INTERNATIONAL STANDARD

ISO 8196-1

IDF 128-1

Second edition
2009-10-01

Milk — Definition and evaluation of the overall accuracy of alternative methods of milk analysis —

Part 1: Analytical attributes of alternative methods

Lait — Définition et évaluation de la précision globale des méthodes alternatives d'analyse du lait —

Partie 1: Attributs analytiques des méthodes alternatives

Reference numbers
ISO 8196-1:2009(E)
IDF 128-1:2009(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

International Dairy Federation
Diamant Building • Boulevard Auguste Reyers 80 • B-1030 Brussels
Tel. + 32 2 733 98 88
Fax + 32 2 733 04 13
E-mail info@fil-idf.org
Web www.fil-idf.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.

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 8196-1|IDF 128-1 was prepared by Technical Committee ISO/TC 34, *Food products*, Subcommittee SC 5, *Milk and milk products*, and the International Dairy Federation (IDF). It is being published jointly by ISO and IDF.

This second edition of ISO 8196-1|IDF 128-1 cancels and replaces the first edition (ISO 8196-1:2000) which has been technically revised.

ISO 8196|IDF 128 consists of the following parts, under the general title *Milk — Definition and evaluation of the overall accuracy of alternative methods of milk analysis*:

— *Part 1: Analytical attributes of alternative methods*

— *Part 2: Calibration and quality control in the dairy laboratory*

— *Part 3: Protocol for the evaluation and validation of alternative quantitative methods of milk analysis*

Foreword

IDF (the International Dairy Federation) is a non-profit organization representing the dairy sector worldwide. IDF membership comprises National Committees in every member country as well as regional dairy associations having signed a formal agreement on cooperation with IDF. All members of IDF have the right to be represented at the IDF Standing Committees carrying out the technical work. IDF collaborates with ISO in the development of standard methods of analysis and sampling for milk and milk products.

The main task of Standing Committees is to prepare International Standards. Draft International Standards adopted by the Action Teams and Standing Committees are circulated to the National Committees for voting. Publication as an International Standard requires approval by at least 50 % of IDF National Committees casting a vote.

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

ISO 8196-1|IDF 128-1 was prepared by the International Dairy Federation (IDF) and Technical Committee ISO/TC 34, *Food products*, Subcommittee SC 5, *Milk and milk products*. It is being published jointly by ISO and IDF.

All work was carried out by the Joint IDF-ISO Action Team on *Automated methods* of the Standing Committee on *Quality assurance, statistics of analytical data and sampling* under the aegis of its project leader, Mr. O. Leray (FR).

This edition of ISO 8196-1|IDF 128-1, together with ISO 8196-2|IDF 128-2 and ISO 8196-3|IDF 128-3, cancels and replaces IDF 128:1985, which has been technically revised.

ISO 8196|IDF 128 consists of the following parts, under the general title *Milk — Definition and evaluation of the overall accuracy of alternative methods of milk analysis*:

— *Part 1: Analytical attributes of alternative methods*

— *Part 2: Calibration and quality control in the dairy laboratory*

— *Part 3: Protocol for the evaluation and validation of alternative quantitative methods of milk analysis*

Introduction

The main purpose of this part of ISO 8196|IDF 128 is to provide definitions of the relevant performance characteristics for quantitatively evaluating the overall accuracy of an analytical method, through the application of proper experimental designs and recommended statistical procedures.

Performance characteristics of an analytical method can be defined as a set of quantitative and experimentally determined values, or criteria, of fundamental importance in assessing the suitability of a method for any given purpose. The general concepts apply to all analytical methods, but special emphasis is given to rapid physico-chemical methods which are currently in use for compositional testing of milk.

In analytical methods where measurements result from combinations of multiple output signals of measurement channels either in series or in parallel (e.g. methods in which multivariate mathematical models are applied), the process of combining the primary raw information is considered as a full part of the method itself. For the purpose of ISO 8196|IDF 128 (all parts), this process is considered as a closed device ("black box"). As such, this process is assumed to be optimized prior to the assessments and evaluations done within the scope of ISO 8196|IDF 128 (all parts).

