine Assigned value

Incorporate uncertainty in score

▪ Z'

▪ En

▪ Zeta

$$z' = (x - X) / \sqrt{\hat{\sigma}^2 + u_x^2}$$

Advise participants of large uncertainty



## Limiting the uncertainty of the Assigned value (X)

$$u(x) = s / \sqrt{11} = 0.3s$$

If  $n \leq 11$  , then cannot meet criterion.

If robust average is used  $n > 1.25*(11)$   
or  $n \geq 14$





## การคำนวณค่าความไม่แน่นอนของค่ากำหนด ของ PT sample

When consensus mean and SD are used to determine performance :

$$\text{standard error for consensus value } u(x) = \frac{SD}{\sqrt{n}}$$

$$U(X) = 2 \times u(x), \quad k=2$$

So with large number of labs, one can have very small uncertainty.



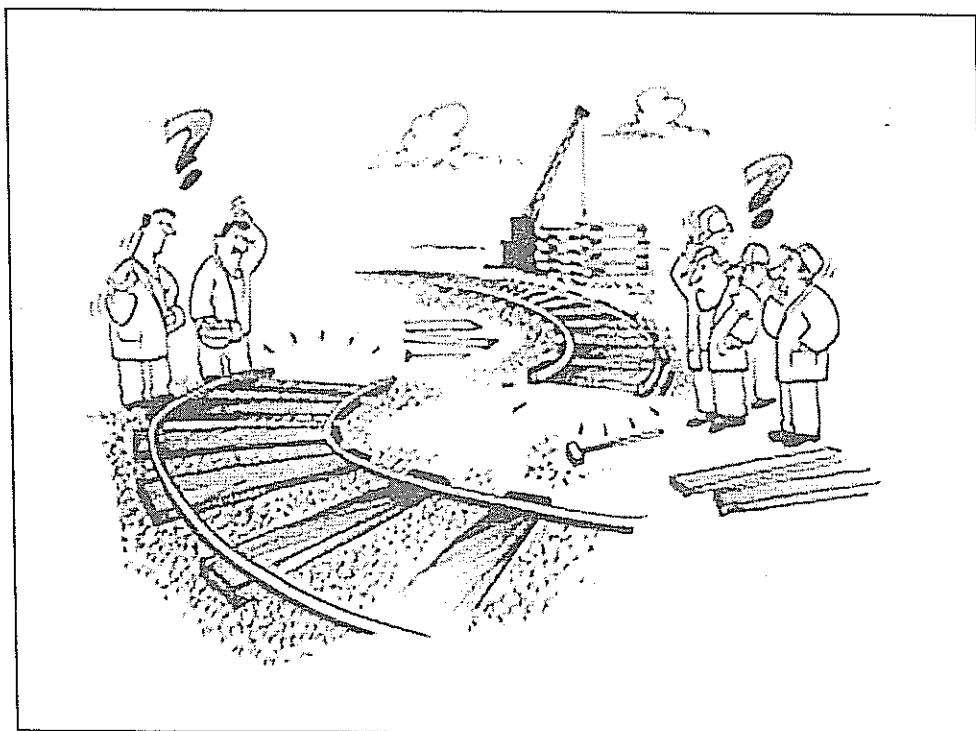
การคำนวณค่าความไม่แน่นอนของ PT sample  
ความสอบกลับได้ของค่ากำหนด

วิทยากร: ดร.พจมาน ทำจิ้น



# *Metrological Traceability & Measurement Uncertainty*

ดร.พอมาน ทำจิ้น

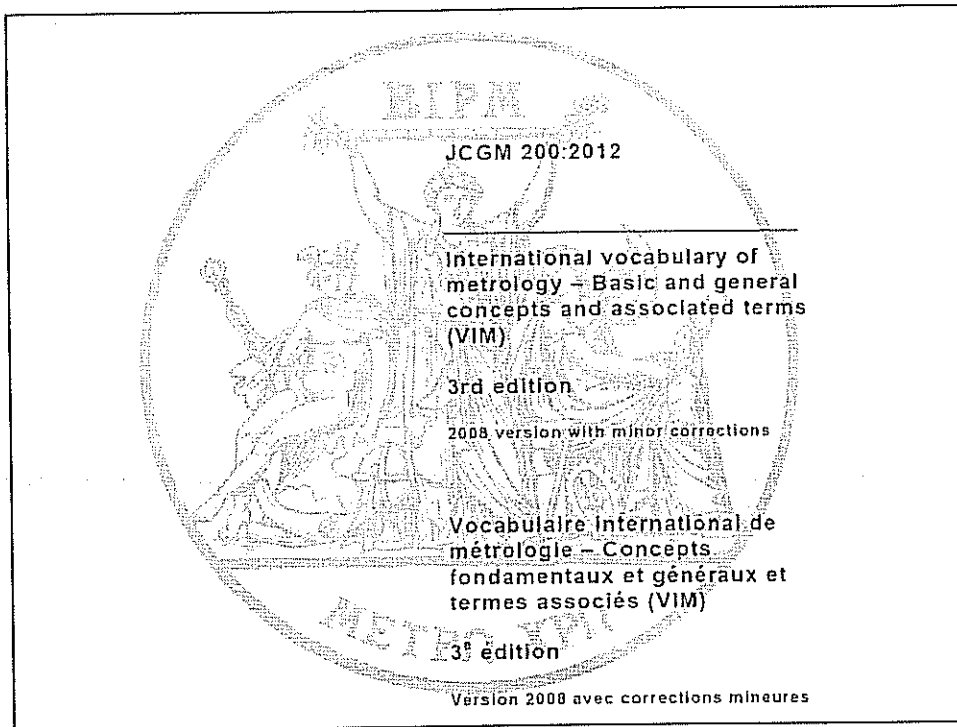


การอบรมเชิงปฏิบัติการ Statistical approach for PT (Testing and Calibration)

วันที่ ๒๘-๓๐ พฤษภาคม ๒๕๕๕

ณ ห้อง ๓๒๐ อาคารสถานศึกษาใหม่ปฏิบัติ กรมวิทยาศาสตร์บริการ

จัดโดย สำนักบริหารและรับรองห้องปฏิบัติการ



## 2.41 (6.10)

### **metrological traceability**

property of a **measurement result** whereby the result can be related to a reference through a documented unbroken chain of **calibrations**, each contributing to the **measurement uncertainty**

การยอมรับเชิงปฏิบัติการ Statistical approach for PT (Testing and Calibration)

วันที่ ๒๘-๓๐ พฤษภาคม ๒๕๕๕

ณ ห้อง ๔๒๐ อาคารสถานศึกษาเคมีปฏิบัติ กรมวิทยาศาสตร์บริการ

จัดโดย สำนักบริหารและรับรองห้องปฏิบัติการ

NOTE 1 For this definition, a 'reference' can be a definition of a **measurement unit** through its practical realization, or a **measurement procedure** including the measurement unit for a **non-ordinal quantity**, or a **measurement standard**.

NOTE 2 Metrological traceability requires an established **calibration hierarchy**.

NOTE 3 Specification of the reference must include the time at which this reference was used in establishing the calibration hierarchy, along with any other relevant metrological information about the reference, such as when the first calibration in the calibration hierarchy was performed.

NOTE 4 For **measurements with more than one input quantity in the measurement model**, each of the **input quantity values** should itself be metrologically traceable and the calibration hierarchy involved may form a branched structure or a network. The effort involved in establishing metrological traceability for each input quantity value should be commensurate with its relative contribution to the measurement result.

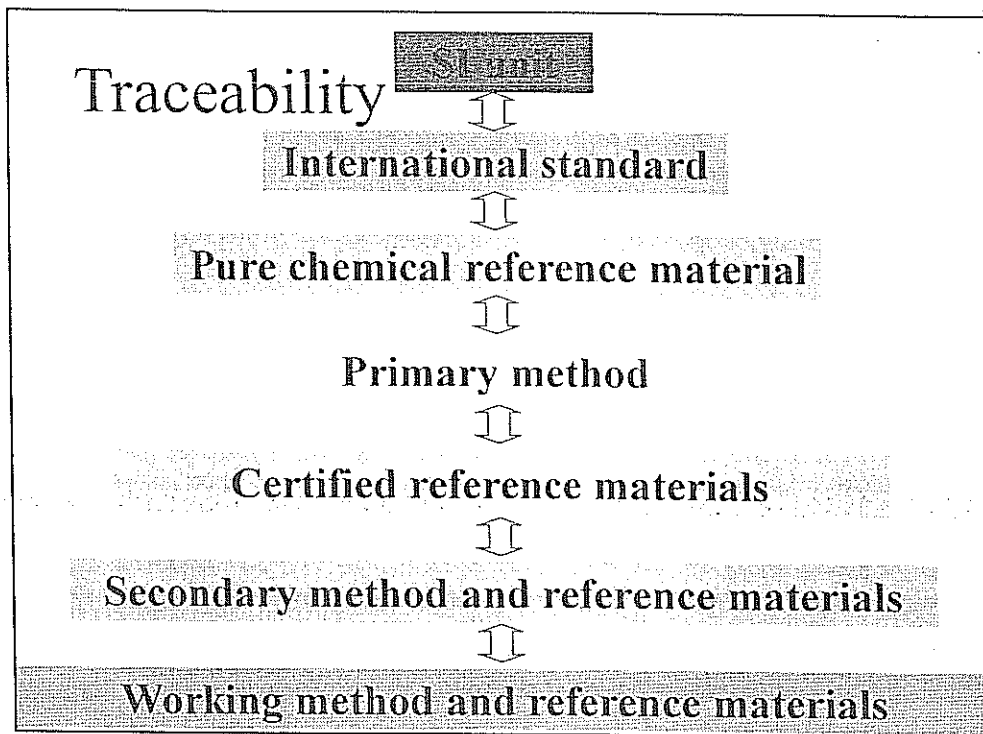
NOTE 5 Metrological traceability of a measurement result does not ensure that the measurement uncertainty is adequate for a given purpose or that there is an absence of mistakes.

NOTE 6 A comparison between two measurement standards may be viewed as a calibration if the comparison is used to check and, if necessary, correct the quantity value and measurement uncertainty attributed to one of the measurement standards.

NOTE 7 The ILAC considers the elements for confirming metrological traceability to be an unbroken **metrological traceability chain** to an **international measurement standard** or a **national measurement standard**, a documented measurement uncertainty, a documented measurement procedure, accredited technical competence, metrological traceability to the SI, and calibration intervals (see ILAC P-10:2002).

NOTE 8 The abbreviated term "traceability" is sometimes used to mean 'metrological traceability' as well as other concepts, such as 'sample traceability' or 'document traceability' or 'instrument traceability' or 'material traceability', where the history ("trace") of an item is meant. Therefore, the full term of "metrological traceability" is preferred if there is any risk of confusion.

