TECHNICAL SPECIFICATION

ISO/TS 16489

First edition
2006-05-15

Water quality — Guidance for establishing the equivalency of results

Qualité de l'eau — Lignes directrices pour la création de l'équivalence des résultats

Reference number
ISO/TS 16489:2006(E)

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
Web www.iso.org

Published in Switzerland

Contents

Page

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.

In other circumstances, particularly when there is an urgent market requirement for such documents, a technical committee may decide to publish other types of normative document:

— an ISO Publicly Available Specification (ISO/PAS) represents an agreement between technical experts in an ISO working group and is accepted for publication if it is approved by more than 50 % of the members of the parent committee casting a vote;

— an ISO Technical Specification (ISO/TS) represents an agreement between the members of a technical committee and is accepted for publication if it is approved by 2/3 of the members of the committee casting a vote.

An ISO/PAS or ISO/TS is reviewed after three years in order to decide whether it will be confirmed for a further three years, revised to become an International Standard, or withdrawn. If the ISO/PAS or ISO/TS is confirmed, it is reviewed again after a further three years, at which time it must either be transformed into an International Standard or be withdrawn.

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/TS 16489 was prepared by Technical Committee ISO/TC 147, *Water quality*, Subcommittee SC 2, *Physical, chemical and biochemical methods*.

Introduction

The methods referred to in this Technical Specification can comprise a standard or reference method, the results of which are to be compared with results generated by an alternative, perhaps more simple, method. Alternatively, a comparison of results produced by an old established method and those produced by a new more modern technique can be undertaken. The methods can be laboratory based or undertaken “on-site” where the samples are taken.

No indication is given to confirm whether either one of the two methods, in terms of bias, is better or worse than the other method, only that the results produced by both methods are considered equivalent or not, in terms of the calculated means, standard deviations and variances. The procedures described are not to be used for, and do not apply to, situations to establish whether two methods can be shown to be equivalent. The procedures apply only to demonstrating equivalency of results.

Since standard deviations and means can vary with concentrations, especially where concentrations vary over several orders of magnitude, the procedures described in Clauses 7 to 9 are only applicable to samples containing a single level of concentration. It would be necessary to repeat the procedures for each concentration level if different concentration levels are encountered, and it is shown that standard deviations and means vary over these concentration levels. It might be that the demonstration of equivalence can only be achieved over relatively small concentration ranges. For multiple concentration levels, the procedures described in Clause 10 might be applicable. In addition, the laboratory will need to show that both methods are suitable and appropriate for the sample matrix and the parameter under investigation, including the level of concentration of the parameter. Also, the experimental data obtained in the comparison of results should reflect the specific application for which equivalence is questioned, as different matrices can lead to different results with the two methods.

Throughout this Technical Specification, it is assumed that results are obtained essentially under repeatability conditions, but it is recognized that this will not always be so. Hence, where appropriate, identical samples are analysed by the same analyst using the same reagents and equipment in a relatively short period of time. Furthermore, a level of confidence of 95 % is assumed. The statistical tests described in this Technical Specification assume that the data to be compared are independent and normally distributed in a Gaussian manner. If they are not, the data might not be suitable for the statistical treatments described and additional data might need to be collected.

The power of the statistical test is greatly enhanced when sufficient data are available for comparisons; i.e. when the numbers of degrees of freedom are available to enable a meaningful interpretation to be made. However, it is recognized that a statistically significant difference might not necessarily infer an important or meaningful difference, and a personal judgement should be made on whether a statistically significant difference is important or meaningful and relevant. Alternatively, a statistical test might not be sufficiently powerful to be able to detect a difference that from a practical point of view could be regarded as important or meaningful.

To aid the analyst, advice is provided as to which clause (and corresponding annex) is applicable to the circumstances surrounding the data that have been generated. It is recognized that when results are compared they can have been generated under a variety of different conditions.

Water quality — Guidance for establishing the equivalency of results

1 Scope

This Technical Specification describes statistical procedures to test the equivalency of results obtained by two different analytical methods used in the analysis of waters. This Technical Specification is not applicable for establishing whether two methods can be shown to be equivalent. The procedures given in this Technical Specification are only applicable to demonstrating the equivalency of results.

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 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*

NOTE A practical guidance document to assist in the use of ISO 5725-2 has been published: see ISO/TR 22971.

3 Terms and definitions

For the purposes of this document, the following terms and definitions apply:

3.1
precision
closeness of agreement between independent test results obtained under repeatability conditions

NOTE 1 Precision depends only on the distribution of random errors and does not relate to the true, specified or accepted value.

NOTE 2 Measurement of precision is usually expressed in terms of imprecision and computed as a standard deviation of the test results. Less precision is reflected by a larger standard deviation.

NOTE 3 “Independent test results” means results obtained in a manner not influenced by any previous result on the same sample. Quantitative measurements of precision depend critically on stipulated conditions.

3.2
repeatability conditions
conditions where independent test results are obtained with the same method on identical test samples in the same laboratory, by the same operator, using the same reagents and equipment within short intervals of time

3.3
analytical method
unambiguously written procedure describing all details required to carry out the analysis of the determinand or parameter, namely: scope and field of application, principle and/or reactions, definitions, reagents, apparatus, analytical procedures, calculations and presentation of results, performance data and test report

4 Overview of the different approaches

Where a sample is analysed in replicate using two methods, then the procedures described in Clause 7 and Annex B may be used. The results should, ideally, be generated by a single analyst, however, it is recognized that different analysts can be involved.

The procedures described in Clause 8 and Annex C might be applicable where, over a period of time, samples are analysed by different analysts using a particular method and these results are compared with results generated using an alternative method that is carried out by one or more analysts. In this case, however, the assumption of repeatability will not be applicable.

Where different analysts are involved in the generation of data, the procedures described in Clause 9 and Annex D may be used. In these cases, the assumption of repeatability will not be applicable. Where identical samples are analysed by one or more analysts using two different methods, the procedures described in Clause 10 and Annex E might be more appropriate. This might be applicable where the same or different concentration levels are indicated.

5 Amount of data

The approach described in this Technical Specification demonstrates the importance that the power of the significance tests lies in the amount of data available as well as the quality (spread) of the data. Throughout this Techncial Specification, it is assumed that the level of confidence is established at 95 %. This might represent a degree of acceptability that is insufficient for certain purposes. This would mean that individual circumstances would merit individual consideration as to whether this Technical Specification, in terms of the confidence level used, should be applied. Confidence levels of 99 % or higher might be, in certain circumstances, more appropriate. In addition, where a statistically significant difference has been suggested by a statistical analysis of the data, there is always a need to question whether this difference is important or relevant, in terms of its suitability and fitness for purpose, and not in terms of its statistical meaning or understanding. This judgement should be based on whether the analytical results are fit for their intended purpose.

For example, with large amounts of data, it is possible to conclude that there is a statistically significant difference between 50,1 and 50,2. Whether this difference is important or meaningful is another matter when deciding on the suitability of the method.

Before any statistical treatment is undertaken, it is always useful to plot a graph of the data. This will provide a visual display of the results, an inspection of which should reveal the amount and quality of data available for comparison. In this way, the number of results and the spread (or range) of the data is easily observed. Figures F.1 to F.6 (Annex F) show illustrative examples of the type of plots that can be produced and the interpretations that can be concluded. Figures F.1 to F.3 show the arithmetic means of the results from a series of determinations undertaken in comparative exercises of two methods and the associated interpretations. Figures F.4 to F.6 show the spread or range of results from a series of determinations and possible interpretations.

From the data, the arithmetic mean (average) $\bar{x}$ of a number, n, of determinations or measurements, x_i, and the standard deviation, s, of numerous repeated determinations obtained under repeatability conditions, are calculated from Equations (1) and (2):

$$\bar{x} = \frac{\sum_{i=1}^{i=n} x_i}{n} \tag{1}$$

$$s = \sqrt{\frac{\sum_{i=1}^{i=n} x_i^2 - \frac{\left(\sum_{i=1}^{i=n} x_i\right)^2}{n}}{n-1}} \tag{2}$$

The square of the standard deviation is known as the variance, namely, s^2.

6 Data comparisons

When the results from two methods are compared, different situations will arise depending upon the circumstances surrounding the manner in which the results are determined. Hence, the comparison will differ for different situations. By way of example, Clauses 7 to 10 describe the different approaches that can be encountered when sets of data are to be compared. In addition, since the comparisons undertaken in this Technical Specification are used to establish whether a difference between sets of data exists, rather than to determine whether one set of data is superior to another, then a two-sided test is carried out, rather than a one-sided test.