ISO 8196-2|IDF 128-2 provides practical details and recommendations for the calibration of instruments and quality control in routine dairy laboratories including checking compliance with a specification value or limit.

ISO 8196-3|IDF 128-3 is intended to complement this part of ISO 8196|IDF 128 as an alternative to the evaluation of new methods to which this part of ISO 8196|IDF 128 cannot apply, e.g. when the organization of interlaboratory studies is hampered by the number of new instruments available, which is too small for such a protocol.

While this part of ISO 8196|IDF 128 and ISO 8196-3|IDF 128-3 are mainly intended for experts to assess new methods of analysis, ISO 8196-2|IDF 128-2 aims to be a guide for routine laboratories using these methods.

ISO 8196|IDF 128 (all parts) only specifies the single linear regression model as a simplified approach to allow users to determine equivalence of an alternative method with a reference method. However, the linear regression approach is valid as a determination of method equivalence only in limited circumstances or if a high correlation between the results of the reference method and the routine method is achieved. If a high correlation is not achieved, recourse should be made to other data handling and measurement error modelling techniques. Although these techniques are referred to, they are not specified in ISO 8196|IDF 128 (all parts).

Milk — Definition and evaluation of the overall accuracy of alternative methods of milk analysis —

Part 1: Analytical attributes of alternative methods

1 Scope

This part of ISO 8196|IDF 128 specifies various performance characteristics that constitute and serve to characterize the overall accuracy of an analytical method. It furthermore establishes general principles for the design of experiments and gives guidelines for the procedures to be used to evaluate these characteristics quantitatively.

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 (all parts), *Statistics — Vocabulary and symbols*

ISO 5725 (all parts), *Accuracy (trueness and precision) of measurement methods and results*

3 Terms and definitions

For the purposes of this part of ISO 8196|IDF 128, the terms and definitions given in ISO 3534 (all parts) and ISO 5725-1 apply, together with the following.

3.1 General terms and definitions

3.1.1
true quantity value
true value of a quantity
true value
quantity value consistent with the definition of a quantity

NOTE 1 In the Error Approach to describing measurement, a true quantity value is considered unique and, in practice, unknowable. The Uncertainty Approach is to recognize that, owing to the inherently incomplete amount of detail in the definition of a quantity, there is not a single true quantity value but rather a set of true quantity values consistent with the definition. However, this set of values is, in principle and in practice, unknowable. Other approaches dispense altogether with the concept of true quantity value and rely on the concept of metrological compatibility of measurement results for assessing their validity.

NOTE 2 In the special case of a fundamental constant, the quantity is considered to have a single true quantity value.

NOTE 3 When the definitional uncertainty associated with the measurand is considered to be negligible compared to the other components of the measurement uncertainty, the measurand may be considered to have an "essentially unique" true quantity value. This is the approach taken by ISO/IEC Guide 98-3:2008[3] and associated documents, where the word "true" is considered to be redundant.

[ISO/IEC Guide 99:2007[4], 2.11]

3.1.2
reference method
anchor method
method of analysis internationally recognized by experts or by agreement between the parties

NOTE 1 Adapted from ISO 21187|IDF 196:2004[1], 3.2.

NOTE 2 A reference method gives the "true value" or "assigned value" of the quantity of the measurand.

3.1.3
alternative method
routine method
method of analysis allowing quantification of the status of a test sample

NOTE 1 Adapted from ISO 21187|IDF 196:2004[1], 3.1.

NOTE 2 An alternative method demonstrates or estimates, for a given category of products, the same measurand as determined using the corresponding **reference method** (3.1.2).

NOTE 3 The alternative method can be either an indirect method — i.e. one not measuring directly the component or the characteristic that it is intended to quantify, but instead one or more quantities or properties which are functionally linked to that component — or a direct method. It can have specific adaptations for user convenience (e.g. speed, automation, miniaturization, cost) that can introduce deviation into the analytical process (e.g. incomplete component or characteristic measurement) and thus prevent direct estimation and give different accuracy.