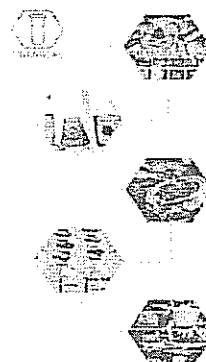


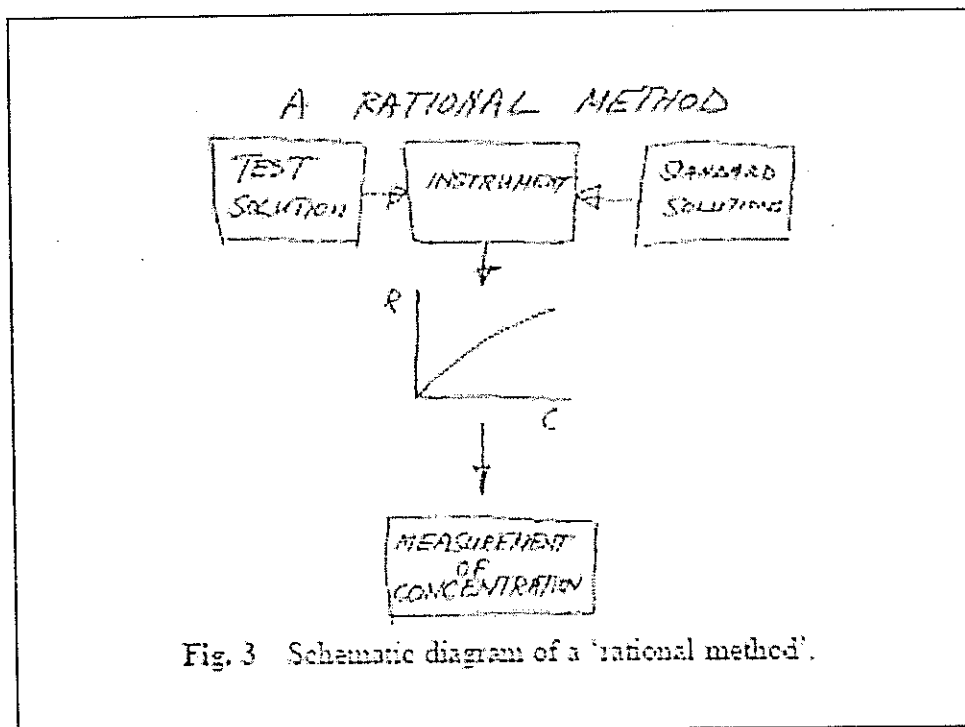
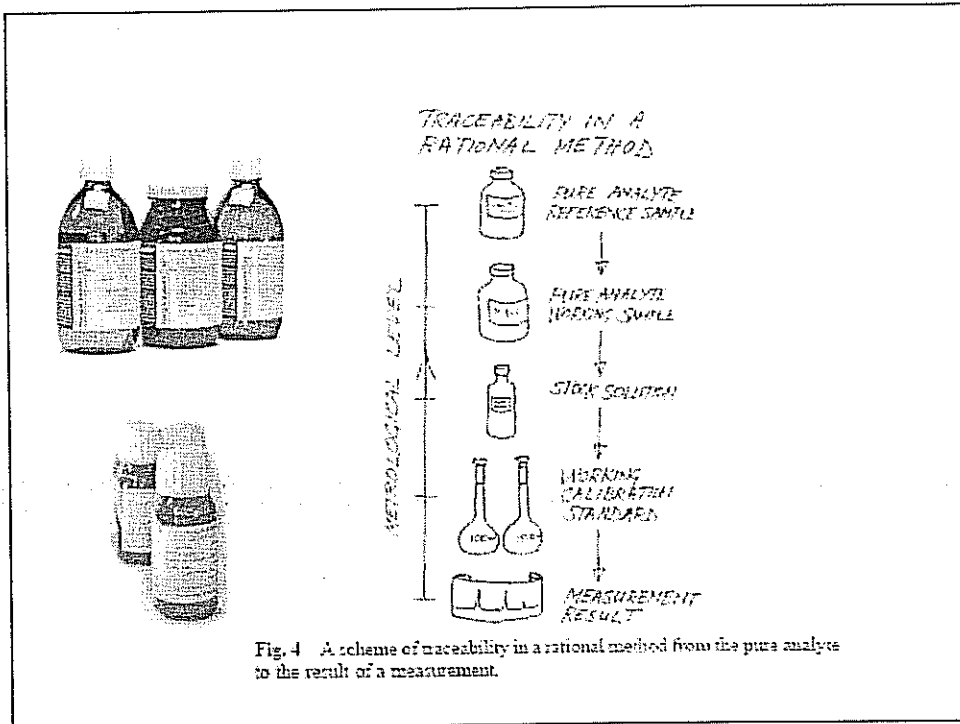


## Primary methods

### Examples

- Gravimetry
- Coulometry
- Titrimetry
- Freezing Point Depression
- Isotope Dilution MS



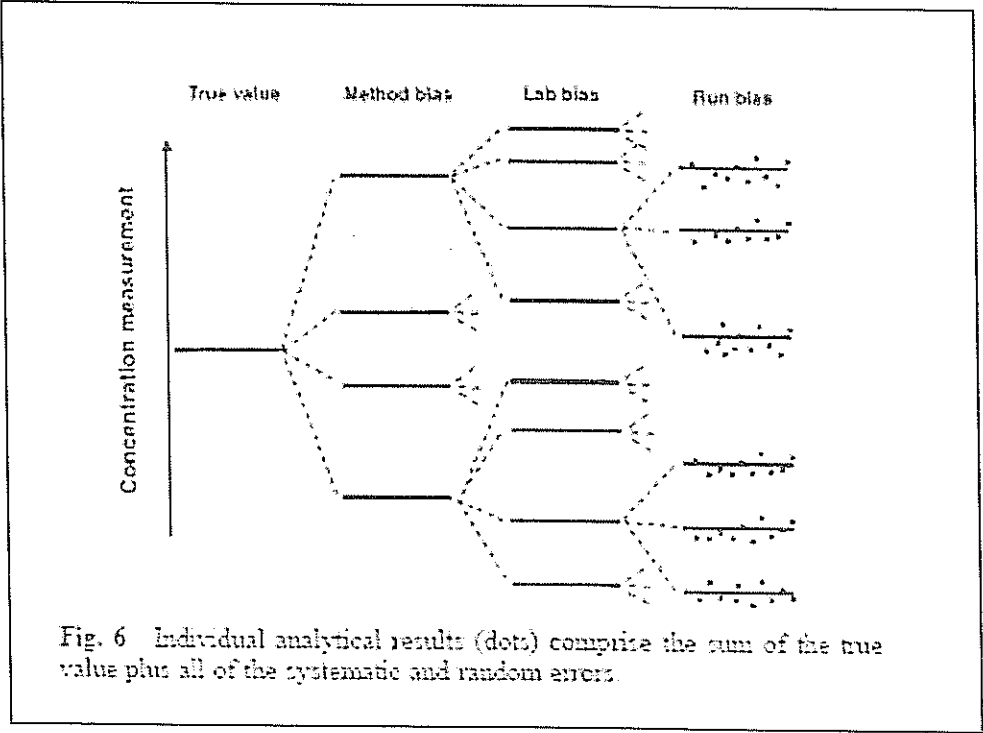
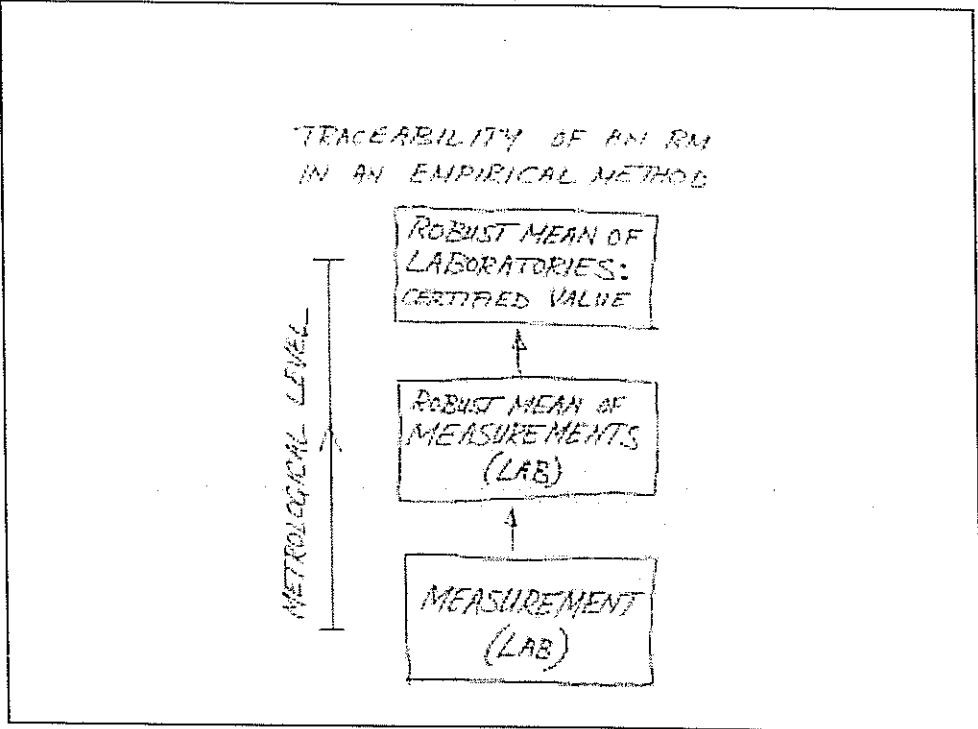


การอบรมเชิงปฏิบัติการ Statistical approach for PT (Testing and Calibration)

วันที่ ๒๘ -๓๐ พฤษภาคม ๒๕๕๕

ณ ห้อง ๓๒๐ อาคารสถานศึกษาเดิมปฏิบัติ กรมวิทยาศาสตร์บริการ

จัดโดย สำนักบริหารและรับรองห้องปฏิบัติการ



การอบรมเชิงปฏิบัติการ Statistical approach for PT (Testing and Calibration)

วันที่ ๒๘-๓๐ พฤษภาคม ๒๕๕๕

ณ ห้อง ๓๒๐ อาคารสถานศึกษาเดิมปฏิบัติ กรมวิทยาศาสตร์บริการ

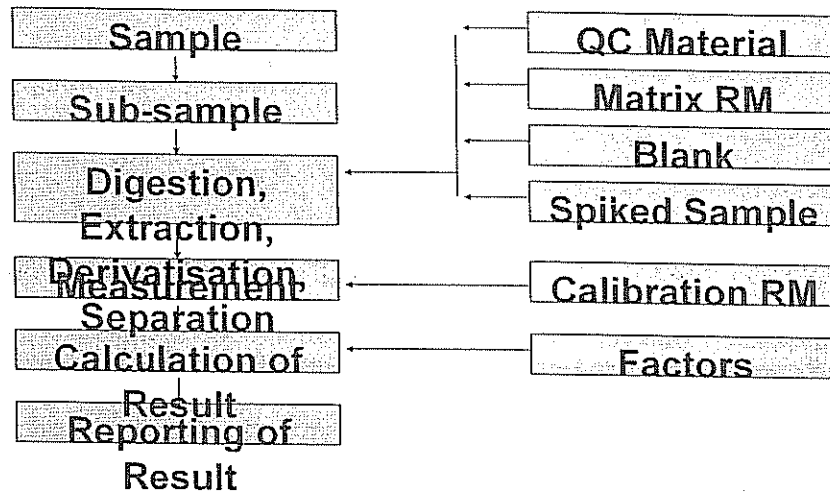
จัดโดย สำนักบริหารและรับรองห้องปฏิบัติการ

