INTERNATIONAL STANDARD

ISO 8196-3
IDF 128-3

Second edition
2022-05

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

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

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

Partie 3: Protocole d'évaluation et de validation des méthodes quantitatives alternatives pour l'analyse du lait

Reference numbers
ISO 8196-3:2022(E)
IDF 128-3:2022(E)

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11

Email: copyright@iso.org
Website: www.iso.org

International Dairy Federation
Silver Building • Bd Auguste Reyers 70/B
B-1030 Brussels
Phone: +32 2 325 67 40
Fax: +32 2 325 67 41
Email: info@fil-idf.org
Website: www.fil-idf.org

Published in Switzerland

Contents

Page

Forewords

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.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular, the different approval criteria needed for the different types of ISO documents should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

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. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation of the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT), see www.iso.org/iso/foreword.html.

This document 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 cancels and replaces the first edition (ISO 8196 | IDF 128-3:2009), which has been technically revised. The main changes are as follows:

— the validation scheme has been simplified for phase II and it is possible to validate a new instrument with the comparison with a previous validated instrument.

A list of all parts in the ISO 8196 | IDF 128 series can be found on the ISO website.

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html.

IDF (the International Dairy Federation) is a non-profit private sector organization representing the interests of various stakeholders in dairying at the global level. IDF members are organized in National Committees, which are national associations composed of representatives of dairy-related national interest groups including dairy farmers, dairy processing industry, dairy suppliers, academics and governments/food control authorities.

ISO and IDF collaborate closely on all matters of standardization relating to methods of analysis and sampling for milk and milk products. Since 2001, ISO and IDF jointly publish their International Standards using the logos and reference numbers of both organizations.

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. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

This document was prepared by the IDF *Standing Committee on Statistics and Automation* and ISO Technical Committee ISO/TC 34, *Food products*, Subcommittee SC 5, *Milk and milk products.* It is being published jointly by ISO and IDF.

The work was carried out by the IDF/ISO Action Team (S14) of the *Standing Committee on Statistics and Automation* under the aegis of its project leader, Dr S. Orlandini (IT).

Introduction

This document is complementary to ISO 8196-1 | IDF 128-1. It describes a protocol for the evaluation of new alternative methods for which ISO 8196-1 | IDF 128-1 cannot apply, e.g. when the organization of interlaboratory studies is hampered by a limited number of new instruments available for study.

The latter is generally the case with dedicated instrumental methods (e.g. milk payment analysis, milk recording analysis) of which the commercialization depends on official approvals for use. An application for such an official approval is to be accompanied by one or more assessments of the relevant performance characteristics.

This document specifies a harmonized protocol for such a method validation by expert laboratories. It lists the evaluation steps and provides a criteria-based approach for the assessment of the performance characteristics, including guidance for checking statistical compliance.

On the basis of such a harmonized protocol, a limited number of evaluations should suffice for a decision by an approval body for the application of the method and/or equipment. Examples with indicative limits are given for the evaluation of a method for the determination of fat, protein, lactose, urea and somatic cell count in milk. The guideline can also be applied to other parameters such as freezing point and pH in milk.

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

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

1 Scope

This document specifies a protocol for the evaluation and validation of alternative quantitative methods of milk analysis. This document is also applicable for the validation of new alternative methods where, due to a limited number of operational instruments, the execution of an interlaboratory study and ISO 8196-1 | IDF 128-1 is not feasible.

The protocol is applicable to milk parameters such as, for example, fat, protein, lactose, urea and somatic cells in milk. It can also be extended to other parameters.

This document also establishes the general principles of a procedure for granting international approvals for the performance of the alternative methods. These principles are based on the validation protocol defined in this document.

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO 3534-1, *Statistics — Vocabulary and symbols — Part 1: General statistical terms and terms used in probability*

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

ISO 8196-1 | IDF 128-1, *Milk — Definition and evaluation of the overall accuracy of alternative methods of milk analysis — Part 1: Analytical attributes of alternative methods*

ISO 8196-2 | IDF 128-2, *Milk — Definition and evaluation of the overall accuracy of alternative methods of milk analysis — Part 2: Calibration and quality control in the dairy laboratory*

ISO/IEC 17025, *General requirements for the competence of testing and calibration laboratories*

3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO 3534-1, ISO 5725-1, ISO 8196-1 | IDF 128-1, ISO 8196-2 | IDF 128-2 and the following apply.