3.2 Terms and definitions on precision

3.2.1
precision
closeness of agreement between independent test/measurement results obtained under stipulated conditions

[ISO 3534-2:2006, 3.3.4]

3.2.2
repeatability limit
r
value less than or equal to which the absolute difference between two final values, each of them representing a series of test results or measurement results obtained with the same method on identical test/measurement items in the same test or measuring facility by the same operator using the same equipment within short intervals of time, is expected to be for a specified probability of 95 %

NOTE Adapted from ISO 3534-2:2006, 3.3.6, 3.3.8, 3.3.9.

3.2.3
reproducibility limit
R
value less than or equal to which the absolute difference between two final values, each of them representing a series of test results or measurement results obtained with the same method on identical test/measurement items in different test or measuring facilities with different operators using different equipment, is expected to be for a specified probability of 95 %

NOTE Adapted from ISO 3534-2:2006, 3.3.11, 3.3.13, 3.3.14.

3.3 Terms and definitions on accuracy

The terms and definitions on accuracy apply specifically to indirect alternative methods for the purpose of this part of ISO 8196|IDF 128.

3.3.1
accuracy
closeness of agreement between a test result or measurement result and the **true value** (3.1.1)

NOTE 1 In practice, the accepted reference value is substituted for the true value.

NOTE 2 The term accuracy, when applied to a set of test or measurement results, involves a combination of random components and a common systematic error or bias component.

NOTE 3 Accuracy refers to a combination of **trueness** (3.3.2) and **precision** (3.2.1).

[ISO 3534-2:2006, 3.3.1]

3.3.2
trueness
closeness of agreement between the expectation of a test result or a measurement result and a **true value** (3.1.1)

NOTE 1 The measure of trueness is usually expressed in terms of bias.

NOTE 2 Trueness is sometimes referred to as “accuracy of the mean”. This usage is not recommended.

NOTE 3 In practice, the accepted reference value is substituted for the true value.

[ISO 3534-2:2006, 3.3.3]

3.3.3
exactness of calibration
closeness of agreement, at each level of the measurand, between an alternative method value and the estimated average true value for all individual samples at the corresponding level

3.3.4
accuracy of estimate
closeness of agreement between the average test result obtained by the reference method and an alternative method on identical materials, provided that the calibration of the alternative method is exact

NOTE Accuracy of estimate measures the part of the systematic error not due to error in calibration.

3.4 Other analytical characteristics

3.4.1
selectivity
property of a method to respond exclusively to the characteristic or analyte to be measured, or degree to which a method can quantify that characteristic or analyte accurately in the presence of interferents

NOTE As a general principle, selectivity is sufficiently good for any interference to be ignored. Selectivity contributes to trueness.

3.4.2
sensitivity
smallest change in concentration which can be measured by an analytical procedure

NOTE Sensitivity is calculated as the ratio of the variation of the method response to the variation of analyte concentration. As this is usually arbitrary, depending on instrumental settings, it is not useful in validation. However, it may be useful in quality assurance procedures to test whether an instrument is performing to a consistent and satisfactory standard.

3.4.3
limit of detection
LOD
minimum amount or concentration of the analyte in a test sample which can be detected reliably but not necessarily quantified, as demonstrated by a collaborative trial or other appropriate validation

[ISO 24276:2006[2], 3.1.6]

NOTE For instrumental methods, sensitivity and limit of detection are usually determined by the sensitivity of the detector and the signal/noise ratio.

3.4.4
limit of determination
limit of quantitation
LOQ
〈analytical procedure〉 lowest concentration or amount of the analyte in a test sample which can be quantitatively determined with an acceptable level of precision and accuracy, as demonstrated by a collaborative trial or other appropriate validation

[ISO 24276:2006[2], 3.1.7]

4 Explanation of the definitions

4.1 Accuracy — General description

The concept applies mainly to the test result of an alternative method calibrated against the accepted value of the reference method or standard material, or when the true value of the component concentration is known.

The accuracy is an index indicating a value of the amount of errors involved, and is ordinarily expressed as the error associated with the method used and calculated under appropriate conditions.

When a single quantitative measurement, x_i, of a specific measurand (or variable) is made with a given method of analysis, that measurement is always an estimate of its true value, μ. The error of the method is given by the difference, $x_i - \mu$. The accuracy is best when the difference, $x_i - \mu$ is the smallest.