Data comparisons can be further complicated by the inclusion of outlier tests to establish whether sets of data contain values that are considered significantly different from the rest of the data. A number of different outlier tests are available and some of these are described in more detail in ISO 5725-2. Other outlier tests may also be used, for example see Annex E. Further consideration of, and the need for, outlier tests are not considered in this Technical Specification but will need to be taken into consideration.

The example comparisons and information contained in Figures F.1 to F.6 and Annexes B to E are for illustrative purposes only. Suitable computer software might be available to facilitate the numerical calculations. In addition, the examples shown are based on limited data to highlight the manner in which the calculations were carried out. They are not presented as actual data comparisons. In reality, many more results would be required before calculations of this type are undertaken. Schematic diagrams outlining the procedures that can be undertaken are shown in Figures G.1 and G.2 in Annex G.

Samples for analysis should be taken using procedures given in relevant International Standards appropriate to the parameter being analysed.

7 Comparison of arithmetic means of two independently obtained sets of data

Under repeatability conditions, analyse a sample in replicate using the two methods. The number of replicate determinations or measurements carried out with each method can be different, but for both methods should be sufficient to provide confidence in the statistical treatment that follows. This may involve 6 to 10 or more repeat determinations. For example, for the analytical method, method i, the following determinations can be obtained, namely x_1, x_2, x_3, $x_4 \ldots x_{n-1}$ and x_n. For the alternative analytical method, method j, the following determinations can be obtained, namely y_1, y_2, $y_3 \ldots y_{m-1}$ and y_m. From these values the corresponding means, standard deviations and variances are calculated, $\bar{x}$, $\bar{y}$, s_i, s_j, s_i^2 and s_j^2 respectively.

To ascertain whether the precision or spread of data (in terms of the variances s_i^2 and s_j^2) obtained from the two methods differ statistically, a statistical F-test should be carried out. This statistical test will show whether there is a statistically significant difference between the two variances. The F-value calculated (F_{calc}) should then be compared with the tabulated or theoretical F-value (F_{tab}) obtained for the corresponding amount of data; i.e. number of degrees of freedom, at the stated level of confidence required, in this case 95 % (see Table A.1). If F_{tab} is less than F_{calc}, then it can be concluded that there is a statistically significant difference between the two variances; i.e. s_i^2 and s_j^2 are not the same and, hence, cannot be regarded as being equivalent.

Under these circumstances, the variances should not be combined to form a single variance value. The method exhibiting the smaller variance is the more precise of the two methods.

If F_{tab} is greater than F_{calc}, then it can be concluded that there is no statistically significant difference between the two variances; i.e. s_i^2 and s_j^2 can be regarded as being similar and, hence, can be regarded as being equivalent. Under these circumstances, the precision of the results generated by both methods can be regarded as being equivalent.

F_{calc} should be calculated as follows:

$$F_{\text{calc}} = \frac{{s_i}^2}{{s_j}^2} \text{ or } F_{\text{calc}} = \frac{{s_j}^2}{{s_i}^2} \tag{3}$$

The equation is always arranged so that a value greater than 1 is obtained.

If no statistically significant difference is indicated for the variances, i.e. if F_{tab} is greater than F_{calc}, then the spread of results from both methods can be regarded as being similar. In such a case, the results from both methods can be combined to produce a pooled or combined standard deviation, s_c, according to Equation (4):

$$s_c = \sqrt{\frac{s_i^2(n-1) + s_j^2(m-1)}{n+m-2}} \tag{4}$$

To ascertain if the arithmetic means, $\bar{x}$, $\bar{y}$, obtained for both methods differ statistically, a t-test should be carried out. This test will show whether there is a statistically significant difference between the two means. The t-value calculated (t_{calc}) should then be compared with the tabulated or theoretical t-value (t_{tab}) obtained for the corresponding amount of data; i.e. number of degrees of freedom, at the stated level of confidence required, in this case 95 % (see Table A.2). If t_{tab} is less than t_{calc}, then it can be concluded that there is a statistically significant difference between the two arithmetic means; i.e. $\bar{x}$ and $\bar{y}$ are not the same, and hence cannot be regarded as being equivalent.

If t_{tab} is greater than t_{calc}, then it can be concluded that there is no statistically significant difference between the two means; i.e. $\bar{x}$ and $\bar{y}$ can be regarded as being similar and, hence, can be regarded as being equivalent. Under these circumstances, the bias of the results generated by both methods can be regarded as being equivalent.

t_{calc} should be calculated as follows:

$$t_{calc} = \frac{\left|(\bar{x} - \bar{y})\right|}{s_c\sqrt{\left(\frac{1}{n} + \frac{1}{m}\right)}} \tag{5}$$

Using these tests, it can be concluded that the precision and bias of the results generated for both methods might or might not be similar. Only if the precision (in terms of s_i^2 and s_j^2) and bias (in terms of $\bar{x}$ and $\bar{y}$) of both sets of results show no statistically significant difference can the results be considered equivalent.

An example of this approach is shown in Annex B.

The use of these statistical tests can also indicate whether the method performance capabilities change significantly over periods of time from those originally established. In these instances, it might be that analytical quality control data can be used and compared over the two time periods rather than considering the data being generated by two different methods.

8 Comparison of population and sample arithmetic means

Over a long period of time, a method might be used by different analysts which provides sufficient information to be established, for example on the overall arithmetic mean, μ, of quality control samples. If a different method is then used by a number of analysts and information gathered on its performance, for a (small) number of determinations, n, the arithmetic mean, $\bar{x}$, and standard deviation, s, can be calculated from results obtained using the new method.

To ascertain whether the results from the new method differ statistically from the results obtained by the old method, a t-test should be carried out. This test will show whether there is a statistically significant difference between the two means, μ and $\bar{x}$. The t-value calculated (t_{calc}) should then be compared with the tabulated or theoretical t-value (t_{tab}) obtained for the corresponding amount of data; i.e. number of degrees of freedom, at the stated level of confidence required (see Table A.2). If t_{tab} is less than t_{calc}, then it can be concluded that there is a statistically significant difference between the two arithmetic means; i.e. μ and $\bar{x}$ are not the same, and hence, cannot be regarded as being equivalent.

If t_{tab} is greater than t_{calc}, then it can be concluded that there is no statistically significant difference between the two means; i.e. μ and $\bar{x}$ can be regarded as being similar, and hence, can be regarded as being

equivalent. Under these circumstances, the bias of the results generated by both methods can be regarded as being equivalent.

On this occasion, t_{calc} should be calculated as follows:

$$t_{calc} = \frac{\left|(\bar{x} - \mu)\right|}{\frac{s}{\sqrt{n}}} \tag{6}$$

An example of this approach is shown in Annex C.

As well as demonstrating the equivalency of results, this test can also be used to ascertain if a method that is used and exhibits a certain bias is deemed acceptable when compared with a target bias value. For example, a method exhibiting a bias of, say, 10,5 % might or might not be statistically acceptable when compared with a target bias value of, say, 10 %. Hence, the actual method performance can be compared to a required level of performance.

9 Analysis of variance

When a new method is proposed, a number of different analysts might be used to generate results or performance data to demonstrate its capability. Under these circumstances, when repeat determinations are made, there will always be some variability in the results and it can be difficult to ascertain if real differences exist between the different sets of data produced by the different analysts. One way this could be undertaken would be to carry out repeated t-tests, as described in Clause 7. Repeated use of this test to compare all combinations of data sets, however, increases the probability of making erroneous conclusions. An easier way is to carry out an analysis of variance (ANOVA) test. This test will help to ascertain whether there are statistically significant differences between the sets of data generated by the different analysts. In other words, the ANOVA test is used to determine the statistical significance of differences in the arithmetic means of different sets of data. The data should be arranged as indicated in Table 1, and then an F-test should be carried out. The F-value calculated (F_{calc}) should then be compared with the tabulated or theoretical F-value (F_{tab}) obtained for the corresponding amount of data i.e. number of degrees of freedom, at the stated level of confidence required, in this case 95 %, (see Table A.1).