In summary, we have characterised most analytical methods currently in use as rational (traceable to the pure analyte) or empirical (traceable to the method). Fully rational methods provide the measurand (*i.e.*, the true value) within the bounds of precision uncertainty, regardless of the nature of the physical

## Establishing Traceability

- Traceability can be established by using a combination of:
  - traceable standards to calibrate the measuring equipment
  - pure substance reference materials
  - appropriate matrix certified reference material
  - an accepted, closely defined procedure
  - comparing results of a primary method

## Unbroken Chain of Comparisons: Within the Method



*Pure Appl. Chem.*, ASAP Article  
doi:10.1351/PAC-REP-07-09-39  
© 2011 IUPAC, Publication date (Web): 15 June 2011

### Metrological traceability of measurement results in chemistry: Concepts and implementation (IUPAC Technical Report)\*

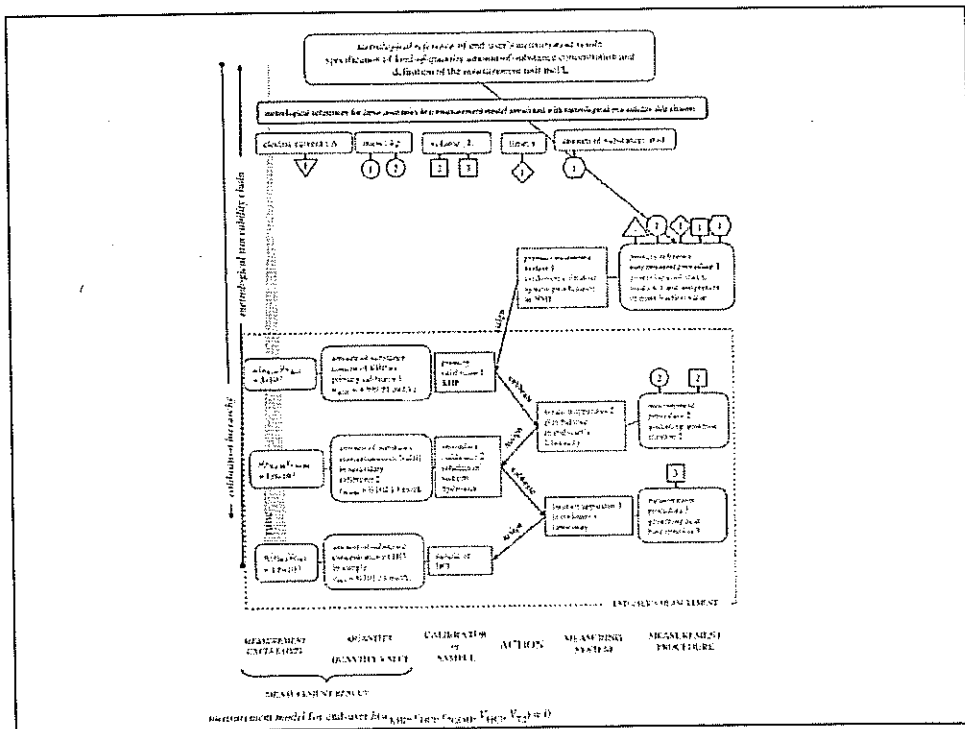
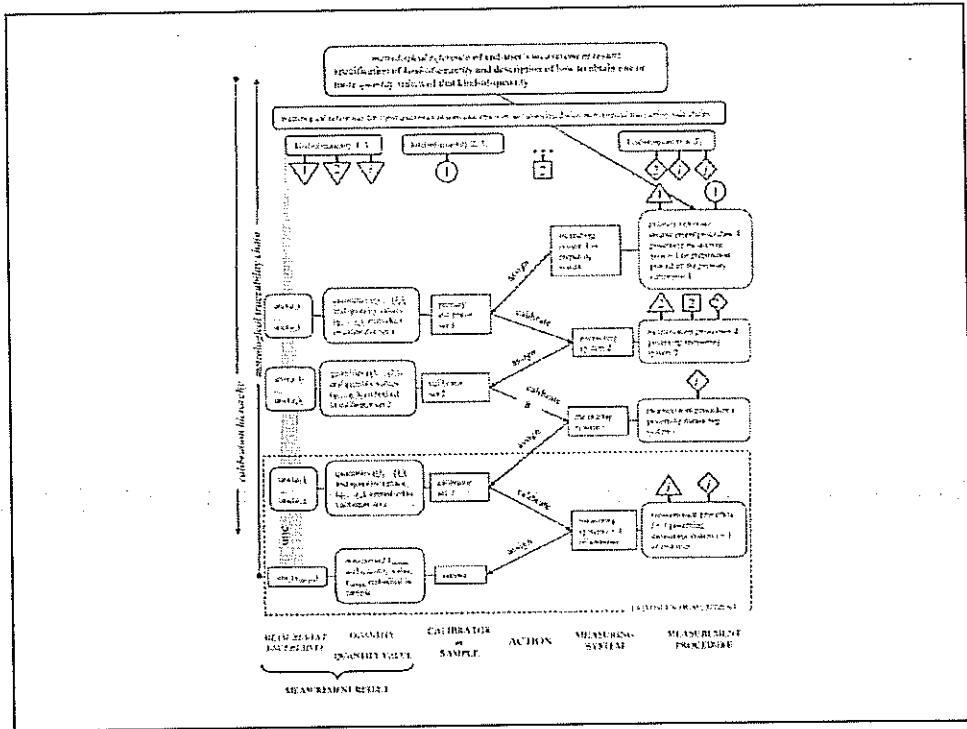
Paul De Bièvre<sup>1,†</sup>, René Dybkaer<sup>2</sup>, Aleš Fajgelj<sup>3</sup>, and  
D. Brynn Hibbert<sup>4</sup>

การอบรมเชิงปฏิบัติการ Statistical approach for PT (Testing and Calibration)

วันที่ ๒๙ -๓๐ พฤษภาคม ๒๕๕๕

ณ ห้อง ๓๒๐ อาคารสถานที่ศึกษาดูงานปฏิบัติ กรมวิทยาศาสตร์บริการ

จัดโดย สำนักบริหารและรับรองห้องปฏิบัติการ

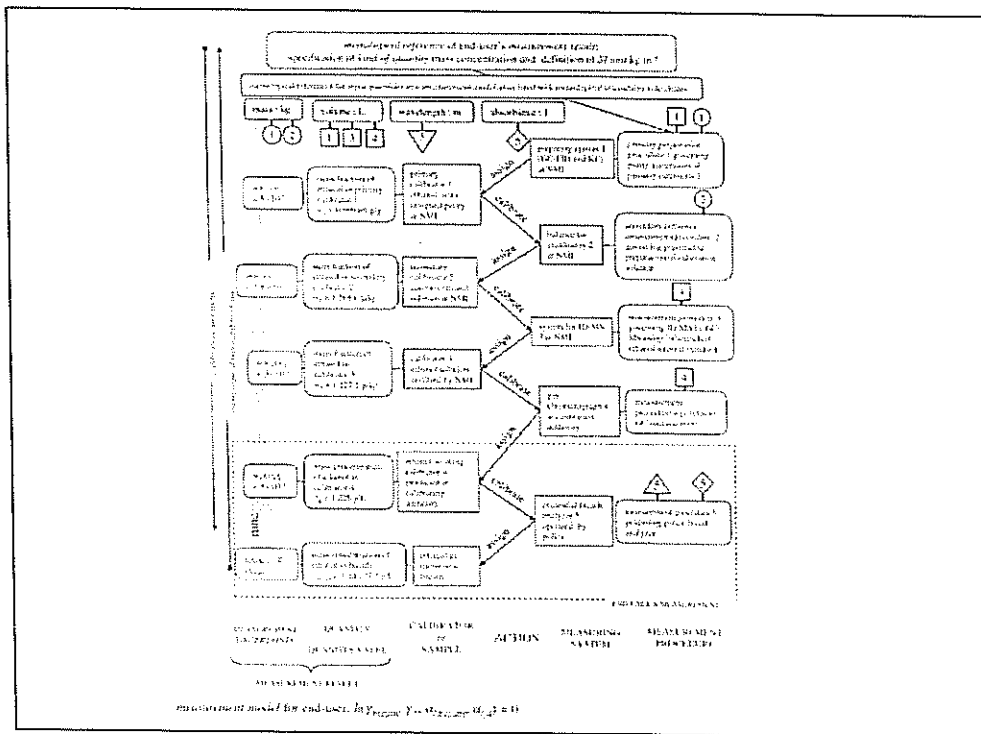
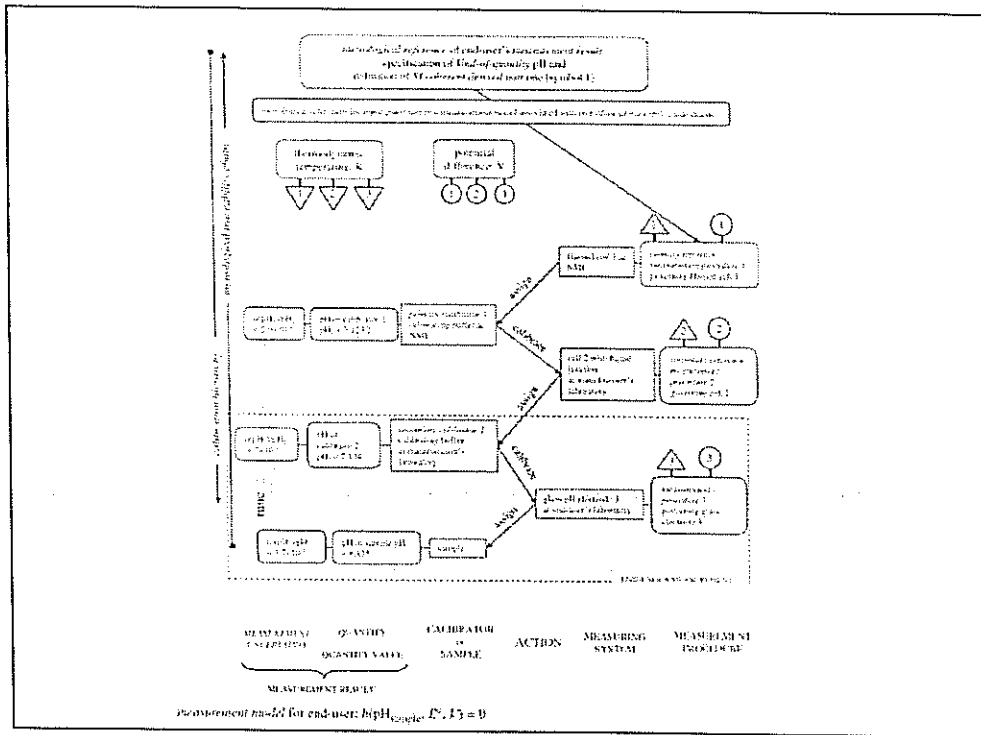