Basically, the aforementioned difference depends on the following major analytical characteristics of the method:

a) precision;

b) trueness;

c) selectivity;

d) sensitivity;

e) limit of detection and limit of determination.

Only the precision and the trueness are considered in this part of ISO 8196|IDF 128.

4.2 Precision

4.2.1 General description

Precision is a general characteristic applicable to all analytical methods. Basically, it covers all types of fortuitous and random errors which cannot be completely avoided and whose main characteristics vary from one test to another (volume delivered by a pipette, environmental conditions, stability of an instrument, electronic noise, etc.). Mistakes, such as misreading or operational failings or, more generally, any value found as outlier with the appropriate tests but worth considering, are not included in precision data.

Obviously, the variability between test results is small when tests are performed within a laboratory under repeatability conditions. It is larger when tests are performed by different laboratories under reproducibility conditions. The latter may be expressed as interlaboratory reproducibility, when only variability is considered.

In order to give quantitative measures of the variability between results under these two extreme situations, precision is expressed in terms of repeatability and reproducibility. Many intermediate conditions are conceivable, e.g. day-to-day variations, between-instrument or operator variations within the same laboratory. Results obtained using different operators and different equipment within the same facility can be used to calculate intra-laboratory reproducibility. Repeatability and reproducibility have been found to deal with most practical cases.

In practice:

a) two single test results obtained within a laboratory under repeatability conditions should be considered suspect if they differ by more than the repeatability limit, r;

b) two single test results obtained by two laboratories under reproducibility conditions should be considered suspect if they differ by more than the reproducibility limit, R.

4.2.2 Mathematical expressions

Derived from the analysis of variance (ANOVA) for data obtained through an interlaboratory trial (see 5.2), the repeatability and reproducibility of a method are expressed for a given range of concentrations of the analyte by:

a) the standard deviation of repeatability, σ_r;

b) the standard deviation of reproducibility, σ_R.

Use of the coefficient of variation, C_V, expressed as a percentage, and given by

$$C_V = \frac{\sigma}{\mu} \times 100 \tag{1}$$

where

σ is the standard deviation;

μ is the mean;

is only recommended whenever the standard deviation varies proportionally with the level of the measurand.

4.3 Trueness

4.3.1 General description

Trueness refers to that portion of the overall error associated with a method at a particular level of the measurand that is not due to the random error of measurement, but is attributable to known or unknown factors which prevent the measurement from reaching the true value.

Sources of the total systematic error or bias for alternative methods, and especially for instrumental methods for milk analysis, may arise from systematic errors in calibration and variation in the physicochemical form of the measured component, or from the influence of interfering factors. This latter source of error is called the "matrix effect".

Therefore, according to the origin of the error and the ability to eliminate the source of the error by adjustment of the calibration, the trueness is split into two components: exactness of calibration and accuracy of estimate or "accuracy". The split into components serves two useful purposes: a) to enable a valid comparison of alternative methods, and b) to give a precise figure for the analytical performance requirements of instruments, and especially for the tolerances against the reference method.

When the adjustment of the calibration is the full responsibility of the operator and, therefore, can be optimized, the exactness of calibration is not considered as a performance characteristic of an alternative method.

However, under practical conditions, it is important to know exactly the actual trueness of the method, i.e. the extent of the average bias and the degree of uncertainty of individual results, whatever the origin of the systematic error may be. An example of the calculation and the evaluation of precision and accuracy parameters is presented in ISO 8196-2|IDF 128-2.

4.3.2 Mathematical expressions

The trueness is calculated from the algebraic differences, d_i, between the measured and true values for a population of samples, and is usually expressed by:

a) the mean of differences or bias, $\overline{d}$;

b) the standard deviation of differences, σ_d.

Exactness of calibration: assuming that the relationship between the average of the reference method values, $\overline{y}_i$, and that of the alternative method results, $\overline{x}_i$, is linear, the actual calibration of an instrument is considered exact if the slope, b, of the linear regression equation, $\overline{y}_i = b\,\overline{x}_i + a$, is equal to 1,000 and the intercept, a, equal to zero. In practice, the slope should not differ statistically from 1,000, nor the mean of the instrumental values, $\overline{x}$, from the mean of the reference values, $\overline{y}$ (mean of the differences, $d_i = \overline{x}_i - \overline{y}_i$, not different from zero).