Table 1 — Statistical significance of differences in the arithmetic means of different sets of data

Replicate determinations	Analysts					
	1	2	3	i	$p-1$	p
1	x_{11}	x_{21}	x_{31}	x_{i1}	$x_{(p-1)1}$	x_{p1}
2	x_{12}	x_{22}	x_{32}	x_{i2}	$x_{(p-1)2}$	x_{p2}
k	x_{1k}	x_{2k}	x_{3k}	x_{ik}	$x_{(p-1)k}$	x_{pk}
$n-1$	$x_{1(n-1)}$	$x_{2(n-1)}$	$x_{3(n-1)}$	$x_{i(n-1)}$	$x_{(p-1)(n-1)}$	$x_{p(n-1)}$
n	x_{1n}	x_{2n}	x_{3n}	x_{in}	$x_{(p-1)n}$	x_{pn}

Equations (7) to (9) should then be used to calculate the test statistics.

$$A = \sum_{i=1}^{i=p} \frac{\left(\sum_{k=1}^{k=n} x_{ik}\right)^2}{n} \tag{7}$$

$$B = \sum_{i=1}^{i=p} \sum_{k=1}^{k=n} x_{ik}^{\,2} \tag{8}$$

$$C = \frac{\left(\sum_{i=1}^{i=p} \sum_{k=1}^{k=n} x_{ik} \right)^2}{N} \tag{9}$$

where the total number of replicates $N = np$.

If the number of replicates for each analyst is not the same, then Equation (10) should be used instead of Equation (7) to calculate the test statistic

$$D = \sum_{i=1}^{i=p} \frac{\left(\sum_{k=l}^{k=n_i} x_{ik} \right)^2}{n_i} \tag{10}$$

where n_i is the number of replicates determined by each analyst.

In addition, the value of N should be calculated accordingly.

To ascertain whether the sets of data differ statistically, an ANOVA table of results should be calculated as shown in Table 2:

Table 2 — ANOVA results

Source of variation	Sum of squares, S	Degrees of freedom	Mean square, M	F_{calc}
Between analysts	$S_1 = A - C$	$p - 1$	$M_1 = S_1/(p - 1)$	M_1/M_0
Within analysts	$S_0 = B - A$	$N - p$	$M_0 = S_0/(N - p)$	
Total	$S_1 + S_0 = B - C$	$N - 1$		
NOTE A, B and C are defined in Equations (7) to (9).				

If F_{tab} is less than F_{calc}, then it can be concluded that there is a statistically significant difference between the sets of data. Hence, one or more of the sets of data is not the same as the remaining sets of data, and hence the whole data set cannot be regarded as being equivalent.

If F_{tab} is greater than F_{calc}, it can be concluded that there is no statistically significant difference between all the sets of data that can be regarded as being similar and hence, equivalent. In such a case, all the results can be combined and the overall mean, standard deviation and variance values calculated. The procedure described in Clause 7 can then be used to ascertain whether the results, in terms of means and variances, obtained for this new method are significantly different from the corresponding values obtained using an existing method.

One disadvantage to this technique is that, if a statistically significant difference is indicated, no information is provided as to which set of data might be different from the remaining sets of data. It might, however, be the case that a graphical plot of the results suggests which set of data is different from the remaining sets. Removing this particular set of data and repeating the procedure might provide confirmation.

An example of this approach is shown in Annex D.

10 Determination of the equivalence of analytical results obtained from samples from different matrices

10.1 General

The procedures described in 10.2 to 10.3 enable the comparison of results to be undertaken using two different analytical methods, provided that there is no statistically significant difference (using the F-test for variances) in the precision between the results of standard solutions obtained by both methods.

The test for the equivalence of analytical results is carried out on the basis of real samples using orthogonal regression, or the difference method.

All analyses should be carried out at the same time, either in one laboratory or two.

For the test of equivalence using orthogonal regression or the difference method, sufficient data will need to be generated. For example, a minimum of 30 samples might need to be analysed. The statistically determined number of samples will vary depending upon different circumstances and there should be sufficient results for carrying out the comparison. The samples should be homogenized and appropriately sub-sampled for each method prior to analysis being carried out.

10.2 Determination of the equivalence of the analytical results of real samples using orthogonal regression

10.2.1 Single/replicated determinations

Single and replicated determinations may be carried out for each sample but the number of replicate determinations, N, should be the same for both methods.

10.2.2 Working range

The working range of the analytical method and the corresponding range of the alternative method need not be the same, but the final reported results (taking into account any sample preparation and dilution steps) should be used in the equivalence test. For the orthogonal regression technique, the lowest and highest final results should differ

— by at least a factor of 5, and

— by at most a factor of 100.

If the factor is smaller, the difference method should be used (see also 10.3). If the factor is larger, the orthogonal regression should be carried out separately for different concentration levels of interest. The concentration range should be divided into at least five smaller equidistant concentration ranges. The analytical results obtained should be distributed evenly over each of these concentration ranges. When results are collated, outlier tests should be used, where necessary, to exclude those results considered to be significantly different from the rest of the data (see Clause 6 and Annex E).

10.2.3 Orthogonal regression

To test the equivalence of the results, an orthogonal regression should be carried out on corresponding pairs of results obtained using both methods; i.e. the results from the first method are plotted on the y-axis whilst the results of the second method are plotted on the x-axis. The slope, b, and intercept, a, of the regression line should be calculated respectively according to Equations (11) and (12):

Slope, b is given by:

$$b = \frac{s_y}{s_x} \tag{11}$$

where

$$s_y = \sqrt{\frac{1}{N-1}\sum\left(y_i - \bar{y}\right)^2}$$

and

$$s_x = \sqrt{\frac{1}{N-1}\sum\left(x_i - \bar{x}\right)^2}$$

Intercept, a, is given by:

$$a = \bar{y} - b \cdot \bar{x} \tag{12}$$

10.2.4 Test for systematic deviations

10.2.4.1 General

If the results from both methods agree, the regression line will possess an intercept value of a = 0 and a slope value of b = 1. If the results from both methods do not agree and deviations are observed, then it should be tested as to whether these differences are proportionally systematic (i.e. are concentration dependent), using the procedure described in 10.2.4.2 or constant-systematic (i.e. are independent of concentration), using the procedure described in 10.2.4.3.

10.2.4.2 Test for proportionally systematic deviations

If the deviations are proportionally systematic, the slope of the regression line will differ significantly from the value of b = 1. The test statistic, χ^2_{calc}, for the χ^2-test should be calculated according to Equation (13):

$$\chi^2_{calc} = N \cdot \ln\left(\frac{s^4 - s_{xy}^2}{s_x^2 s_y^2 - s_{xy}^2}\right) \tag{13}$$

where

s is a combined estimator for the standard deviation of the pairs (x_i, y_i), calculated as follows:

$$s = \sqrt{\frac{1}{2}\cdot\left(s_x^2 + s_y^2\right)}$$

s_{xy} is an estimator for the variance of the determinations x_i, and y_i, calculated as follows:

$$s_{xy} = \frac{1}{N-1}\sum\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)$$

For a sufficiently large number of samples, the statistical χ^2-test should be carried out. The χ^2-value calculated, χ^2_{calc}, should then be compared with the tabulated or theoretical χ^2 value (χ^2_{tab}) at the stated level of confidence required, in this case 95 %, i.e. χ^2_{tab} = 3,8. If the calculated test statistic, χ^2_{calc}, is larger than the tabulated value (χ^2_{tab}) the deviation is proportionally systematic, i.e. there is no equivalence. If the calculated test statistic, χ^2_{calc}, is less than or equal to the tabulated value (χ^2_{tab}) then no proportionally systematic deviation is detectable.

10.2.4.3 Test for constant-systematic deviations

If results are indicative of a constant-systematic deviation, the regression line is displaced but still retains a slope value of b = 1; i.e. the regression line is parallel to, but displaced from, the line at 45° passing through the origin. The displacement, $\bar{d}$, corresponds to the difference between the mean values:

$$\bar{d} = \bar{x} - \bar{y} \tag{14}$$

The existence of a constant-systematic deviation should be tested using the paired t-test, as described in 10.3. The test statistic should be calculated as follows:

$$t_{\text{calc}} = \frac{|\bar{x} - \bar{y}|}{s_d}\sqrt{N} \tag{15}$$

where

$$s_d = \sqrt{\frac{1}{N-1}\sum\left(d_i - \bar{d}\right)^2}$$

$$d_i = x_i - y_i$$

$$\bar{d} = \frac{\sum d_i}{N}$$

The calculated test statistic, t_{calc}, should be compared with the tabulated or theoretical value, t_{tab}, of the t-test distribution. If the calculated test statistic, t_{calc}, is larger than the tabulated value, t_{tab}, the deviation is constant-systematic; i.e. there is no equivalence. If the calculated test statistic, t_{calc}, is less than or equal to the tabulated value, t_{tab}, no constant-systematic deviation is detectable; i.e. the results are equivalent.