ISO and IEC maintain terminological databases for use in standardization at the following addresses:

— ISO Online browsing platform: available at https://www.iso.org/obp

— IEC Electropedia: available at https://www.electropedia.org/

3.1
validation of alternative method
verification of the performance of an alternative method on whether it is adequate for the intended use

3.2
measurand
quantity intended to be measured

Note 1 to entry: A measurand may be a milk component (e.g. fat and protein), a physical characteristic (e.g. freezing point) or a biological element (e.g. somatic cells).

Note 2 to entry: Adapted from ISO/IEC Guide 99:2007, 2.3.

3.3
quantitative method
method of analysis whereby the result is an amount of a quantity, a concentration or a value of a *measurand* (3.2) determined either directly or on a test portion

3.4
methods comparison study
study performed by an *expert laboratory* (3.6) of an alternative method against the reference method or a comparison method/instrument under test bed conditions

3.5
interlaboratory study
study of the performance of an alternative method on one or more "identical" laboratory samples of homogeneous, stable materials under documented conditions in several laboratories and under the control of an *organizing laboratory* (3.7)

Note 1 to entry: The data interpretation should be performed in collaboration with *expert laboratory* (3.6).

3.6
expert laboratory
laboratory having qualified staff and equipment to perform a *methods comparison study* (3.4)

Note 1 to entry: The expert laboratory is specialized in analytical evaluations and shall conform to ISO/IEC 17025 as well as having relevant experience in the area of application.

3.7
organizing laboratory
laboratory having staff with statistical expertise and qualified staff and necessary equipment to prepare the samples to perform an *interlaboratory study* (3.5)

Note 1 to entry: The organizing laboratory shall operate in conformity with ISO/IEC 17025 for the method used to check the homogeneity of the samples.

3.8
national approval
authorization of the use of a method for defined purposes in a country, generally for reasons of collective interest and/or having an official character, delivered by an approval body

3.9
international approval
authorization of the use of a method for defined purposes at international level, generally for reasons of collective interest and/or having an official character, delivered by an approval body for the benefit of stakeholders

4 General principles for the validation of alternative methods

4.1 Validation protocol

4.1.1 General

The validation protocol comprises two phases as specified in 4.1.2 and 4.1.3.

4.1.2 Phase I

A methods comparison study includes the assessment of the performance characteristics of the alternative method under validation. A comparison of the alternative method against the reference method under test bed conditions is required. In cases where the instrument under evaluation has the same analytical principle and only minor technical changes from the previously validated version, the comparison can be done between the two instruments, considering the results of the oldest version as an anchor to evaluate the results of the new instrument generation.

This part of the evaluation shall be carried out by an expert laboratory.

4.1.3 Phase II

A method confirmation study under routine testing conditions is initiated after a successful Phase I. It is recommended to examine at least two instruments, for national approval, or three instruments, for international approval.

Depending on the purpose, the approval body can decide whether two or three instruments are to be examined and whether the instruments are to be located in the same laboratory or in different laboratories and geographies under routine testing conditions. A test period of a minimum of two months is recommended for Phase II or to organize an interlaboratory study associated with the data collection from routine laboratories. For this phase, detailed steps are described in 5.3.2.

4.1.4 National approval

Based on the content of submitted reports, a competent body can grant a national approval, indicating sufficient quality in measurement results and adequateness of the alternative method for the proposed purpose.

4.1.5 International approval

Approval bodies or international organizations can grant an international approval. International approval can be granted based on three single national validations or the results of Phase I performed in an expert laboratory and the results from a method confirmation study or an interlaboratory study as described in 4.1.3.

4.2 Field of validity of the approval

This protocol is applicable to the validation of alternative methods for the quantitative compositional analysis and somatic cell count determination in raw milk from cow, sheep, goat and buffalo. The validation study shall be conducted separately for the milk of each species. When a component under validation occurs with unusual concentrations (e.g. Jersey breed with high fat and protein content) the evaluation should be carried out over the whole relevant range of the concerned component.

The method and/or instrument should be evaluated with the configuration as offered by the concerned manufacturer. If the configuration changes, it should be proven in an independent way that it does not influence the precision and the accuracy beyond acceptable limits.

Carefully note and report all characteristics of both the milk products analysed, the calibration model(s) version and the configuration(s) of the alternative method assessed.