การอบรมเชิงปฏิบัติการ Statistical approach for PT (Testing and Calibration)

วันที่ ๒๘ -๓๐ พฤษภาคม ๒๕๕๕

ณ ห้อง ๓๒๐ อาคารสถานศึกษาเคมีปฏิบัติ กรมวิทยาศาสตร์บริการ

จัดโดย สำนักบริหารและรับรองห้องปฏิบัติการ



การอบรมเชิงปฏิบัติการ Statistical approach for PT (Testing and Calibration)

วันที่ ๒๘-๓๐ พฤษภาคม ๒๕๕๕

ณ ห้อง ๓๒๐ อาคารสถานศึกษาที่มีปฏิบัติ กรมวิทยาศาสตร์บริการ

จัดโดย สำนักบริหารและรับรองห้องปฏิบัติการ

## Metrological Traceability of Analytical Results

*In order for results to be comparable we need traceability.*

In this leaflet we would like to point out to analytical laboratories what traceability is and how a lab can demonstrate traceability of its results using an example of a mercury determination.

### Example - Mercury in tuna fish

A measurement result (mass fraction) of mercury in a tuna fish sample is  $4.03 \pm 0.11$  mg/kg. The result is reported as total Hg on dry weight basis (105 °C, 12 h determined on a separate sample portion) and the measurement uncertainty is reported with a 95 % level of confidence ( $k=2$ ).

*In this case mercury was determined with a mercury analyzer (LAS cold vapour) after a microwave digestion.*

*The samples are weighed on a balance with a calibration certificate relating the weight measured to the SI unit kg.*

*The acid digest is diluted in a volumetric flask where the manufacturer supplies the traceability of the volume of the flask to a national standard.*

*The calibration curve was made using a CRM (Certified Reference Material), a mercury standard with a certificate stating a quantity value of  $0.998 \pm 0.005$  mg/kg ( $k=2$ ) and with traceability to pure mercury.*

*The method is validated using an appropriate matrix CRM with a total mercury concentration of  $1.97 \pm 0.04$  mg/kg ( $k=2$ ). This validation is a check on the method performance.*



### Traceability shown by the laboratory

The evidence required by the laboratory to demonstrate traceability for the mercury result is shown below:

1. concentration of the Hg solution - a certificate of the CRM solution
2. mass of sample - calibration certificate of the balance
3. volume of volumetric flask - calibration certificate of the manufacturer
4. drying temperature - calibration of oven
5. digestion conditions - check according to specifications
6. drying time - ordinary clock or stopwatch

*Point 1 would need special attention to assure the quality and traceability for the calibration standard.*

*Traceability for points 2, 3 and 6 is easily achieved with adequate uncertainty using commercial equipment.*

*Points 4 and 5 need additional attention by the lab.*

The use of the matrix CRM in the method validation is vital but not part of the traceability, since it is not used for calibration. If this CRM is used for recovery correction it should be part of the traceability. However, the uncertainty associated with this matrix CRM may have to be included in the uncertainty budget.

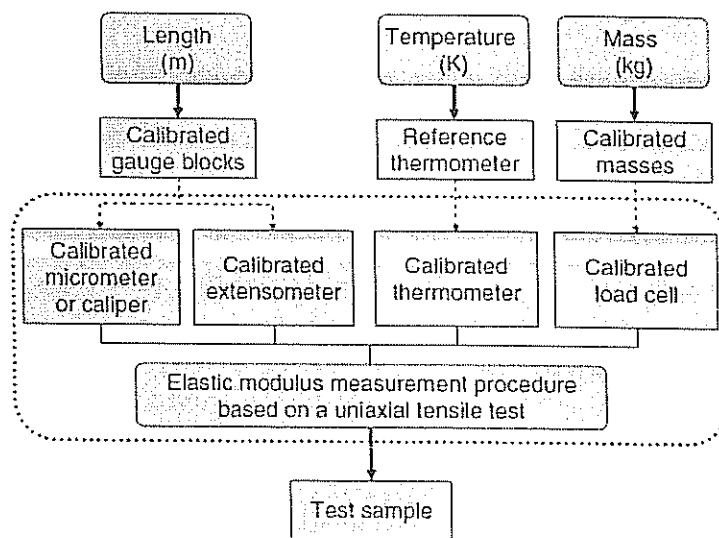
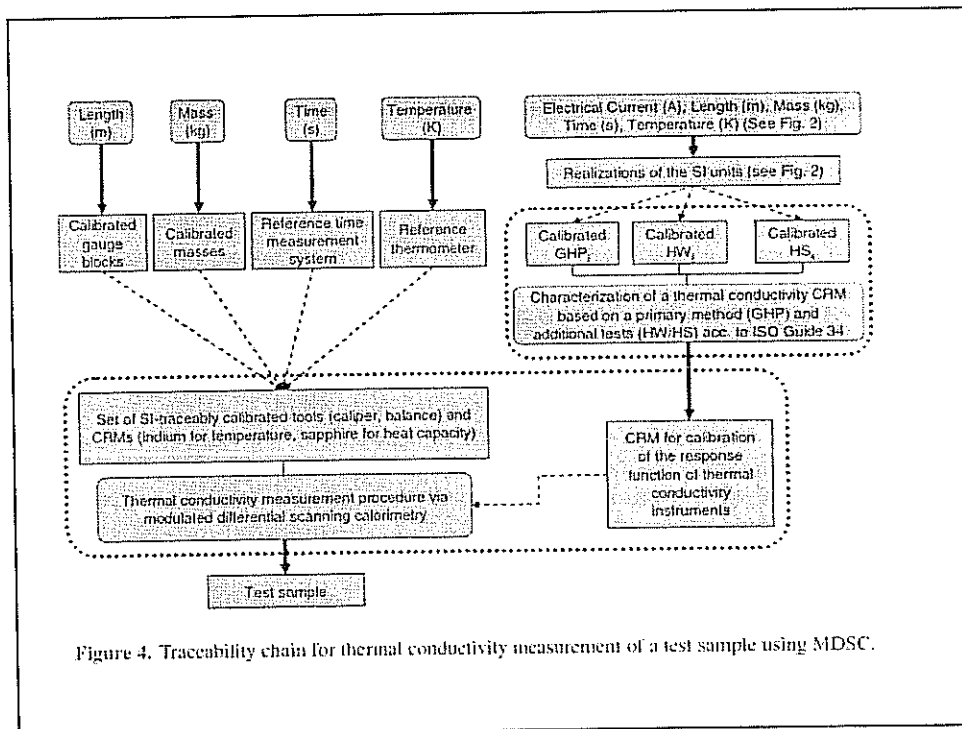
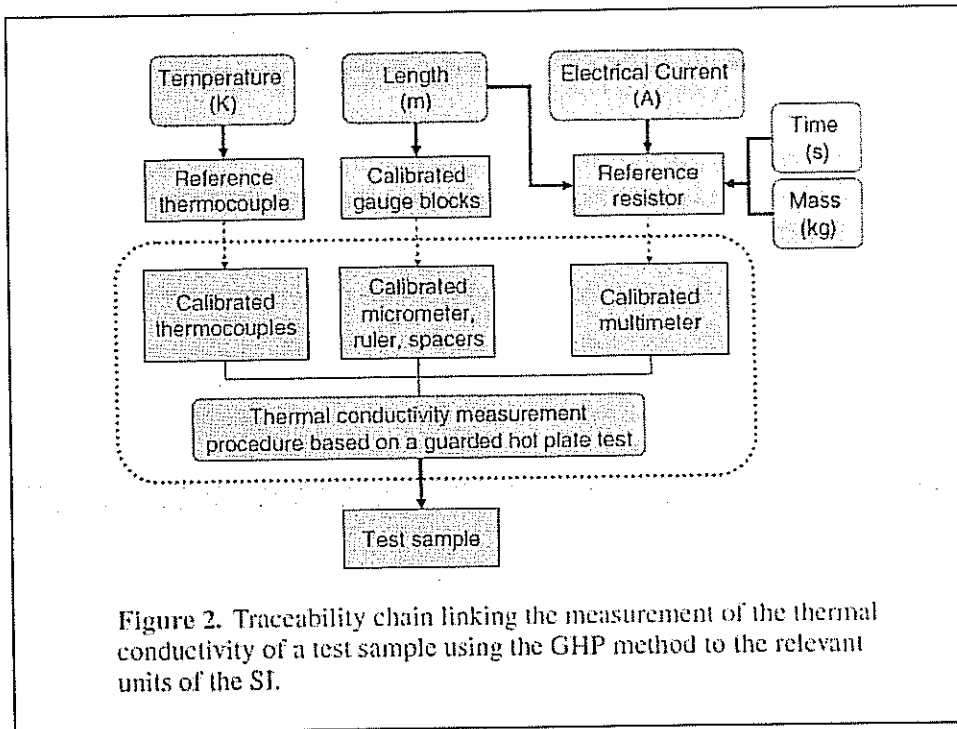


Figure 1. Traceability chain linking the measurement of the elastic modulus of a test sample using a uniaxial tensile test to the relevant units of the SI.



การอบรมเชิงปฏิบัติการ Statistical approach for PT (Testing and Calibration)

วันที่ ๒๘-๓๐ พฤษภาคม ๒๕๕๕

ณ ห้อง ๓๒๐ อาคารสถานศึกษาเคมีปฏิบัติ กรมวิทยาศาสตร์บริการ

จัดโดย สำนักบริหารและรับรองห้องปฏิบัติการ

**Eurachem** 

**CITAC**   
Co-Operation on International Traceability in Analytical Chemistry

**EURACHEM / CITAC Guide**

# **Traceability in Chemical Measurement**

**A guide to achieving comparable results  
in chemical measurement**

## **2. Cadmium Release from Ceramic Ware**

### **Specify the measurand and acceptable uncertainty**

The amount of released cadmium from ceramic ware is determined using atomic absorption spectrometry. The procedure employed is the empirical method BS 6748. The AAS spectrometer needs a detection limit below 0.02 mg/l Cadmium. The acceptable standard uncertainty for this empirical method is 10% (expressed as relative standard deviation).