10.3 Evaluation according to the difference method

10.3.1 Formation of the differences

Two methods might be used by a number of different analysts to obtain results on identical portions or aliquots of the same sample. In order to ensure that variations in results due to sample stability are minimized, the analyses should be carried out at the same time. In this way, a series of results, x_1, y_1; x_2, y_2; x_3, y_3; … x_{n-1}, y_{n-1}; and x_n, y_n are obtained using the two methods. From these results, the difference in results for each sample should be calculated, namely d_1, d_2, d_3, … d_{n-1} and d_n. From the calculated paired differences, the arithmetic mean, $\bar{d}$, and standard deviation, s_d, are calculated.

Hence

$$d_i = x_i - y_i \tag{16}$$

10.3.2 Determination of equivalence (paired t-test)

To ascertain whether the arithmetic mean of the difference, $\bar{d}$, obtained using the two methods differs statistically from zero, a t-test should be carried out. This test will show whether there is a statistically significant difference. The t-value calculated (t_{calc}) should then be compared with the tabulated or theoretical t-value (t_{tab}) obtained for the corresponding amount of data; i.e. number of degrees of freedom, at the stated level of confidence required, in this case 95 %, (see Table A.2). If t_{tab} is less than t_{calc}, then it can be concluded that there is a statistically significant difference between the arithmetic mean of the differences and zero; i.e. $\bar{d}$ and 0 are not the same and, hence, cannot be regarded as being equivalent.

If t_{tab} is greater than t_{calc}, then it can be concluded that there is no statistically significant difference between the mean of the differences and zero; i.e. $\bar{d}$ and 0 can be regarded as being similar, and hence, can be regarded as being equivalent. Under these circumstances, the bias of the results generated by both methods can be regarded as being equivalent.

The equation used to calculate t_{calc} is

$$t_{\text{calc}} = \frac{|\bar{d}|}{\frac{s_d}{\sqrt{n}}} \tag{17}$$

Examples of the approaches described in this clause are shown in Annex E.

11 Reporting

When a comparison of results has been undertaken and these are considered or regarded as being equivalent, the procedure (as described in this Technical Specification) used to demonstrate this equivalency should be reported. In addition, the corresponding number of degrees of freedom and confidence level should be reported.

Annex A
(informative)

Statistical tables

Table A.1 — Two-sided F-test values (F_{tab}) at 95 % confidence level

Degrees of freedom for the smaller variance	Degrees of freedom for the larger variance																
	1	2	3	4	5	6	7	8	9	10	15	20	30	40	60	120	∞
1	647,8	799,5	864,2	899,6	921,8	937,1	948,2	956,7	963,3	968,6	984,9	993,1	1 001	1 006	1 010	1 014	1 018
2	38,51	39,00	39,17	39,25	39,30	39,33	39,36	39,37	39,39	39,4	39,43	39,45	39,46	39,47	39,48	39,49	39,50
3	17,44	16,04	15,44	15,10	14,88	14,73	14,64	14,54	14,47	14,42	14,25	14,17	14,08	14,04	13,99	13,95	13,90
4	12,22	10,65	9,98	9,60	9,36	9,20	9,07	8,98	8,90	8,84	8,66	8,56	8,46	8,41	8,36	8,31	8,26
5	10,01	8,43	7,76	7,39	7,15	6,98	6,85	6,76	6,68	6,62	6,43	6,33	6,23	6,18	6,12	6,07	6,02
6	8,81	7,26	6,60	6,23	5,99	5,82	5,70	5,60	5,52	5,46	5,27	5,17	5,07	5,01	4,96	4,90	4,85
7	8,07	6,54	5,89	5,52	5,29	5,12	4,99	4,90	4,82	4,76	4,57	4,47	4,36	4.,31	4,25	4,20	4,14
8	7,57	6,06	5,42	5,05	4,82	4,65	4,53	4,43	4,36	4,30	4,10	4,00	3,89	3,84	3,78	3,73	3,67
9	7,21	5,71	5,08	4,72	4,48	4,32	4,20	4,10	4,03	3,96	3,77	3,67	3,56	3,51	3,45	3,39	3,33
10	6,94	5,46	4,83	4,47	4,24	4,07	3,95	3,85	3,78	3,72	3,52	3,42	3,31	3,26	3,20	3,14	3,08
11	6,72	5,26	4,63	4,28	4,04	3,88	3,76	3,66	3,59	3,53	3,33	3,23	3,12	3,06	3,00	2,94	2,88
12	6,55	5,10	4,47	4,12	3,89	3,73	3,61	3,51	3,44	3,37	3,18	3,07	2,96	2,91	2,85	2,79	2,72
13	6,41	4,97	4,35	4,00	3,77	3,60	3,48	3,39	3,31	3,25	3,05	2,95	2,84	2,78	2,72	2,66	2,60
14	6,30	4,86	4,24	3,89	3,66	3,50	3,38	3,29	3,21	3,15	2,95	2,84	2,73	2,67	2,61	2,55	2,49
15	6,20	4,77	4,15	3,80	3,58	3,41	3,29	3,20	3,12	3,06	2,86	2,76	2,64	2,59	2,52	2,46	2,40
20	5,87	4,46	3,86	3,51	3,29	3,13	3,01	2,91	2,84	2,77	2,57	2,46	2,35	2,29	2,22	2,16	2,09
25	5,69	4,29	3,69	3,35	3,13	2,97	2,85	2,75	2,68	2,61	2,41	2,30	2,18	2,12	2,05	1,98	1,91
30	5,57	4,18	3,59	3,25	3,03	2,87	2,75	2,65	2,57	2,51	2,31	2,20	2,07	2,01	1,94	1,87	1,79
40	5,42	4,05	3,46	3,13	2,90	2,74	2,62	2,53	2,45	2,39	2,18	2,07	1,94	1,88	1,80	1,72	1,64
60	5,29	3,93	3,31	3,01	2,79	2,63	2,51	2,41	2,33	2,27	2,06	1,94	1,82	1,74	1,67	1,58	1,48
120	5,15	3,80	3,23	2,89	2,67	2,52	2,39	2,30	2,22	2,16	1,94	1,82	1,69	1,61	1,53	1,43	1,31
∞	5,02	3,69	3,12	2,79	2,57	2,41	2,29	2,19	2,11	2,05	1,83	1,71	1,57	1,48	1,39	1,27	1,00

Table A.2 — Two-sided *t*-test values at 95 % confidence level

Degrees of freedom	t_{tab}	Degrees of freedom	t_{tab}	Degrees of freedom	t_{tab}
1	12,706	15	2,131	29	2,045
2	4,303	16	2,120	30	2,042
3	3,182	17	2,110	31	2,039
4	2,776	18	2,101	32	2,037
5	2,571	19	2,093	33	2,034
6	2,447	20	2,086	34	2,032
7	2,365	21	2,080	35	2,030
8	2,306	22	2,074	40	2,021
9	2,262	23	2,069	50	2,009
10	2,228	24	2,064	60	2,000
11	2,201	25	2,060	100	1,985
12	2,179	26	2,056	120	1,980
13	2,160	27	2,052	∞	1,960
14	2,145	28	2,048		

Annex B
(informative)

Example of a comparison of arithmetic means of two independently obtained sets of data

This annex gives an example of comparing the arithmetic means of two independently obtained sets of data. Methods i and j are to be compared after a number of replicate determinations are carried out on samples containing relatively similarly concentrations. The results obtained are set out in Table B.1.