5 Technical protocol for the validation

5.1 Course of operations

Whatever the alternative method, a standard measurement process can be represented schematically as shown in Figure A.1. Each step corresponds to a source of error that can contribute to the overall uncertainty of the method. The evaluation protocol and experimental designs are constructed to fit the sequence of signal treatment and to permit verification that they are set up in such a way that precision and accuracy of the method can respond to the limits required in practice.

It is necessary for each step of the evaluation described in the following paragraphs to fulfil the appropriate limits for each analytical criterion before starting the next step.

The methods comparison study (Phase I) defines the minimum assessment sequence to be carried out.

The method confirmation/interlaboratory study (Phase II) provides complementary information on the method performance under routine use conditions.

5.2 Methods comparison study (Phase I)

5.2.1 General

The evaluation is to be carried out with test results expressed in standardized units of the reference method. For methods covering large ranges of measured values (i.e. wider than one log unit), it is recommended to split the range into levels, each of maximum width one log unit, so as to obtain a minimum of three levels and to perform statistical calculations separately on each level. Where appropriate, a logarithmic transformation of the data can be applied, see 5.2.2.

NOTE 1 For instance, for fat in commercial milk, distinction can be made between skim milk, semi-skimmed milk and whole milk. For raw milk, natural fat and protein ranges are often related to the species, which are then to be assessed by separate evaluations (see 4.2). Somatic cells in raw milk typically cover a range of several log units.

Evaluation results should conform to the specifications stated in the following paragraphs. For general dairy industry purposes, limits for the different analytical characteristics mentioned have been extracted or derived from existing International Standards.

Annex B summarizes these limits for fat, protein (crude protein, true protein and casein), lactose, urea, somatic cells, freezing point and pH as indicative limits obtained from proficiency tests.

NOTE 2 For liquid milk during milking or processing, there can be different assessment criteria for in-line and at-line analyses systems.

5.2.2 Compulsory assessments for the validation

5.2.2.1 Assessment of preliminary instrumental fittings

5.2.2.1.1 General

Before starting any further assessment, basic criteria indicating a proper functioning of the method or the instrument require verification. These criteria are repeatability, intralaboratory reproducibility carry-over and linearity.

5.2.2.1.2 Precision (repeatability and intralaboratory reproducibility)

The method used should present a stable measurement signal that conforms to the precision requirements. If not, the analyser is either not functioning correctly (and should not be used) or its precision is not appropriate for the objective of the analysis. Hence, the instantaneous stability (repeatability) and the signal level stability shall be assessed prior to any other characteristics.

The precision should be evaluated at three different concentration levels of the component measured: low, medium, and high.

During the day, analyse pilot milk samples in triplicate (n = 3) every 15 min to 20 min of instrument activity without any change in the calibration in order to obtain results from a minimum of 20 pilot samples analysed for each level ($q \geq 20$). Preferably, the instrument should be operated under conditions as close as possible to routine circumstances. Sufficient numbers of samples should be processed to keep the instrument running between the periodic checks.

Estimate for each pilot:

a) s_r, the standard deviation of repeatability;

b) s_p, the standard deviation of mean pilots;

c) s_c, the standard deviation between time periods;

d) s_{Rintra}, the standard deviation of intralaboratory reproducibility.

For each time period (i = 1,2, ... q), calculate the pilot sample mean $\overline{x_j}$ and the standard deviation of the mean pilot s_j over the q replicate measurements, as shown by Formulae (1) and (2):

$$\bar{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij} \tag{1}$$

$$s_j = \left(\frac{1}{n-1}\sum_{i=1}^{n}(x_{ij} - \bar{x}_j)^2\right)^{1/2} \tag{2}$$

where

n is the number of replicates at each time period (typically n = 3).

The overall repeatability standard deviation of this pilot is found by averaging these s_j^2 over all the q time periods in the day, as shown by Formula (3):

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

where

q is the number of time periods.

and the standard deviation of mean pilots, as shown by Formula (4):

$$s_p = \sqrt{\left[\frac{1}{q-1}\sum_{j=1}^{q}\left(\bar{x}_j - \bar{\bar{x}}^2\right)\right]} \tag{4}$$

where $\bar{\bar{x}} = \frac{1}{q}\sum_{i=1}^{q}\overline{x_i}$

The corrected standard deviation between time periods (for this pilot) is given by Formula (5):

$$s_c = (s_b^2 - s_r^2/n)^{1/2} \tag{5}$$

with $s_c = 0$ if $s_c < 0$.