### **Establish the procedure to determine the Cadmium release from ceramic ware**

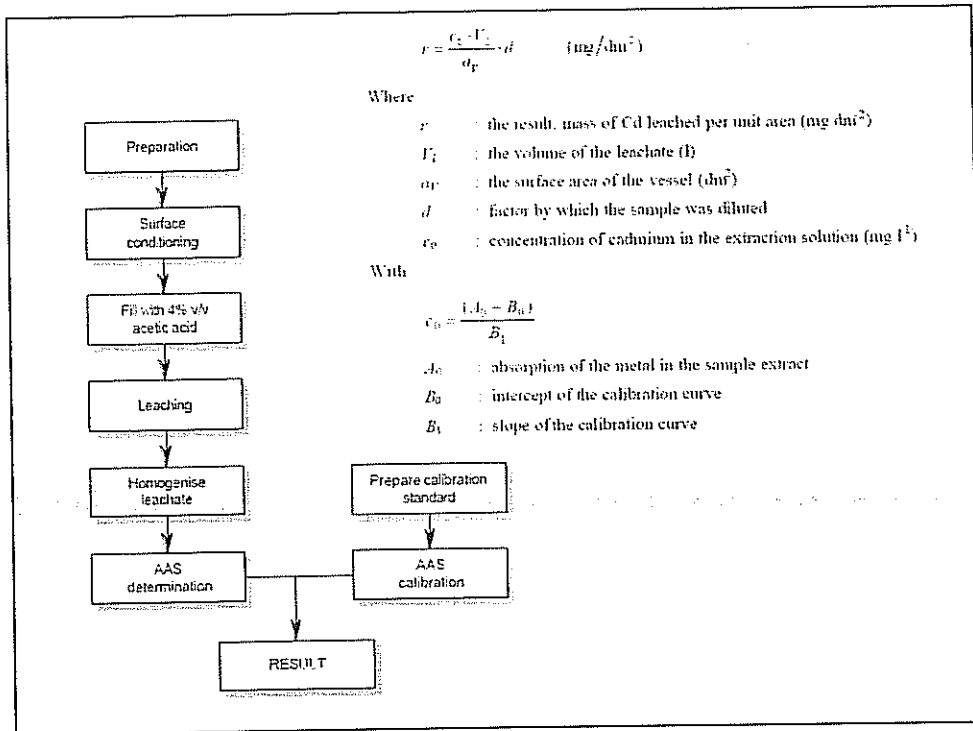
The complete procedure is given in British Standard BS 6748:1986 "Limits of metal release from ceramic ware, glass ware, glass ceramic ware and vitreous enamel ware" and this forms the specification for the measurand. Only a general description is given here.

การอบรมเชิงปฏิบัติการ Statistical approach for PT (Testing and Calibration)

วันที่ ๒๘ -๓๐ พฤษภาคม ๒๕๕๕

ณ ห้อง ๓๒๐ อาคารสถานที่ศึกษามีปฏิบัติ กรมวิทยาศาสตร์บริการ

จัดโดย สำนักบริหารและรับรองห้องปฏิบัติการ



### Choosing the references

The equation of the measurand has the seven parameters concentration of Cadmium in the extract solution, volume, area, dilution factor, correction factor for acid concentration, for time and for temperature. In order to establish traceability of the result is necessary to establish the traceability of these parameters with adequate uncertainty.

- **Concentration**  $c_0$ ,  $A_0$ ,  $B_0$  and  $B_1$  relate the concentration of the extraction solution, which has the largest contribution to the overall uncertainty, to the concentration of the calibration solutions, establishing traceability to the calibration solutions. These calibration solutions were obtained by diluting the reference solution of (500 ± 1) mg l<sup>-1</sup> cadmium in 4% v/v acetic acid. The reference solution is traceable to NIST SRM solution according to the certificate of the manufacturer. NIST has shown its measurement capabilities for determining the concentration of cadmium in solution in a key comparison at CIPM. The dilution steps were done using volumetric glassware, whose manufacturer specifies the value of the volume and its tolerance. Details were also available about the procedure used to establish traceability to the SI. The calibration solutions were measured using atomic absorption spectrometry and then the absorption values and the concentration of the calibration solution were employed to calculate the intercept ( $B_0$ ) and slope ( $B_1$ ) of the calibration curve by least square linear regression. These activities achieve traceability for  $c_0$ .

การอบรมเชิงปฏิบัติการ Statistical approach for PT (Testing and Calibration)

วันที่ ๒๘-๓๐ พฤษภาคม ๒๕๕๕

ณ ห้อง ๓๒๐ อาคารสถานศึกษาเคมีปฏิบัติ กรมวิทยาศาสตร์บริการ

จัดโดย สำนักบริหารและรับรองห้องปฏิบัติการ

- $V_L$  is the volume of the leachate. It is measured using a measuring cylinder. The volume determined by the measuring cylinder is adequately controlled by manufacturing tolerances according to the glassware standards referenced, so purchase from a reputable source according to specification is sufficient. As usual, however, a brief check on laboratory glassware on receipt, if only against similar equipment, is prudent to guard against the occasional, if rare, manufacturing error.
- The length measurement is done by placing a mark on the vessel employing a ruler to check the distance of 1 or 6 mm. This is not a critical measurement, so specific calibration of the ruler is unnecessary. As a matter of ordinary prudence, however, the ruler was checked on initial receipt in the laboratory using a calibrated vernier caliper (available for other measurements).
- $a_V$  is the surface area of the vessel. It is measured using a ruler checked as above.
- $d$  is a factor by which the sample was diluted. It is not used in this determination, therefore no traceability statement is needed.
- *Acid concentration.* The British Standard BS 6748:1986 specifies the values for the acid concentration, based on glacial acetic acid of stated purity. Because the influence of changes in acid concentration is small (the uncertainty estimate is based on the manufacturer's purity range), no further measures are required for sufficient traceability to the SI.

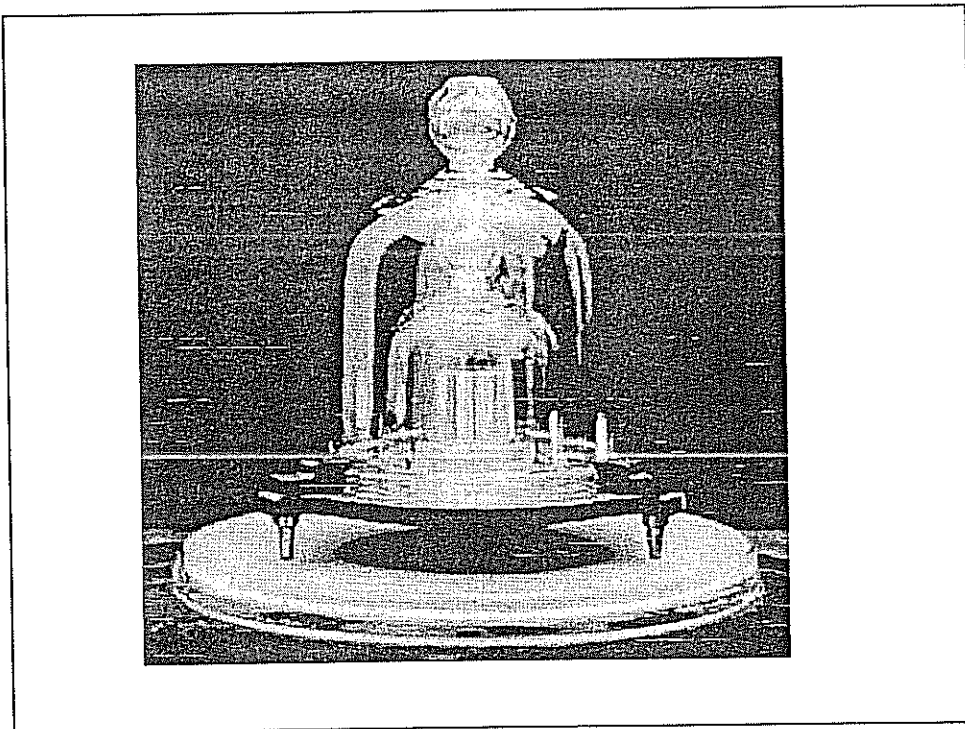
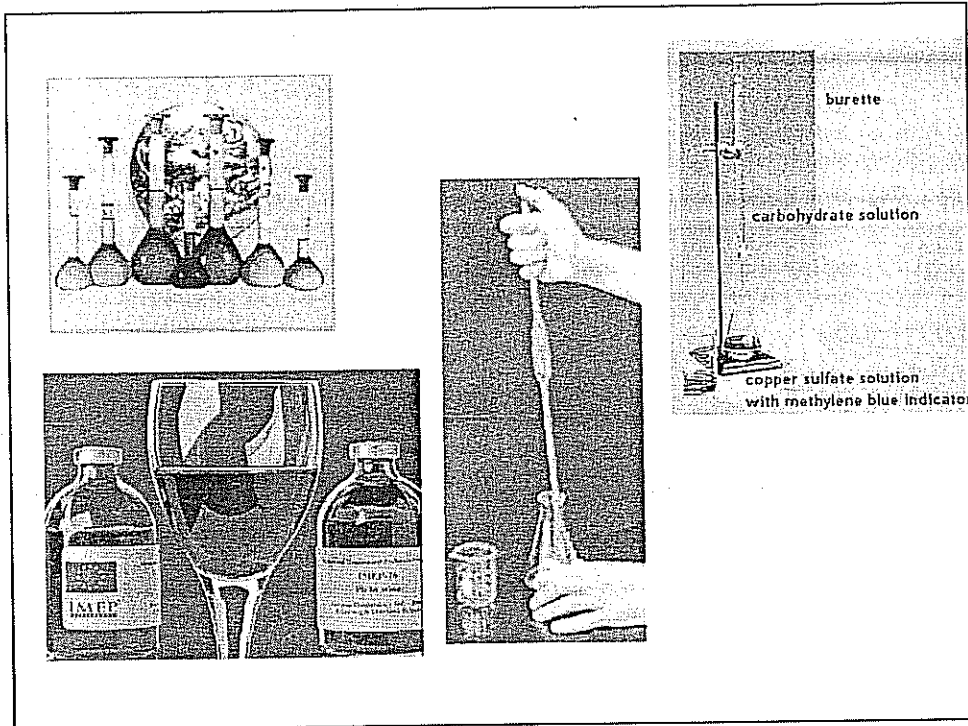
- *Time.* The duration of the leaching process is specified in the BS and has to be controlled. Because the time has such a minor influence on the combined standard uncertainty, it is sufficient to control the duration with a normal laboratory clock, which need only be checked occasionally against an appropriate time signal.
- *Temperature.* BS 6748:1986 quotes a temperature range of  $22 \pm 2$  °C. Because the temperature influence is the second largest contribution to the overall uncertainty, it is necessary to control and measure it with a thermometer, which is checked and calibrated by the manufacturer every two years. The manufacturer has accreditation to perform this calibration service. In this way traceability to the SI is established.