Table B.1 — Results

Replicates	Method i	Method j	Difference
1	4,0	4,4	
2	4,8	4,7	
3	4,0	4,3	
4	5,0	4,7	
5	4,6	4,9	
6	4,7	4,7	
7	4,2	4,5	
8	4,9	5,0	
9	3,9	4,6	
10	4,8	4,5	
Mean	4,49	4,63	0,14
Standard deviation	0,420	0,216	

From these results, the following can be calculated:

$F_{calc} = 0{,}420^2/0{,}216^2 = 3{,}77$

With (10 − 1) degrees of freedom for method i and (10 − 1) degrees of freedom for method j

$F_{tab} = 4{,}03$

Since F_{tab} is greater than F_{calc}, it can be concluded that there is no statistically significant difference between the two variances. Hence, a combined standard deviation can be calculated. Thus

$s_c = \{[0{,}420^2 \times (10 - 1) + 0{,}216^2 \times (10 - 1)]/(10 + 10 - 2)\}^{½} = 0{,}334$

Using this value, $t_{calc} = |4{,}63 - 4{,}49|/[0{,}334\{(1/10) + (1/10)\}^{½}] = 1{,}447$.

Using Table A.2, with (10 + 10 − 2) degrees of freedom, $t_{tab} = 2{,}101$.

Since t_{tab} is greater than t_{calc} it can be concluded that there is no statistically significant difference between the two arithmetic means. The results can, therefore, be regarded as being equivalent.

Thus, a statistically significant difference would not be demonstrated using these two methods until the difference in the mean values exceeded 0,295.

If the standard deviation of both methods is calculated from all 20 results (i.e. s_{ij} = 0,333), and this value is used to calculate t_{calc}, a value of 0,940 is produced. Comparing this value with a t_{tab} value of 2,093 from Table A.2 with (20 − 1) degrees of freedom, the same interpretation is obtained; i.e. it is concluded that there is no statistically significant difference between the mean values of both methods.

Annex C
(informative)

Example of a comparison of population and sample arithmetic means

This annex gives an example of comparing population and sample arithmetic means.

A long-serving method with a mean for quality control samples of 22,7 is to be compared with a new method producing a corresponding mean of 23,5 from 10 results with a standard deviation of 0,9. For this comparison, data are obtained using samples with similar concentrations that are analysed by a number of different analysts for both methods. Thus

$$t_{calc} = |23,5 - 22,7|/[0,9/(10)^{1/2}] = 2,81$$

Using Table A.2, with (10 − 1) degrees of freedom, $t_{tab} = 2,262$.

Since t_{tab} is less than t_{calc}, it can be concluded that there is a statistically significant difference between the two arithmetic means and the results cannot be regarded as equivalent.

Thus, the difference in the mean values should not exceed 0,64 for it to be concluded that there was no statistically significant difference.

Annex D
(informative)

Example of an analysis of variance

This annex gives an example of comparing the results from a number of different analysts.

A new method is to be considered for the replacement of an old method. Before this can proceed, the new method is used by a number of different analysts to establish its capability. Each analyst carries out replicate determinations and these results are compared to ascertain if there are any statistically significant differences. The results obtained are set out in Table D.1.

Table D.1 — Results

Replicates	Analysts						
	1	**2**	**3**	**4**	**5**	**6**	
1	52	55	51	53	54	51	
2	49	54	52	55	52	53	
3	50	53	53	58	58	57	
	151	162	156	166	164	161	Total = 960

The corresponding values of A, B and C [see Equations (7) to (9)] are:

A = $(151^2/3) + (162^2/3) + (156^2/3) + (166^2/3) + (164^2/3) + (161^2/3)$
= 51 251,33

B = $(52^2 + 49^2 + 50^2) + (55^2 + 54^2 + 53^2) + (51^2 + 52^2 + 53^2) +$
$(53^2 + 55^2 + 58^2) + (54^2 + 52^2 + 58^2) + (51^2 + 53^2 + 57^2)$
= 51 310

C = $960^2/(3 \times 6)$
= 51 200

From these values, an ANOVA table is calculated as shown in Table D.2.

Table D.2

Sum of squares	Degrees of freedom	Mean square	F_{calc}
$S_1 = A - C$ = 51,33	$p - 1$ = 5	$M_1 = S_1 / (p - 1)$ = 10,266	(10,266/4,889)
$S_0 = B - A$ = 58,67	$N - p$ = 12	$M_0 = S_0 / (N - p)$ = 4,889	2,1
$S_1 + S_0 = B - C$ = 110	$N - 1$ = 17		

From Table A.1, with (6 – 1) degrees of freedom for the larger mean square and (18 – 6) degrees of freedom for the smaller mean square, F_{tab} = 3,89.

Since F_{tab} is greater than F_{calc}, then it can be concluded that there is no statistically significant difference between the sets of data. Hence, all the sets of data are similar, and can be considered equivalent. In this case, all the results can be combined and overall mean, standard deviation and variance values calculated.

The procedure described in Annex C can then be used to compare results from two methods.

Annex E
(informative)

Examples of a comparison of results from samples of different matrices

E.1 Introduction

When using the orthogonal regression or difference method technique, the precision of both methods in matrix-free standard solutions should be regarded as being equivalent for the two analytical methods. This can be shown by using the F-test.

E.2 Orthogonal regression

E.2.1 Example

Table E.1 shows the analytical results of 35 different real samples that have been analysed using two different analytical methods.

Table E.1 — Analytical results of analytical method (x_i) and alternative analytical method (y_i)

i	x_i, mg/l	y_i, mg/l	$y_i / x_i = Q_i$	i	x_i, mg/l	y_i, mg/l	$y_i / x_i = Q_i$
1	0,27	0,5	1,85	19	1,8	2,0	1,11
2	0,29	0,6	2,07	20	2,0	2,1	1,05
3	0,45	1,0	2,22	21	2,1	2,6	1,24
4	0,48	0,8	1,67	22	2,2	2,9	1,32
5	0,49	1,1	2,24	23	2,3	2,3	1,0
6	0,73	1,1	1,51	24	2,3	2,7	1,17
7	0,74	1,0	1,35	25	2,5	1,6	0,64 [a]
8	0,79	1,3	1,65	26	2,6	2,9	1,12
9	0,81	3,8	4,69 [a]	27	2,7	3,6	1,33
10	1,0	1,4	1,40	28	2,7	3,3	1,22
11	1,1	1,5	1,36	29	2,8	3,3	1,18
12	1,2	1,8	1,50	30	2,8	3,3	1,18
13	1,3	1,5	1,15	31	3,3	3,6	1,09
14	1,3	2,2	1,69	32	3,4	3,9	1,15
15	1,4	2,0	1,43	33	3,4	4,0	1,18
16	1,6	2,3	1,44	34	3,4	4,0	1,18
17	1,7	1,9	1,12	35	3,5	3,8	1,09
18	1,8	1,6	0,89				
	s_x = 1,011	s_y = 1,060					
	$\bar{x}$ = 1,836	$\bar{y}$ = 2,221					

[a] See text in E.2.2.