The accuracy of an alternative method is given by the critical values or tolerance limits within which the true value is estimated to lie with a given probability when the alternative method is exactly calibrated.

Accuracy is expressed by the residual standard deviation, s_{yx}, of the differences between the true value, $\overline{y}_i$, and the estimated mean reference value, $\hat{y}_i$, obtained from the regression of the actual calibration function (see Figure A.1). It reflects, in a mathematical expression, the selectivity and the specificity of the method in predicting the true values of the component to be measured.

In practice, for any individual sample, the mean of test result values, $\overline{x}_i$, given by an alternative method exactly calibrated, should be considered suspect if the difference between $\overline{x}_i$ and the mean of the reference value $\overline{y}_i$ is outside the statistical tolerance limits.

5 Assessment of precision and accuracy

5.1 General

It should be borne in mind that both alternative and reference methods have to be standardized. At the least, very accurate detailed procedures of the methods shall be available before undertaking an overall accuracy evaluation.

5.2 Precision: Interlaboratory trial

5.2.1 General

For details concerning the organization and the statistical analysis, see ISO 5725-1 and ISO 5725-2.

In order to determine the precision (repeatability and reproducibility) of a test method, an interlaboratory trial or collaborative study is organized in which test portions of a certain number of test samples covering the normal range of variation of the measurand are analysed in replicate by several laboratories using the same standard method. Variation of results, both within and between laboratories, enables the precision of the method to be estimated.

5.2.2 Experimental design

5.2.2.1 Preparation and performance of such a trial should be conducted under precise conditions, briefly summarized in 5.2.2.2 to 5.2.2.6.

5.2.2.2 The number of participating laboratories should be as large as possible ($n_p \geqslant 8$). In practice, with high-cost instrumental methods, it may be difficult to find equipment in numerous laboratories.

5.2.2.3 The method should have been standardized or at least adequate instructions should have been written.

5.2.2.4 Assessing the reproducibility of an instrumental method implies that instruments are identically calibrated; otherwise reproducibility reflects the variations in calibration rather than the real reproducibility of the instrument. One major source of difference between calibrations is the lack of uniformity of test results obtained with the reference method.

To overcome that difficulty and to obtain a valid estimate of reproducibility, the laboratory responsible for the collaborative study should distribute standard materials (e.g. milk samples) to each of the participating laboratories for calibration of the different instruments.

5.2.2.5 Milk samples should be as homogeneous as possible and no physical or chemical change in the samples should occur before analysis.

5.2.2.6 Laboratories should perform at least duplicate tests under repeatability conditions.

5.2.3 Statistical analysis

The fullest possible ANOVA model is given in this subclause.

Such a model is suitable provided it fulfils the following requirements.

1) Within-laboratory (error) variances are homogeneous and there is no outlier result [see ISO 5725-2 for specific statistical tests].

2) The values of r and R are independent of the levels tested. If not, use a one-way ANOVA for each level or make an appropriate data transformation.

3) The laboratory bias may vary by level, leading to a significant interaction (laboratory × level) effect.

4) Laboratories involved are considered as a random selection of the total laboratory population. That variable is assumed to be normally distributed with a variance, σ_L^2. It is called the between-laboratory variance, and includes the random and systematic differences between laboratories. These systematic differences, which are unavoidable, should be small.

5) The component levels are also considered as a random selection of the total population of sample levels with a variance, σ_s^2. This variance does not enter into the calculation of precision data.

Under the aforementioned conditions a two-way ANOVA using a cross-classification and a random model allows the calculation of the variance of repeatability, σ_r^2, and the variance of reproducibility, σ_R^2 (see Table 1).