การอบรมเชิงปฏิบัติการ Statistical approach for PT (Testing and Calibration)

วันที่ ๒๘-๓๐ พฤษภาคม ๒๕๕๕

ณ ห้อง ๓๒๐ อาคารสถานศึกษาเคมีปฏิบัติ กรมวิทยาศาสตร์บริการ

จัดโดย สำนักบริหารและรับรองห้องปฏิบัติการ



การยอมรับเชิงปฏิบัติการ Statistical approach for PT (Testing and Calibration)

วันที่ ๒๘ -๓๐ พฤษภาคม ๒๕๕๕

ณ ห้อง ๓๒๐ อาคารสถานศึกษาที่มีปฏิบัติ กรมวิทยาศาสตร์บริการ

จัดโดย สำนักบริหารและรับรองห้องปฏิบัติการ

## Standard uncertainty of the assigned value (ตาม ISO 13528)

### 5 Determination of the assigned value and its standard uncertainty

#### 5.1 Choice of method of determining the assigned value

#### 5.2 Formulation [see ISO/IEC Guide 43-1:1997, A.1.1, item a)]

##### 5.2.2 Standard uncertainty $u_X$ of the assigned value

When the assigned value is calculated from the formulation of the test material, the standard uncertainty is estimated by combination of uncertainties using the approach described in the *Guide to the expression of uncertainty in measurement*. For example, in chemical analyses the uncertainties will usually be those associated with gravimetric and volumetric measurements.

The limitation of this method (in chemical analysis) is that care is needed to ensure that:

- the base material is effectively free from the added constituent, or that the proportion of the added constituent in the base material is accurately known;
- the constituents are mixed together homogeneously (where this is required);
- all sources of error are identified (e.g. it is not always realized that glass absorbs mercury compounds, so that the concentration of an aqueous solution of a mercury compound can be altered by its container);
- there is no interaction between the constituents and the matrix.

The *Guide to the expression of uncertainty in measurement* gives guidance on the evaluation of measurement uncertainties.

### 5.3 Certified reference values [see ISO/IEC Guide 43-1:1997, A.1.1 item b)]

#### 5.3.1 General

When the material used in a proficiency test is a certified reference material (CRM), its certified reference value is used as the assigned value  $X$ .

#### 5.3.2 Standard uncertainty $u_X$ of the assigned value

When a certified reference material is used as the test material, the standard uncertainty of the assigned value is derived from the information on uncertainty provided on the certificate.

The limitation of this approach is that it can be expensive to provide every participant in a proficiency test with a sample of a certified reference material.

การอบรมเชิงปฏิบัติการ Statistical approach for PT (Testing and Calibration)

วันที่ ๒๘-๓๐ พฤษภาคม ๒๕๕๕

ณ ห้อง ๓๒๐ อาคารสถานศึกษาเฉลิมปฏิบัติ กรมวิทยาศาสตร์บริการ

จัดโดย สำนักบริหารและรับรองห้องปฏิบัติการ

#### 5.4 Reference values [see ISO/IEC Guide 43-1:1997, A.1.1 item c)]

##### 5.4.1 General

In this approach, samples of the test material that is to be the reference material (RM) are prepared first, ready for distribution to the participants. A number of the samples are then selected at random and tested along with certified reference materials, in one laboratory, using a suitable measurement method, and under repeatability conditions (as defined in ISO 3534-2). The assigned value  $X_{RM,1}$  of the test material is then derived from a calibration against the certified reference values of the CRMs.

##### 5.4.2 Standard uncertainty $u_X$ of the assigned value

When the assigned value of a test material is derived from the results of a series of tests on that material and on CRM, the standard uncertainty of the assigned value is derived from the test results, and the uncertainties of the certified reference values of the CRM. If the test material and the CRM are not similar (in matrix, composition and level of results), then the uncertainty arising from this is also to be included.

This method allows the assigned value to be established in a manner that is traceable to the certified values of the CRMs, with a standard uncertainty that can be calculated, and avoids the cost of distributing the CRM to all the participants. These are good reasons for preferring it to other methods. However, the method assumes that there are no interactions between the materials used and the test conditions.

The example in 5.4.3 illustrates how the required uncertainty may be calculated in the simple case when the assigned value of a test material is established by direct comparison with a single CRM.

The CRM described in the example in 5.3 may be used to determine the assigned value for an RM which is another, similar, aggregate. This determination requires a series of tests to be carried out, in one laboratory, on samples of the two aggregates, using the same measurement method, and under repeatability conditions. Let

$X_{CRM}$  is the assigned value for the CRM

$X_{RM}$  is the assigned value for the RM

$D_i$  is the difference (RM - CRM) between the average results for the RM and the CRM on the  $i^{\text{th}}$  samples

$\bar{D}$  is the average of the differences  $D_i$

Then

$$X_{RM} = X_{CRM} + \bar{D} \quad (3)$$

The standard uncertainty of the assigned value of the RM may be calculated as:

$$u_{X_{RM}} = \sqrt{u_{X_{CRM}}^2 + u_{\bar{D}}^2} \quad (4)$$

Table 1 gives an example of data that might be obtained in such a series of tests, and shows how the standard uncertainty  $u_{\bar{D}}$  of the difference is calculated.

With these results,

$$X_{RM} = 21,62 + 1,73 = 23,35 \text{ LA units} \quad (5)$$

$$u_{X_{RM}} = \sqrt{0,26^2 + 0,24^2} = 0,35 \text{ LA units} \quad (6)$$

where 0,26 is the standard uncertainty of the assigned value of the CRM (given in the example in 5.3), and 0,24 is the standard uncertainty of  $\bar{D}$ .

การอบรมเชิงปฏิบัติการ Statistical approach for PT (Testing and Calibration)

วันที่ ๒๘-๓๐ พฤษภาคม ๒๕๕๕

ณ ห้อง ๓๒๐ อาคารสถานศึกษาคณะวิทยาศาสตร์ ธรรมวิทย์มหาวิทยาลัยศรีนครินทรวิโรฒ

จัดโดย สำนักบริหารและรับรองห้องปฏิบัติการ



Table 1 — Calculation of the average difference between a CRM and a RM, and of the standard uncertainty of this difference

Sample	RM		CRM		Difference in average values RM - CRM LA units
	Test 1 LA units	Test 2 LA units	Test 1 LA units	Test 2 LA units	
1	20,5	20,5	19,0	18,0	2,00
2	21,1	20,7	19,8	19,9	1,05
3	21,5	21,5	21,0	21,0	0,50
4	22,3	21,7	21,0	20,8	1,10
5	22,7	22,3	20,5	21,0	1,75
6	23,6	22,4	20,3	20,3	2,70
7	20,9	21,2	21,5	21,8	-0,60
8	21,4	21,5	21,9	21,7	0,35
9	23,5	23,5	21,0	21,0	2,50
10	22,3	22,9	22,0	21,3	0,55
11	23,5	24,1	20,8	20,6	3,10
12	22,5	23,5	21,0	22,0	1,50
13	22,5	23,5	21,0	21,0	2,00
14	23,4	22,7	22,0	22,0	1,05
15	24,0	24,2	22,1	21,5	2,30
16	24,5	24,4	22,3	22,5	2,05
17	24,8	24,7	22,0	21,9	2,60
18	24,7	25,1	21,9	21,9	3,00
19	24,9	24,4	22,4	22,6	2,15
20	27,2	27,0	24,5	23,7	3,00
Average difference, $\bar{D}$					1,73
Standard deviation					1,07
Standard uncertainty of $\bar{D}$ (standard deviation / $\sqrt{20}$ )					0,24
NOTE The data are measurements of the mechanical strength of aggregate, obtained from the Los Angeles (LA) test.					

## 5.5 Consensus values from expert laboratories [see ISO/IEC Guide 43-1:1997, A.1.1 item d)]

### 5.5.1 General

### 5.5.2 Standard uncertainty $u_X$ of the assigned value

When each of  $p$  expert laboratories reports a measurement  $y_i$  on the test material together with an estimate  $u_i$  of the standard uncertainty of the measurement, and the assigned value  $X$  is calculated as a robust average using Algorithm A, the standard uncertainty of the assigned value  $X$  is estimated as:

$$u_X = \frac{1,25}{p} \sqrt{\sum_{i=1}^p u_i^2} \quad (7)$$

When the expert laboratories do not report standard uncertainties, or when the uncertainties are not validated independently (e.g. by a laboratory accreditation body), the standard uncertainty of the assigned value shall be estimated as described in 5.6.

การอบรมเชิงปฏิบัติการ Statistical approach for PT (Testing and Calibration)

วันที่ ๒๘-๓๐ พฤษภาคม ๒๕๕๕

ณ ห้อง ๓๒๐ อาคารสถานศึกษามหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี

จัดโดย สำนักบริหารและรับรองห้องปฏิบัติการ

## 5.6 Consensus value from participants [see ISO/IEC Guide 43-1:1997, A.1.1 item e)]

### 5.6.1 General

#### 5.6.2 Standard uncertainty $u_X$ of the assigned value

When the assigned value is derived as a robust average calculated using Algorithm A, the standard uncertainty of the assigned value  $X$  is estimated as:

$$u_X = 1.25 \times s^* / \sqrt{p} \quad (8)$$

where  $s^*$  is the robust standard deviation of the results calculated using Algorithm A in Annex C. (Here a "result" for a participant is the average of all their measurements on the test material.)

The limitations of this approach are that:

- a) there may be no real consensus amongst the participants;
- b) the consensus may be biased by the general use of faulty methodology and this bias will not be reflected in the standard uncertainty of the assigned value calculated as described above.

Neither of these conditions is rare in the determination of trace constituents.

การอบรมเชิงปฏิบัติการ Statistical approach for PT (Testing and Calibration)

วันที่ ๒๘-๓๐ พฤษภาคม ๒๕๕๕

ณ ห้อง ๓๒๐ อาคารสถานศึกษาเคมีปฏิบัติ กรมวิทยาศาสตร์บริการ

จัดโดย สำนักบริหารและรับรองห้องปฏิบัติการ

การประเมินสมรรถนะของห้องปฏิบัติการ  
การคำนวณค่า Z-score, En และอื่นๆ

วิทยากร: ผศ.ดร.ครรชิต จุดประสงค์  
นางจันทร์รัตน์ วรรณพวิทย์



## Part IV.



# Evaluation of laboratory performance

*Asst. Prof. Kunchit Judprasong\**

*Assoc. Prof. Prapasri Puwastien\**

*Chantarat Vorasapavit\*\**

\*Institute of Nutrition, Mahidol University, Salaya, Putthamonthon 4,  
Nakhon Pathom 73170, THAILAND  
Tel: 0 2800 2380 ext . 324

\*\* Department of Science and Service

[kunchit.jud@mahidol.ac.th](mailto:kunchit.jud@mahidol.ac.th)

Statistical approach for PT (Testing and Calibration)  
(APLAC-PTB approved work plan)  
28-30 พฤษภาคม 2555



## Laboratory performance study

**Objective:** to assess the performance of a  
laboratory (ies) [or analyst (s)]

**Activities:** a group of laboratories analyse one  
or more components of one or more homogenous and stable  
test-samples by the methods used or selected by each  
laboratory.