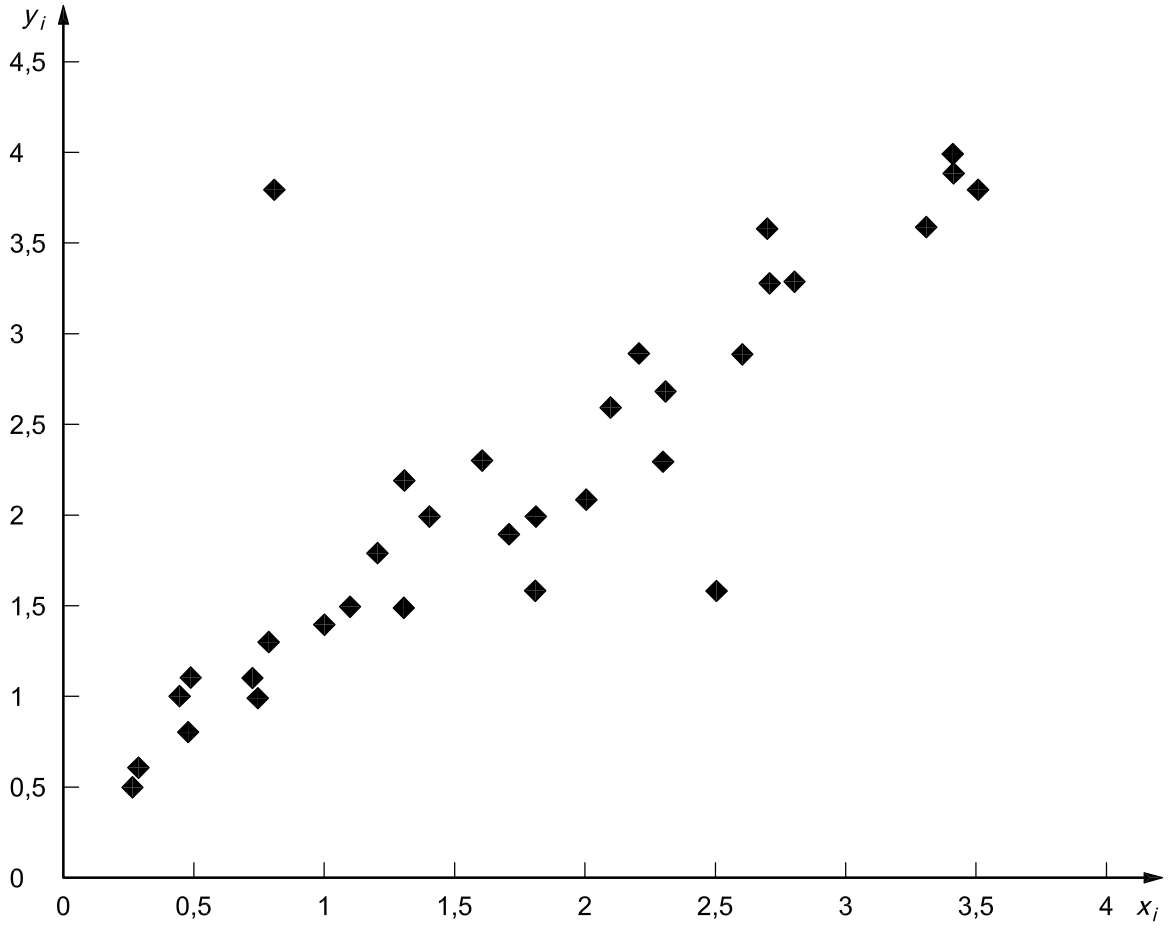

Figure E.1 — Graph of the data pairs listed in Table E.1

E.2.2 Outliers

Outliers are found using Grubbs-test for quotients of single pairs of values for data pair 9, where Q_i is 4,69 and the calculated test statistic (PG) from Equations (E.1) to (E.4) is 4,43. This value is greater than the tabulated value (2,811) based on 95 % confidence level and 35 degrees of freedom, indicating an outlier which is rejected from the rest of the data and excluded from further calculations when means and standard deviations are re-calculated. Using the same procedure, data pair 25 is shown not to be an outlier.

$$Q_i = \frac{y_i}{x_i} \tag{E.1}$$

$$\overline{Q} = \frac{\sum Q_i}{N} \tag{E.2}$$

$$s_Q = \sqrt{\frac{\sum \left(Q_i - \overline{Q}\right)^2}{N-1}} \tag{E.3}$$

$$\text{PG} = \frac{\left|Q^* - \overline{Q}\right|}{s_Q} \tag{E.4}$$

E.2.3 Orthogonal regression

slope: b = 1,049

intercept: a = 0,294

Test for proportionally systematic deviations

s = 1,035 6

s_{xy} = 1,019 3

χ^2_{calc} = 0,807 8

χ^2_{tab} = 3,8 (from appropriate tables based on 95 % confidence level and 1 degree of freedom)

As χ^2_{calc} is less than 3,8, no proportional systematic deviation is detectable.

Test for constant-systematic deviations

$\bar{d}$ = -0,384

s_d = 0,325 9

t_{calc} = 6,873

t_{tab} = 2,03 (from t-test tables with 95 % confidence and 33 degrees of freedom)

As t_{calc} is greater than 2,03, a constant-systematic deviation was detected.

Thus, it can be concluded that the results obtained using the two methods cannot be regarded as being equivalent.

E.3 Difference method

E.3.1 Difference method where concentration levels may be different

E.3.1.1 Example

Table E.2 shows the analytical results from 30 different real samples which have been analysed using two analytical methods.

Table E.2 — Analytical results of method (x_i) and alternative analytical method (y_i)

i	x_i mgl/	y_i mg/l	d_i mg/l	i	x_i mg/l	y_i mg/l	d_i mg/l
1	1,03	1,07	–0,04	16	2,00	2,29	–0,29
2	1,18	1,34	–0,16	17	2,08	2,55	–0,47[a]
3	1,20	1,22	–0,02	18	2,13	2,18	–0,05
4	1,21	1,30	–0,09	19	2,19	2,12	0,07
5	1,27	1,33	–0,06	20	2,26	2,65	–0,39
6	1,34	1,31	0,03	21	2,35	2,61	–026
7	1,36	1,47	–0,11	22	2,48	2,41	0,07
8	1,42	1,66	–0,24	23	2,55	2,73	–0,18
9	1,43	1,57	–0,14	24	2,75	2,8	–0,05
10	1,58	1,70	–0,12	25	2,76	2,44	0,32
11	1,69	1,96	–0,27	26	2,89	2,95	–0,06
12	1,75	1,69	0,06	27	2,99	2,91	0,08
13	1,78	1,52	0,26	28	3,01	3,18	–0,17
14	1,96	1,92	0,04	29	3,1	3,05	0,05
15	1,99	1,91	0,08	30	3,16	3,21	–0,05
			$\bar{d}$ = –0,07				
			s_d = 0,172				

[a] See text in E.3.1.2.

E.3.1.2 Outliers

Outliers are determined using the Grubbs-test for data pair 17 where d_i is –0,47 (the highest difference) and the calculated test statistic (PG) from Equations E.5 to E.7 is 2,313. This value is less than the tabulated value (2,745) based on 95 % confidence level and 30 degrees of freedom, indicating no outlier.