Table 1 — ANOVA of precision data

Source of variation	Sum of squares	Degrees of freedom	Mean square	Expected mean square
Sample (level)[a]	S_s	$\nu_s = q - 1$	$\overline{S}_s = S_s/(q-1)$	$s_e^2 + ns_{Ls}^2 + nps_s^2$
Laboratory	S_L	$\nu_L = p - 1$	$\overline{S}_L = S_L/(p-1)$	$s_e^2 + ns_{Ls}^2 + nqs_L^2$
Laboratory × sample	S_{Ls}	$\nu_{Ls} = (q-1)(p-1)$	$\overline{S}_{Ls} = S_{Ls}/(q-1)(p-1)$	$s_e^2 + ns_{Ls}^2$
Error	S_e	$\nu_e = pq(n-1)$	$\overline{S}_e = S_e/pq(n-1)$	s_e^2

[a] Samples representing q different levels are sent to p laboratories which perform n replicate tests at each level.

From Table 1 can be derived:

a) s_r^2, since the estimate, s_e^2, of the variance of error, σ_e^2, is an unbiased estimate of σ_r^2;

b) s_R^2, which is the sum of the variance between laboratories, s_L^2, the variance of interaction laboratory × sample, s_{Ls}^2, and the variance of repeatability, s_r^2.

That leads to:

$$s_R^2 = s_L^2 + s_{Ls}^2 + s_r^2 \tag{2}$$

where

s_L represents the standard deviation of laboratory systematic errors (mean biases);

s_{Ls} represents the part of the error related to levels (slope, deviation from linearity).

Information on the level standard deviation is given by s_s.

Finally, $r = 2{,}83 s_r$ and $R = 2{,}83 s_R$ are the quantitative expressions of the repeatability and the reproducibility limits of the method.

For a simple case for repeatability: when a set of q milk samples is analysed in duplicate by one laboratory, w_i being the absolute difference between duplicates, the standard deviation of repeatability can be calculated using:

$$s_r = \left(\frac{1}{2q} \sum_{i=1}^{q} w_i^2 \right)^{1/2} \tag{3}$$

5.2.4 Peculiarities of instrumental milk analysis

The assumption that precision is identical over the whole range of variation of the milk components is usually acceptable, but only within the normal range of concentration (e.g. for raw milk between 2,5 % mass fraction and 5,0 % mass fraction fat). When considering a wider variation, it is possible either to use a logarithmic transformation of data or to split the results into three or more levels and calculate the precision values for each of the split levels. For most methods of analysis, the repeatability value is proportional to its levels and can be expressed as a CV. Different figures are possible for the reproducibility value.

Routinely, milk samples are analysed once without rinsing between consecutive samples. Consequently, a certain amount of milk from previous samples can contaminate subsequent samples. The extent of contamination, called the "carry-over effect", is here conventionally included in the repeatability value of the instrument, provided that this carry-over effect is acceptable.

NOTE Usually the carry-over effect is less than 1 % mass fraction.

5.3 Accuracy

5.3.1 General

As pointed out in 4.3.1, accuracy is the major component of the trueness, characterizing the importance of the systematic errors specific to an alternative method.

5.3.2 Experimental design

Evaluation of method accuracy may be conducted separately by one or more laboratories having a good knowledge of the reference method.

Prior to such evaluation, the repeatability, the linearity of the signal over the normal range of concentration of the component, and all major instrumental parameters should be checked.

For purposes of the evaluation, it is not essential to adjust very closely the intercept and slope of the instrument when it is known that a small deviation in calibration does not affect the accuracy value.

A set of samples ($q > 40$) covering the normal range of variation of the component should be analysed at least in duplicate by using the reference method and the alternative method.

5.3.3 Statistical analysis

5.3.3.1 Analysis of variance

A total of q samples are analysed in replicate using the reference method, resulting in values y_i, and the alternative method, resulting in values x_i. The reference values obtained are considered here as the dependent variables and the alternative method values as independent.

According to the equation

$$s_{yx}^2 = s_y^2 (1 - r^2) \tag{4}$$

this method allows an estimate of the deviation from the regression, s_{yx}, not being dependent on the exactness of the instrument calibration (i.e. not dependent on the variance of the x values). It expresses the accuracy in the same unit as the reference method (e.g. in grams per 100 g of milk).

The ANOVA (Table 2) of the linear regression equation

$$y = bx + a \tag{5}$$

calculated from the means of replicates of each sample, allows the variance from the regression, σ_{yx}^2, to be estimated.