The overall standard deviation of intralaboratory reproducibility for this pilot is shown by Formula (6):

$$s_{R\text{intra}} = \sqrt{s_r^2 + s_c^2} \tag{6}$$

The values obtained for s_R and $s_{R\text{intra}}$ should conform to the limits stated in Annex B.

The stability of the method response during the analyses of the pilot sample can be visualized by plotting the means $\bar{x}_j$ of the different three pilots means versus the time. See the example in Clause C.1.

5.2.2.1.3 Carry-over effect

5.2.2.1.3.1 Strong differences in component concentrations between two successively analysed samples can influence the result of the second.

Differences can be caused by incomplete rinsing of the flow system and the measuring cell by liquid circulation and contamination by the stirring device. Automatic correction of results is acceptable within certain limits, provided it can be proven that there is a systematic transfer of a small quantity of material from one measurement to the next.

Automated analysers for liquids often allow automatic correction to compensate for the overall carry-over effect when necessary. Carry-over shall be clearly distinguished from rinsing efficiency.

5.2.2.1.3.2 The overall carry-over effect should be assessed including the correction factors either set in the instrument or obtained using the method supplied by the manufacturer. It should not exceed the values stated per component.

Limits are defined from the prerequisite that carry-over effect should not produce an error higher than the repeatability of the method. Hence, limits for the carry-over ratio (COR), L_C, should fulfil the condition $L_C \leq (r/\Delta L_{range}) \times 100$ where r is the repeatability limit at the level of the bias measured and ΔL_{range} is the difference between the maximum and the minimum concentration in the range of interest. For components where repeatability is not constant over the measuring range, the COR limits are set based on the levels of best repeatability (e.g. somatic cell counting). Common limits for COR are in the range 1 % to 2 %.

5.2.2.1.3.3 The rinsing efficiency of the flow system shall be assessed separately by running tests without any correction (correction factor set to zero) in manual mode that bypasses the stirrer. The carry-over should not exceed 1 % as given in ISO 9622 | IDF 141 or 2 % as given in ISO 13366-2 | IDF 148-2.

5.2.2.1.3.4 Analyse two samples, with high and low concentrations of prior distribution in series of test portions. Repeat, as many times, as necessary (see below) the analytical sequence in terms of component concentration, low, low, high, high, in order to obtain N_C sets of results, L_{L1}, L_{L2}, L_{H1} and L_{H2}. The minimum number of sequences, N_C, should be 20.

NOTE For components where repeatability is not constant over the measuring range and for levels with high repeatability, more numerous sequences can be required. Alternative numbers of sequences can be calculated by $N_C \geq [r \times 100/(L_C \Delta L_{test})]^2$ where ΔL_{test} is the range between high and low concentration samples (equal to or greater than ΔL_{range}).

5.2.2.1.3.5 Method requirements for samples: Prepare a sufficient number of test portions from each low and high concentration laboratory sample prior to analysis in order to analyse each test portion only once. The low and high concentration laboratory samples should preferably be milks or liquid products with similar viscosity to those routinely analysed.

Ensure that individual component concentrations differ considerably. For milk, this can, for instance, be achieved by using natural separation (creaming for fat), artificial separation (ultrafiltration for protein, microfiltration for somatic cells) or addition (lactose and urea).

For biochemical component determinations, the low and high concentrations of the laboratory samples should, preferably, be extreme values in the measuring range.

Sufficiently large ranges are recommended to easily differentiate carry-over effects from random error. The minimum range needed, $\Delta L_{test} = L_H - L_L$, can be calculated according to $\Delta L_{test} \geq r \times 100/(L_C\sqrt{N_C})$ where r and L_C are the stated limits and N_C is the number of sequences applied (see Annex B).

For milk components or criteria covering large ranges of concentration, e.g. from 10 to 1 000, the ratio of carry-over error may not be constant over the whole range. This should be verified by assessing the carry-over at different concentrations.

In such case, it is recommended to choose a level L_{Hi} at the median of each part, i, previously defined in the whole range. A minimum number of two levels in the medium and high concentration range is needed that can be extended to three for particularly wide ranges.

Indication for somatic cell counting in individual animal milk, the definition of three levels, at about 500 x 10^3 cells/ml, 1 000 x 10^3 cells/ml and 1 500 x 10^3 cells/ml, is recommended.