**Outcome:** status of laboratory performance, selection of  
methods for specific analysis, development of consensus  
values of analytes



## การคำนวณค่า Z-score

- Z-scores rely on the results being normally distributed
- Microbiological results are known to be non normally distributed (Poisson distribution)
  - Log-transformation prior to calculating z-scores

### PT-8: Protein in Rice (g/100g)

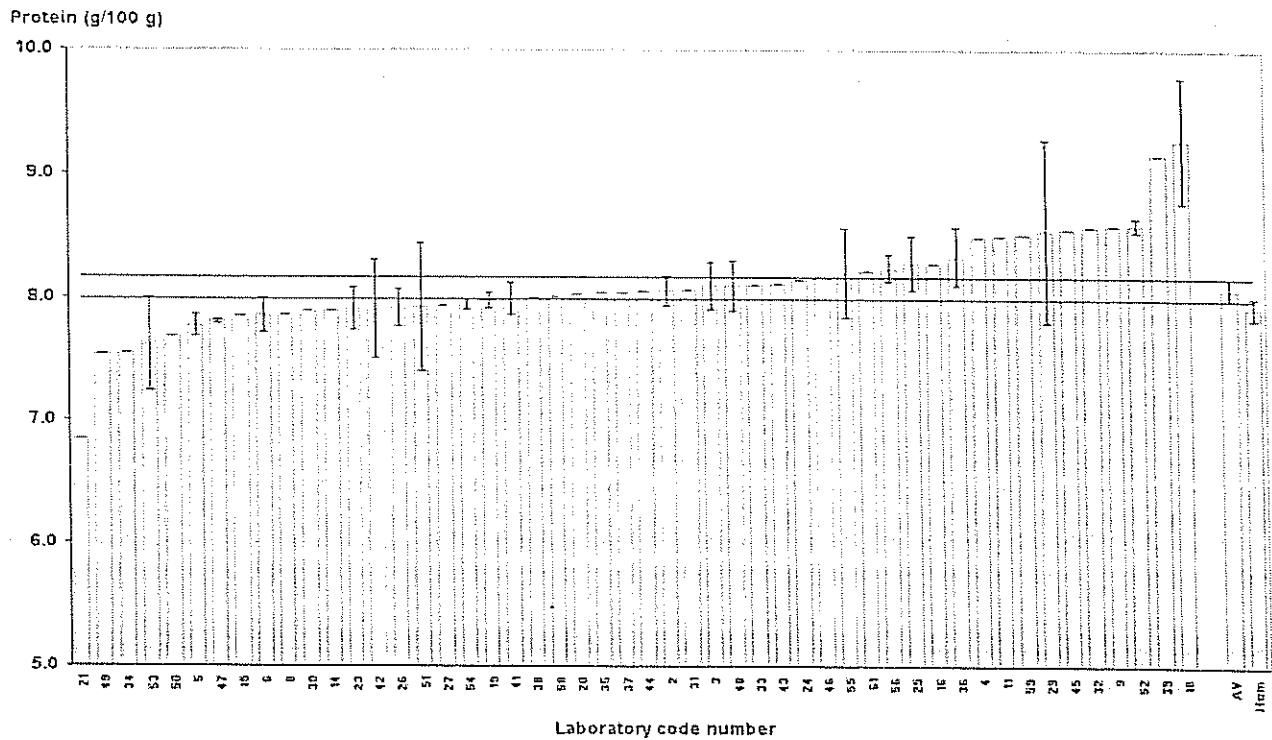
Results submitted from participating laboratories (N=50)



Laboratory Code No	Result			Uncertainty	
	A	B	Mean	+	-
21	6.82	6.88	6.85	+	-
49	7.55	7.53	7.54	+	-
34	7.67	7.43	7.55	+	-
53	7.57	7.67	7.62	+	0.37
50	7.70	7.67	7.69	+	-
5	7.83	7.71	7.77	+	0.09
47	7.77	7.84	7.81	+	0.02
15	7.81	7.89	7.85	+	-
6	7.90	7.81	7.86	+	0.14
8	7.80	7.91	7.86	+	-
30	7.88	7.90	7.89	+	-
14	7.85	7.94	7.90	+	-
23	7.90	7.91	7.91	+	0.17
42	7.91	7.90	7.91	+	0.40
26	7.91	7.92	7.92	+	0.15
51	7.98	7.86	7.92	+	0.52
27	7.91	7.96	7.94	+	-
54	7.95	7.94	7.95	+	0.04
19	7.97	7.98	7.98	+	0.06
41	8.00	7.98	7.99	+	0.13
38	8.03	7.97	8.00	+	-
58	7.97	8.04	8.01	+	-
20	8.04	8.01	8.03	+	-
35	8.03	8.04	8.04	+	-
37	8.02	8.06	8.04	+	-
44	7.76	8.33	8.05	+	-
2	8.05	8.06	8.06	+	0.12

Laboratory Code No	Result			Uncertainty	
	A	B	Mean	+	-
31	8.09	8.02	8.06	+	-
3	8.10	8.09	8.10	+	0.19
48	8.15	8.04	8.10	+	0.20
33	8.11	8.10	8.11	+	-
43	8.12	8.10	8.11	+	-
24	8.12	8.16	8.14	+	-
46	8.17	8.15	8.16	+	-
55	8.21	8.18	8.20	+	0.36
61	7.82	8.61	8.22	+	-
56	8.28	8.20	8.24	+	0.11
25	8.33	8.22	8.28	+	0.22
16	8.29	8.27	8.28	+	-
36	8.35	8.32	8.34	+	0.24
4	8.47	8.51	8.49	+	-
11	8.55	8.44	8.50	+	-
59	8.63	8.39	8.51	+	-
29	8.52	8.55	8.54	+	0.74
45	8.57	8.52	8.55	+	-
32	8.54	8.60	8.57	+	-
9	8.34	8.82	8.58	+	-
52	8.56	8.61	8.59	+	0.06
39	9.23	9.07	9.15	+	-
18	9.24	9.30	9.27	+	0.51
AV	Assigned value		8.07	+	0.09
Hom	Homogeneity		7.92	+	0.09

# Individual data of protein in Rice-1 (N=50) (with uncertainty values from 21 lab)



## Laboratory performance by estimation of robust z-score

### 1. Within laboratory variation (see example)

$$z\text{-score}_{\text{within}} = \frac{x - \text{median}_{(\text{diff})}}{\text{NIQR}_{(\text{diff})}}$$

$x$  = Difference between reported values of A and B from each laboratory/SQRT 2

Median = Median of the difference between reported values of A and B ( $x$ )

NIQR = Normalised Inter-quartile Range  
= (Quartile 3-Quartile 1)  $\times$  0.7413

### Interpretation

$|z| \leq 2$  Satisfactory result ("a")

$2 < |z| < 3$  Questionable result ("w")

$|z| \geq 3$  Unsatisfactory result ("ww")

w: within laboratory z-score



## 2. Between laboratory variation (see example)

$$\text{Robust z-score} = \frac{x_i - \bar{X}}{SD}$$

where  $x$  = average value of reported data from each participant

$\bar{X}$  = robust mean ( $x^*$ ) obtained by ISO 13528

or median of the mean values obtained from participants

SD = robust standard deviation ( $s^*$ ) obtained by ISO 13528

or predicted SD of Horwitz

### Interpretation

$|z| \leq 2$  Satisfactory (acceptable) result ("a")

$2 < |z| < 3$  Questionable (warning) result ("b")

$|z| \geq 3$  Unsatisfactory (unacceptable) result ("bb")

b: between laboratory z-score



## 2. Between laboratory variation (con't)

- In case of participating laboratories < 10

$$\text{Modified z-score } MZ = \frac{|x_i - \mu|}{\sigma}$$

where  $x_i$  = average value of reported data from each participant

$\mu$  = robust mean ( $x^*$ ) obtained by iterative calculation

(downweighting procedure)

$\sigma$  = robust standard deviation = MAD/0.6745

MAD = median absolute deviation = median  $[|x_i - \text{median}(x_i)|]$

### Interpretation

$|z| \leq 2$  Satisfactory (acceptable) result ("a")

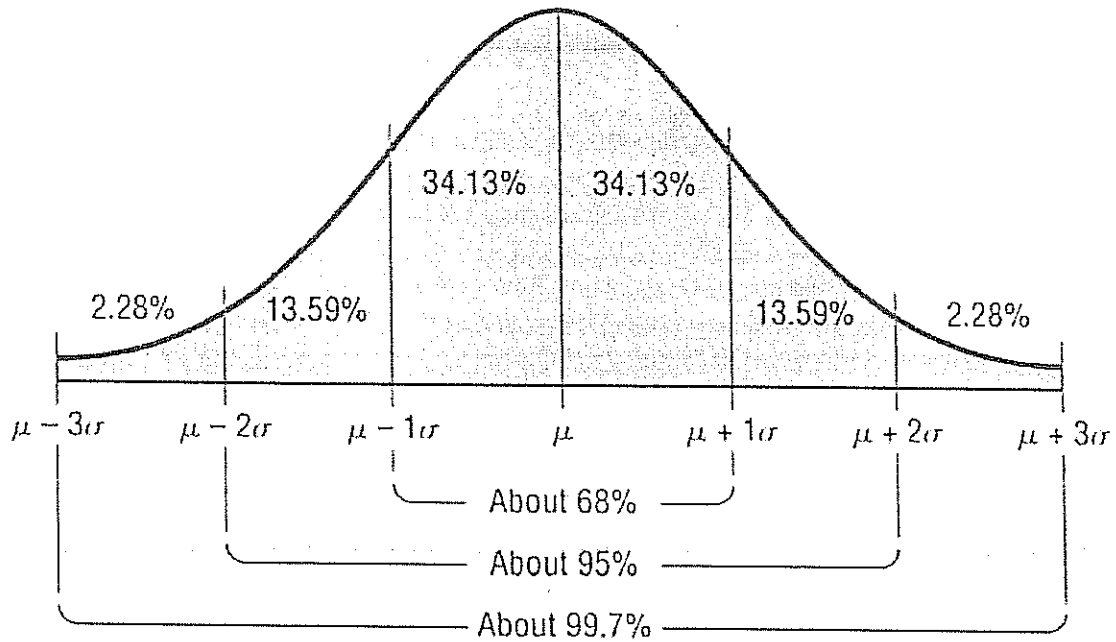
$2 < |z| < 3$  Questionable (warning) result ("b")

$|z| \geq 3$  Unsatisfactory (unacceptable) result ("bb")

b: between laboratory z-score



# Interpreting Z-scores



ผศ.ดร.ครรชิต จุดประสงค์ & รศ.ดร.ประกาศศรี กุลเสถียร Chantarot Vorasapavit

9

LABORATORY PERFORMANCE STUDY: evaluate against assigned value of test material  
Assigned value (robust mean + robust SD) of protein in Rice-1 =  $8.07 \pm 0.25$  g/100g (% CV = 3.1)