Table 2 — ANOVA of the regression of reference method versus alternative method results

Source of variation	Sum of squares	Degrees of freedom	Mean square	Expected mean square
Linear regression	S_{l}	$\nu_{\mathrm{l}} = 1$	$\overline{S}_{\mathrm{l}}$	
Deviation from regression	S_{yx}	$\nu_{yx} = q - 2$	$\overline{S}_{yx}/(q-2)$	s_{yx}^2

For each selected value $\overline{x}_i$ of the alternative method, there is a corresponding normal distribution of reference values, $\overline{y}_i$, with a mean equal to $b\overline{x}_i + a$, and a standard deviation, s_{yx}, which is an unbiased estimate of the standard deviation of accuracy.

5.3.3.2 Accuracy

According to the definition of accuracy, the value $\pm 1{,}96 s_{yx}$ is the quantitative expression for the accuracy of an alternative method.

However, the aforementioned statement is not fully correct since the regression coefficient is an estimated value with a variance, σ_{b}^2. Therefore, in order to simplify the expression for accuracy, σ_{b}^2, is assumed to be negligible, and the 95 % confidence interval of individual estimated values of y is assumed to be constant at each level.

In the simple case, when the calibration is exact and the instrument is perfectly calibrated ($b = 1{,}000$ and $a = 0$), the standard deviation of the algebraic differences between instrumental and reference results, s_{di}, is an estimate of the standard deviation of accuracy.

At a difference of one degree of freedom, s_{di} is equal to the standard deviation from the regression, s_{yx}:

$$s_{yx} \approx s_d \left[\sum \frac{\left(d_i - \overline{d}\right)^2}{q-1} \right]^{1/2} \tag{6}$$

5.3.3.3 Correlation coefficient

Using the correlation coefficient between the reference and the alternative method results as a quantitative expression of the accuracy of an alternative method can be misleading. That is because the correlation coefficient is closely related to the variance of the dependent variable, y (the reference method).

Therefore, for a given s_{yx}^2 value, the larger the variance of y, the better the correlation coefficient:

$$r_{xy}^2 \approx 1 - \frac{s_{yx}^2}{s_y^2} \tag{7}$$

It is not recommended to use the correlation coefficient as a single statistical indicator for accuracy. The standard deviation from regression, s_{yx}, is more meaningful and a better estimate of accuracy than the correlation coefficient, especially when that is 0,97 to 0,98 or higher. However, this latter parameter can be used, along with the standard deviation from regression, provided that the range of concentration for the measured component with which it is associated is clearly stated.

Otherwise, the correlation coefficient is relevant and may be used in quality assurance procedures as an additional indicator to verify whether the calibration sample set characteristics, s_{yx} and s_y, provide sufficient conditions for an appropriate estimation of the calibration line.

5.3.4 Accuracy of instrumental milk analysis

It has been clearly demonstrated that the accuracy of instrumental methods currently in use for milk analysis is influenced to a rather large extent by variations in the composition of the milk samples and, therefore, by the degree of homogeneity of the population of test samples (e.g. test samples from individual animals or pooled milk).

That means that, according to the experimental conditions, more than one accuracy value may be found. Therefore, it is of the utmost importance for purposes of evaluation to carry out experiments under the different conditions corresponding to the various applications of the methods.