$$\bar{d} = \frac{\sum d_i}{N} \tag{E.5}$$

$$s_d = \sqrt{\frac{\sum \left(d_i - \bar{d}\right)^2}{N-1}} \tag{E.6}$$

$$\mathrm{PG} = \frac{\left|d^* - \bar{d}\right|}{s_d} \tag{E.7}$$

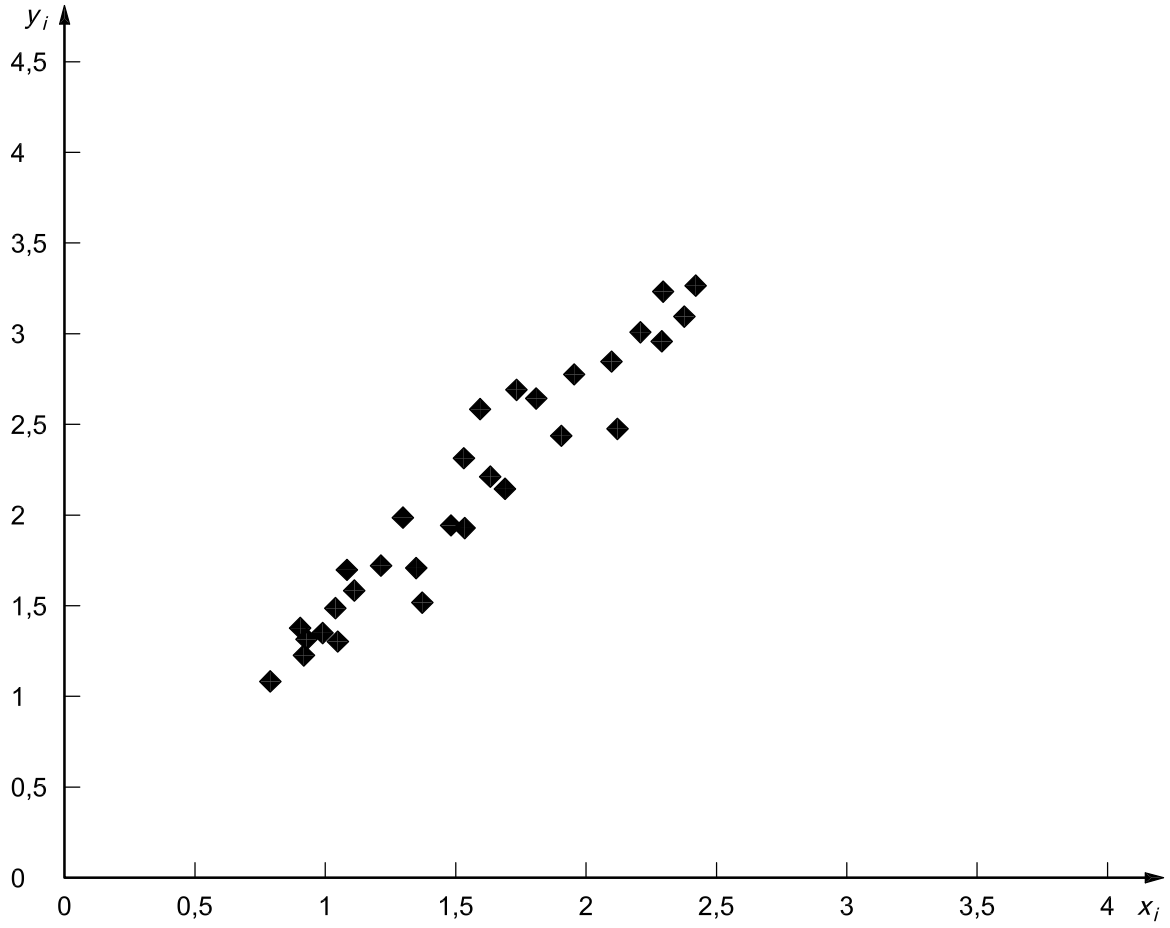

Figure E.2 — Graph of the data pairs listed in Table E.2

E.3.1.3 Paired *t*-test

t_{calc} = 2,292

t_{tab} = 2,04 (from tables based on 95 % confidence with 30 degrees of freedom)

As t_{calc} is greater than t_{tab}, there is a statistically significant difference (at the 95 % confidence level) between the results using both methods.

Thus, it can be concluded that the results from both methods cannot be regarded as being equivalent.

E.3.2 Difference method where concentration levels are similar

As an example of comparing arithmetic means from paired results, consider the following. Two methods are to be compared, see Table E.3. Each sample (comprising a similar matrix) is divided into two aliquots and each aliquot is analysed by one of the methods. The samples would need to contain similar concentrations and be analysed by different analysts.

Table E.3 — Results

Sample	Method 1	Method 2	Difference
1	2,52	3,17	−0,65
2	3,13	5,00	−1,87
3	4,33	4,03	0,30
4	2,25	2,38	−0,13
5	2,79	3,68	−0,89
6	3,04	2,94	0,10
7	2,19	2,83	−0,64
8	2,16	2,18	−0,02
Mean of differences			−0,475
Standard deviation of the differences			0,700

and

t_{calc} = |0,475|/[0,700/(8)$^{1/2}$]

= 1,919

From Table A.2, with (8 − 1) degrees of freedom, t_{tab} = 2,365.

Since t_{tab} is greater than t_{calc}, it can be concluded that there is no statistically significant difference between the mean of the differences and zero, and hence the results can be regarded as being equivalent.

Thus the absolute mean of the differences should exceed 0,585 for it to be concluded that there was a statistically significant difference.

For an alternative approach, if the results from both methods were treated as described in Clause 7 (Annex B), then a similar conclusion would be reached.

Annex F
(informative)

Illustrative examples of graphic plots

These sets of data are for illustrative purposes only and represent results obtained using two methods, a) and b), where statistical comparisons are undertaken in order to establish whether the two sets of results, in terms of means and variances, can be regarded as being equivalent or not. Whilst statistical treatments indicate a pass/fail or positive/negative situation, an interpretation of the data comparison should always be undertaken with caution, bearing in mind the importance of the decision to be made and the resulting implications.

For the two sets of results shown in Figure F.1, a statistical comparison (using a t-test) of both mean values might show that $t_{tab} > t_{calc}$ indicating that there is no statistically significant difference between the means of both methods. Hence, the two mean values are similar and can be regarded as being equivalent. (The bias values of both methods are thus similar.) A statistical comparison (using the F-test) of the variances of both methods would need to be carried out in order to establish whether the precision or the spread of data, in terms of the standard deviations or variances, can be considered equivalent or not. However, there might be insufficient data to draw a positive conclusion.

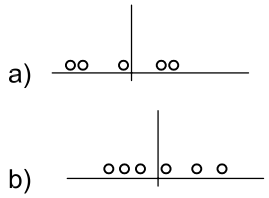

Figure F.1

For the two sets of results shown in Figure F.2, a statistical comparison (using a t-test) of both mean values can show that $t_{tab} < t_{calc}$ indicating that there is a statistically significant difference. Hence, the two mean values are different and cannot, therefore, be regarded as being equivalent.

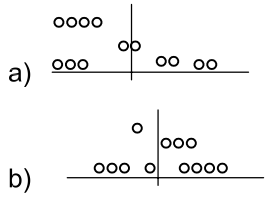

Figure F.2

The mean values of the two sets of results shown in Figure F.3 might or might not be different. Although a statistical conclusion would be obtained, judgement would need to be made on whether this was meaningful or not. The amount of data, i.e. the number of results from both methods, is probably not sufficient to reach a positive interpretation.

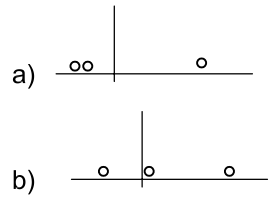

Figure F.3

For the two sets of results shown in Figure F.4, a statistical comparison of both mean values might show that $t_{tab} < t_{calc}$ indicating that there is a statistically significant difference and the two mean values are different. However, whether this statistical difference is important or meaningful should be questioned, in light of the amount of data available and the purpose to which the results can be used.

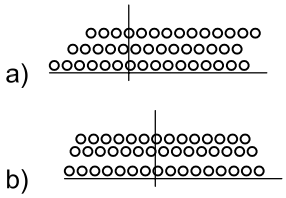

Figure F.4

For the precision or spread of data, in terms of variance or standard deviation, for the two sets of results shown in Figure F.5, a statistical comparison (using the F-test) of both variances might show that $F_{tab} > F_{calc}$ indicating that there is no statistically significant difference, and the precision or spread of data is not different for both sets of results. If no statistical difference is indicated then the standard deviations from both methods can be combined to produce a single (pooled) standard deviation. A statistical comparison (using a t-test) of both mean values will then indicate whether the mean values from both methods are similar and can be regarded as equivalent or not.

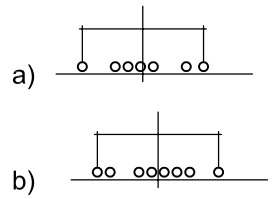

Figure F.5

For the precision or spread of data, in terms of variance or standard deviation, for the two sets of results shown in Figure F.6, a statistical comparison (using the F-test) of both variances might show that $F_{tab} < F_{calc}$ indicating that there is a statistically significant difference and the spread of data is different for both sets of results. However, a statistical comparison (using a t-test) of the mean values for both sets of results might or might not indicate a statistically significant difference between the means. If a statistical comparison of the mean values for both methods shows that there is a significant difference between the means, then the results cannot be considered equivalent. If a statistical comparison of the mean values for both methods shows that there is no significant difference between the means, an interpretation of the data could be that whilst the results are not equivalent, one method is more precise than the other method.

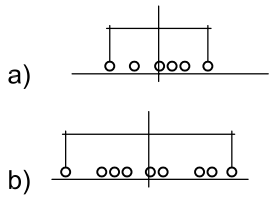

Figure F.6

The spread of data, in terms of standard deviation or variance, for the two sets of results shown in Figure F.7 might or might not be different. Although a statistical conclusion would be obtained, judgement would need to be made on whether this was meaningful or not. The amount of data, i.e. the number of results from both methods, is probably not sufficient to reach a positive interpretation.