Laboratory Code No	Result		Di	Si	z-score			Conclusion based on (1)	Values for consensus
	A	B			(B-A)/SQRT2	(A+B)/SQRT2	Within		
2	8.05	8.06	0.01	11.39	0.29	0.05	-0.06	a	8.06
3	8.10	8.09	-0.01	11.45	0.00	0.20	0.10	a	8.10
4	8.47	8.51	0.03	12.01	0.73	1.68	1.68	a	8.49
8	7.97	7.99	0.01	11.29	-0.44	-0.24	-0.36	a	7.98
18	9.24	9.30	0.04	13.11	1.02	4.62	4.80	bb	
19	7.97	7.95	-0.01	11.26	0.29	-0.25	-0.38	a	7.98
20	8.04	8.01	-0.02	11.35	-0.29	-0.07	-0.18	a	8.03
31	8.09	8.02	-0.05	11.39	-0.88	0.05	-0.05	a	8.05
32	8.54	8.6	0.04	12.12	1.02	1.98	2.00	a	8.57
33	8.11	8.1	-0.01	11.46	0.00	0.24	0.14	a	8.11
34	7.67	7.43	-0.17	10.68	-3.35	-1.85	-2.08	wvb	
35	8.03	8.04	0.01	11.35	0.29	-0.03	-0.14	a	8.04
36	8.35	8.32	-0.02	11.79	-0.29	1.10	1.06	a	8.34
37	8.02	8.06	0.03	11.37	0.73	-0.01	-0.12	a	8.04
38	8.03	7.97	-0.04	11.31	-0.73	-0.16	-0.28	a	8.00
39	9.23	9.07	-0.11	12.94	-2.19	4.16	4.32	wbb	
41	8.00	7.98	-0.01	11.30	-0.15	-0.20	-0.32	a	7.99
42	7.91	7.9	-0.01	11.18	0.00	-0.52	-0.65	a	7.91
43	8.12	8.1	-0.01	11.47	-0.15	0.25	0.16	a	8.11
44	7.76	8.33	0.40	11.39	8.46	0.01	-0.10	wv	
45	8.67	8.52	-0.04	12.08	-0.59	1.89	1.90	a	8.55
50	7.70	7.67	-0.02	10.87	-0.29	-1.34	-1.54	a	7.69
51	7.98	7.86	-0.06	11.20	-1.60	-0.46	-0.60	a	7.92
52	8.56	8.61	0.04	12.14	0.88	2.04	2.06	b	
53	7.57	7.67	0.07	10.78	1.50	-1.59	-1.80	a	7.62
54	7.85	7.94	-0.01	11.24	0.00	-0.37	-0.50	a	7.95
55	8.21	8.19	-0.02	11.59	-0.29	0.57	0.50	a	8.20
61	7.82	8.51	0.56	11.62	11.67	0.65	0.58	ww	
No of results	50	50	50	50					40
Median	8.03	8.04	-0.01	11.37					8.07
O1	7.90	7.91	-0.03	11.12					0.23
O3	8.29	8.31	0.03	11.69					2.88
IQR (O3-O1)	0.39	0.40	0.07	0.51					
Normalised IOP			0.05	0.38					
Robust CV				3.31					

Final consensus value for protein in Rice-1 calculated from laboratories with accepted values =  $8.07 \pm 0.23$  g/100g (mean  $\pm$  SD), with N=40, %CV=2.9

← Show only some data

(1) based on median  $\pm$  MIOR; (2) based on assigned value estimated according to ISO 13628

# Laboratory performance by estimation of En score\*

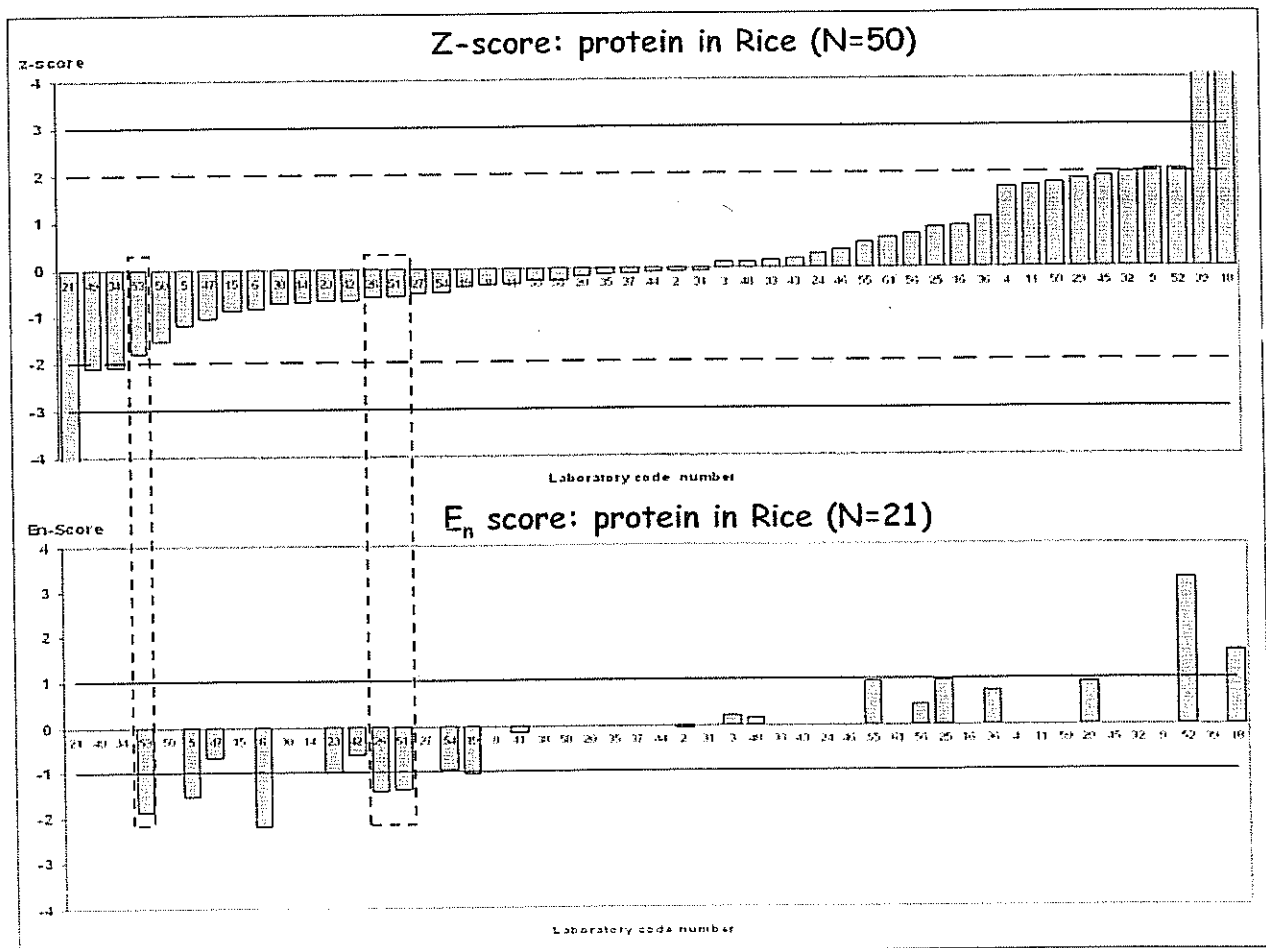


- PT-8 trial study:
- request uncertainty values of moisture, protein and iron
  - calculate  $E_n$  score and evaluate the laboratory performance using the  $E_n$  score

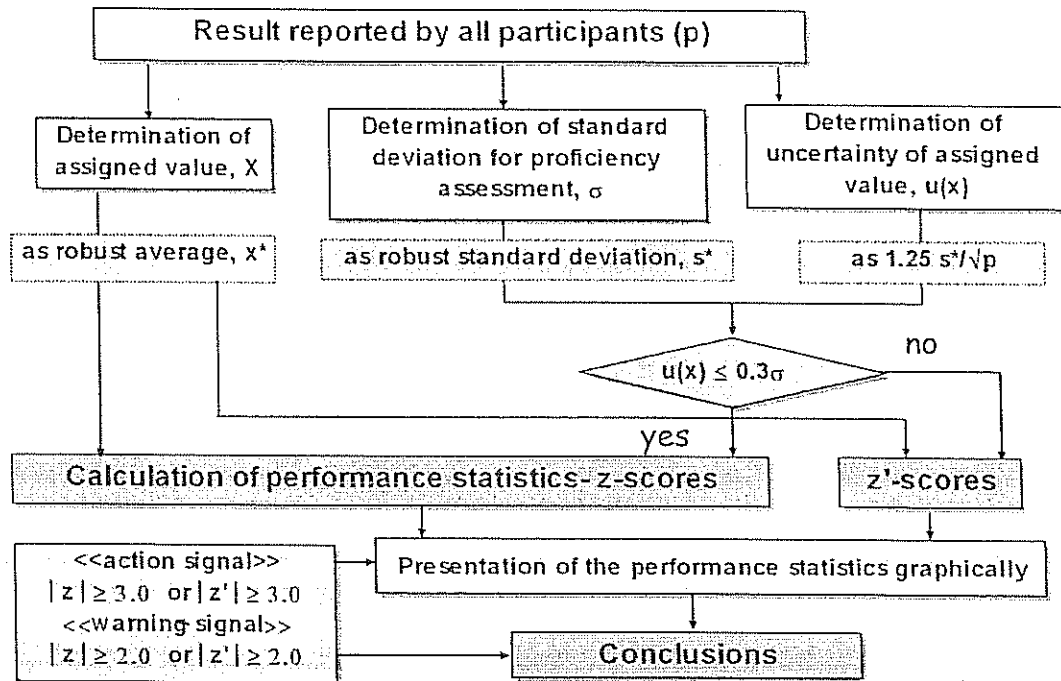
$$E_n = \frac{x_{lab} - x_{ref}}{\sqrt{(U_{lab}^2 + U_{ref}^2)}}$$

$[E_n \text{ score}] \leq 1$ : reported results agree with assigned values  
 $[E_n \text{ score}] > 1$ : reported results is different from the assigned value and the reported uncertainty values are underestimated

- where
- $x_{lab}$  = mean result submitted by participants
  - $x_{ref}$  = robust mean of the assigned value derived from ISO 13528
  - $U_{lab}$  = expanded uncertainty of  $x_{lab}$
  - $U_{ref}$  = expanded uncertainty of  $x_{ref}$  derived from ISO 13528



## Summary



## Technical meeting after PT

### Objective :

- Present final report to participants
- Discuss , exchange experience, share problems and solutions
- Conduct activities to improve analytical performance
- Plan for future activities



# Exercise

## Evaluation of laboratory performance