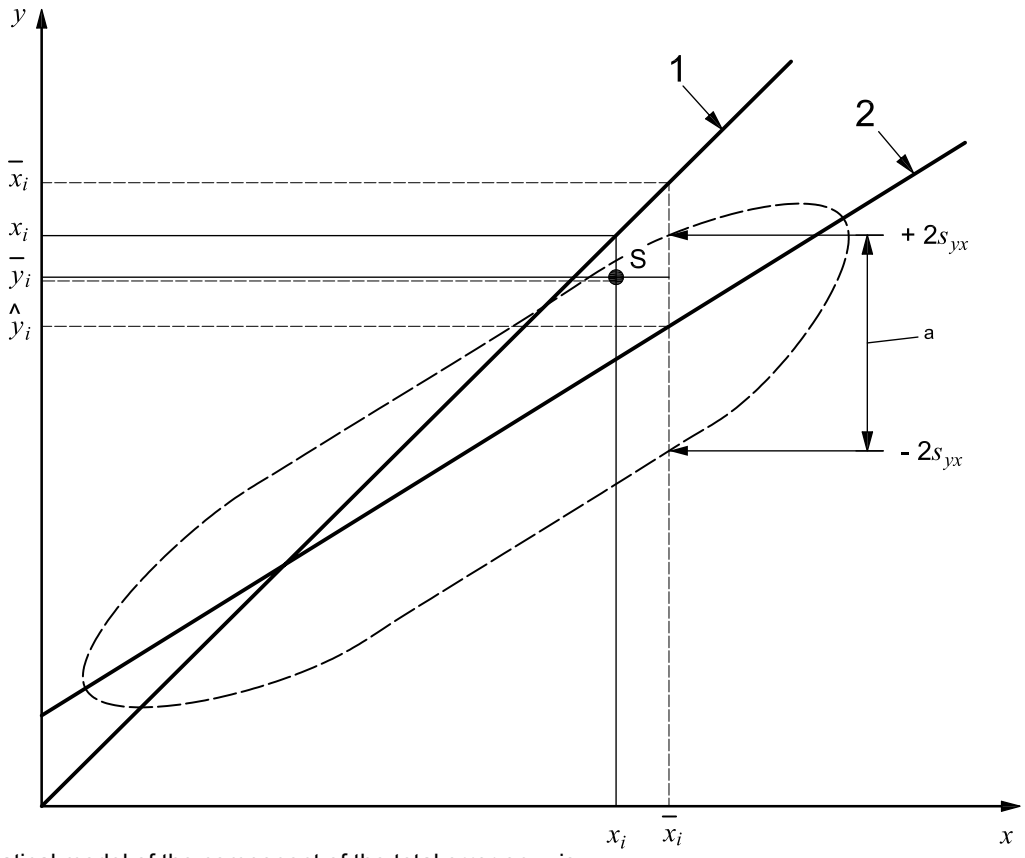
5.3.5 Breakdown of overall accuracy

Figure A.1 illustrates the various criteria involved in the overall accuracy of an alternative method.

The ellipse encloses the scatter diagram for the results of a given population of samples. It corresponds to the 95 % confidence interval within which the true value is expected to lie for any result of the alternative method.

Annex A
(normative)

Illustration of the various criteria involved in the overall accuracy of an alternative method



The mathematical model of the component of the total error on x_i is:

$$(x_i - \bar{y}_i) = (x_i - \bar{x}_i) + \begin{cases} (\bar{x}_i - \bar{y}_i) \\ (\bar{x}_i - \hat{y}_i) + (\hat{y}_i - \bar{y}_i) \end{cases}$$

Key

1	theoretical line, $y = x$	$\bar{x}_i$	arithmetic mean of several determinations of sample S with the instrument	$(x_i - \bar{y}_i)$	overall accuracy
2	observed line, $y = bx + a$	y	reference method	$(x_i - \bar{x}_i)$	repeatabillity
S	"true value" of the component to be measured for sample S	$\hat{y}_i$	estimated mean reference value for all the samples with an instrument level, $\bar{x}_i$	$(\bar{x}_i - \bar{y}_i)$	trueness
x	indirect method	$\bar{y}_i$	"true value" of the component to be measured for sample S	$(\bar{x}_i - \hat{y}_i)$	exactness of calibration
x_i	single instrumental value	s_{yx}	standard deviation from the regression	$(\hat{y}_i - \bar{y}_i)$	accuracy

a Accuracy.

Figure A.1 — Breakdown of criteria in the overall accuracy of an alternative method

Bibliography

[1] ISO 21187|IDF 196:2004, *Milk — Quantitative determination of bacteriological quality — Guidance for establishing and verifying a conversion relationship between routine method results and anchor method results*

[2] ISO 24276:2006, *Foodstuffs — Methods of analysis for the detection of genetically modified organisms and derived products — General requirements and definitions*

[3] ISO/IEC Guide 98-3:2008, *Uncertainty of measurement — Part 3: Guide to the expression of uncertainty in measurement* (*GUM:1995*)

[4] ISO/IEC Guide 99:2007, *International vocabulary of metrology — Basic and general concepts and associated terms (VIM)*

ICS 67.100.10

Price based on 13 pages