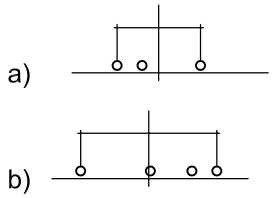

Figure F.7

Annex G
(informative)
Schematic diagrams

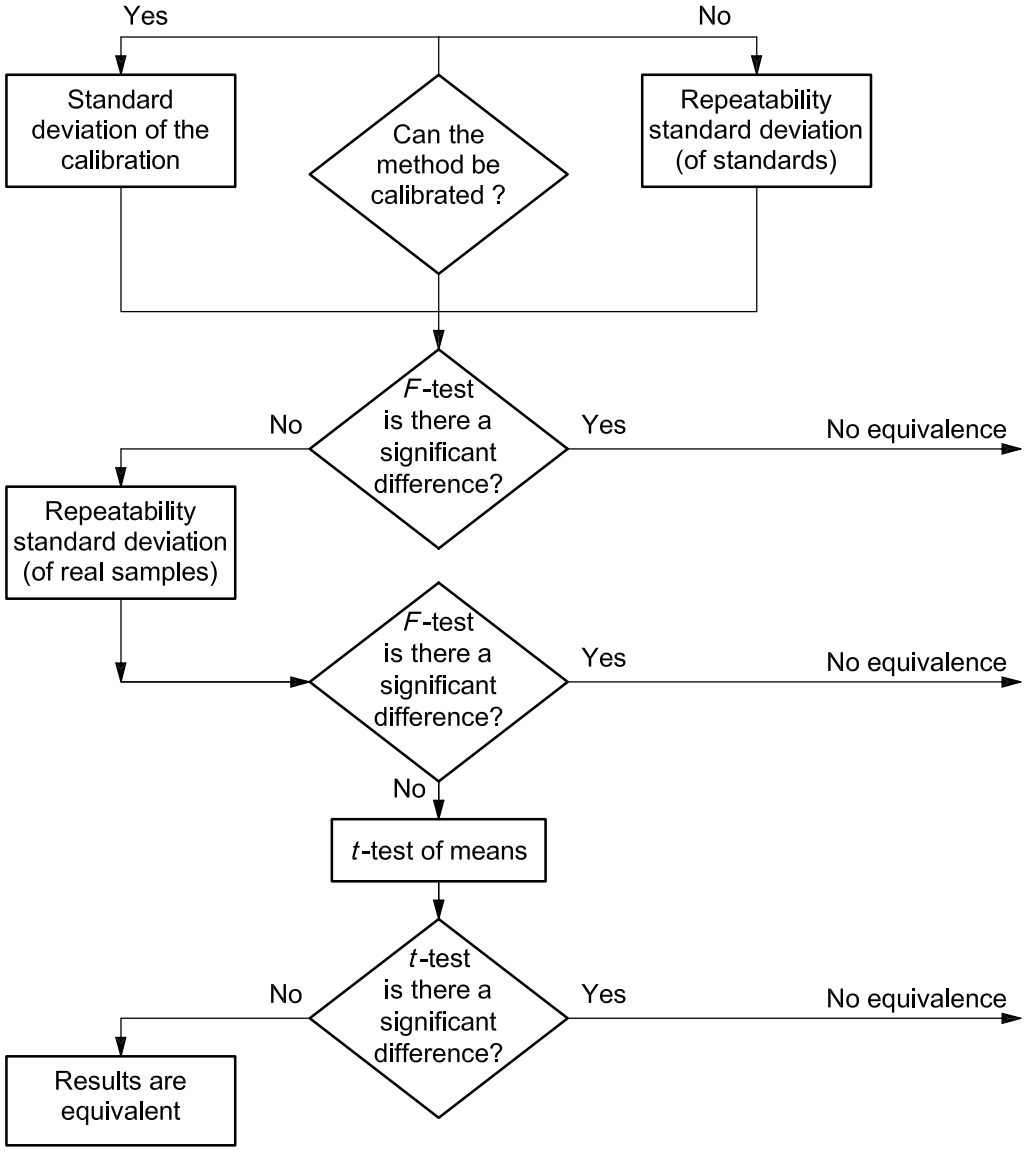

Figure G.1 — Schematic diagram where concentration levels are similar

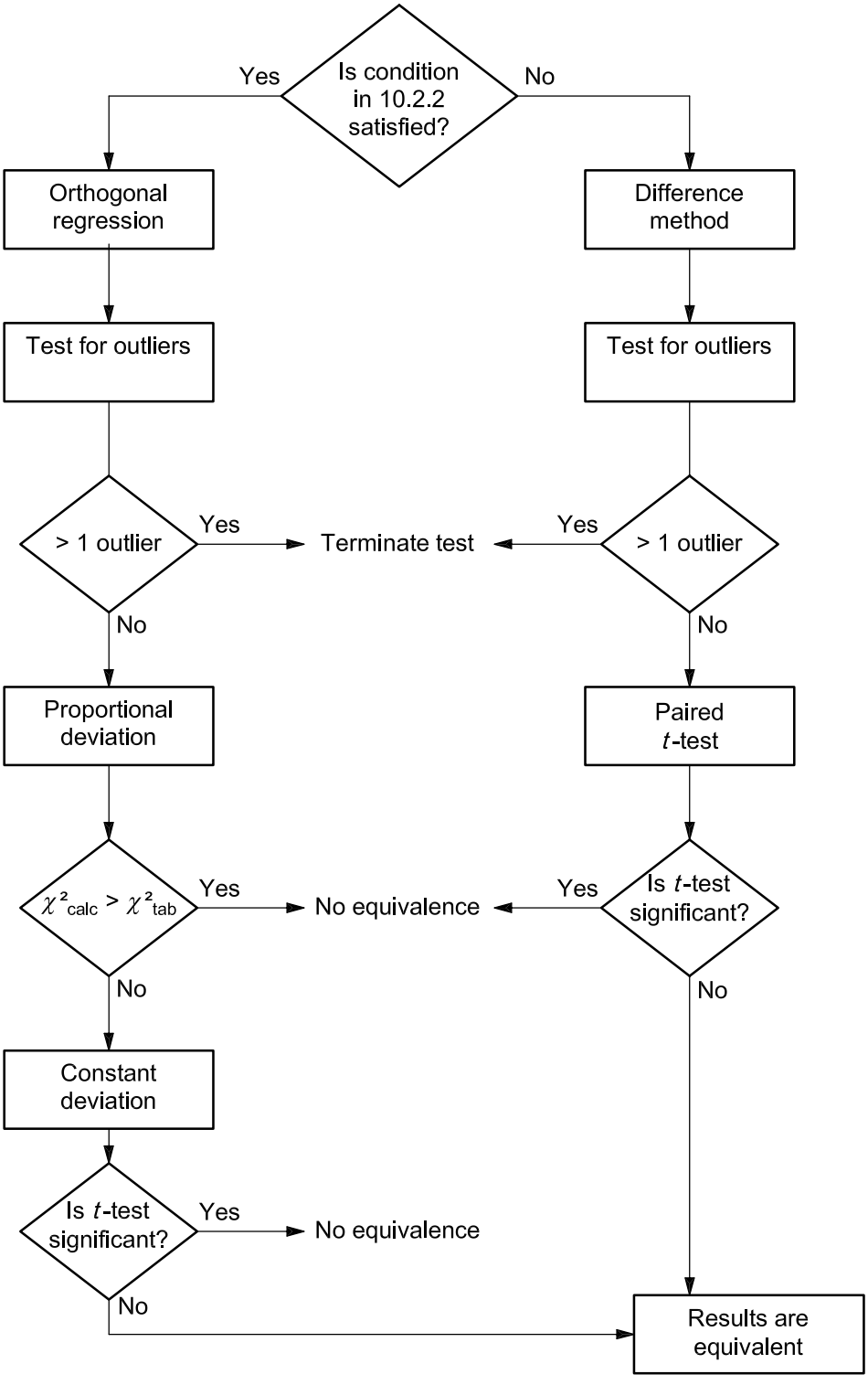

Figure G.2 — Schematic diagram for equivalency test for different matrices

Bibliography

[1] ISO/TR 22971, *Accuracy (trueness and precision) of measurement methods and results — Practical guidance for the use of ISO 5725-2:1994 in designing, implementing and statistically analysing interlaboratory repeatability and reproducibility results*

[2] Statistics in Context, *VAM Bulletin*, (1997) **16**, pp. 20-22

[3] Statistics in Context, *VAM Bulletin*, (1997) **17**, pp. 18-21

[4] Statistics in Context, *VAM Bulletin*, (1998) **18**, pp. 18-21

[5] Statistics in Context, *VAM Bulletin*, (1998) **19**, pp. 22-27

[6] Statistics in Context, *VAM Bulletin*, (1999) **20**, pp. 28-31

[7] Statistics in Context, *VAM Bulletin*, (1999) **21**, pp. 28-32

[8] BURKE, S. *Scientific Data Management*, (1997) **1**(1) pp. 32-38

[9] BURKE, S. *Scientific Data Management*, (1998) **2**(1) pp. 36-41

[10] BURKE, S. *Scientific Data Management*, (1998) **2**(2) pp. 33-40

[11] BURKE, S. *Scientific Data Management*, (1998) **2**(3) pp. 28-35

ICS 13.060.45

Price based on 29 pages