5.2.2.1.3.6 Calculation: Calculate the mean of the differences, $d_{LLi} = L_{L1i} - L_{L2i}$ and $d_{LHi} = L_{H2i} - L_{H1i}$, $d_{LL,}$ d_{LH} and the mean difference of concentration, $d_{\acute{A}} = \bar{L}_{H2} - \bar{L}_{L2}$

The COR can be obtained by using Formulae (7) and (8):

$$C_{H/L} = \left(\sum L_{L1} - \sum L_{L2}\right) \times 100 / \left(\sum L_{H2} - \sum L_{L2}\right) = \left(\bar{L}_{L1} - \bar{L}_{L2}\right) \times 100 / \left(\bar{L}_{H2} - \bar{L}_{L2}\right) \quad (7)$$

$$C_{L/H} = \left(\sum L_{H2} - \sum L_{H1}\right) \times 100 / \left(\sum L_{H2} - \sum L_{L2}\right) = \left(\bar{L}_{H2} - \bar{L}_{H1}\right) \times 100 / \left(\bar{L}_{H2} - \bar{L}_{L2}\right) \quad (8)$$

The two should not exceed the limit, L_C, in the test condition stated for the component reported in Annex B.

5.2.2.1.4 Linearity

5.2.2.1.4.1 General

According to the classical definition of an indirect method, the instrument signal should result from a characteristic of the component measured and thereby allow the definition of a simple relationship to the component concentration.

Linearity expresses the constancy of the ratio between the increase in the concentration of a milk component and the corresponding increase of the alternative method result. Therefore, linearity of the measurement signal is in most cases essential to maintain a constant sensitivity over the measuring range and to allow easy handling of calibration and fittings. Moreover, it allows in routine (to some extent) measurements beyond the calibration range through linear extrapolation.

The method is specified in 5.2.2.1.4.2 to 5.2.2.1.4.4.

5.2.2.1.4.2 Samples

Linearity can be assessed using sets of 8 to 15 samples with component concentrations evenly distributed over the measuring range.

a) Samples should preferably be milks or liquids of similar physical characteristics (i.e. density, viscosity), e.g. by combining (weighing) a high content sample, L_H, and a low content sample, L_L.

b) Concentrations should vary in regular intervals. Depending on the component, that can for instance be achieved by natural separation (creaming for milk fat), artificial separation (ultrafiltration for protein, microfiltration for somatic cells) and recombination, or by using pure solutions (lactose

and urea). For the PLS calibration model, it is suggested to add pure substance (powder, e.g. lactose) to the milk.

c) The linearity assessment range should be congruent with the concentration range for the validation study (see Annex B).

d) Reference values for linearity samples can be established from either the mixing ratio or the theoretical concentrations as calculated from the concentrations of the initial samples. Depending on the alternative method, they should be obtained from volume by volume mixing ratios where analysis is performed on a milk volume (volumetric intake measurement) and from mass by mass mixing ratios where analysis is applied to a weighed milk portion (see Annex D).

5.2.2.1.4.3 Analyses

If N_L is the total number of replicates to execute for each sample, analyse first, in order of increasing concentrations, half replicates ($N_L/2$) and second, in order of decreasing concentrations, half replicates in ($N_L/2$), so as to obtain the total replicate number N_L relevant for the measurand (see Tables B.1, B.2 and B.3).

5.2.2.1.4.4 Calculation and assessment

Calculate the linear regression formula $y = bx + a$ (where y = instrument and x = reference) and the residuals e_i ($e_i = y_i - bx_i - a$) from the means of replicates and the theoretical reference.

Plot the residuals, e_i, on the ordinate against theoretical concentrations on the abscissa. Visual inspection of the data points usually yields sufficient information about the linearity of the signal.

Any deviation from linearity or obvious trend in the data in this plot indicates a potential problem and should lead to further investigation of the method, as detailed below.

Any residual obviously being out of the current distribution (outlier) should lead to deletion of that result and repetition of the calculation before applying further tests.

Calculate the ratio of the residual range to the signal values range, shown by Formula (9):

$$\frac{\Delta_e}{\Delta_L} = \frac{(e_{max} - e_{min})}{(L_{max} - L_{min})} \quad (9)$$

where

e_{max} is the numerical value of the upper residual;

e_{min} is the numerical value of the lower residual;

L_{max} is the numerical upper mean value measured with the instrument;

L_{min} is the numerical lower mean value measured with the instrument.

NOTE 1 Limits are defined from the prerequisite that deviation from linearity will not produce a larger error than the repeatability of the method over the usual measuring range. Hence, limits of the relative linearity bias, $L_{\Delta e/\Delta L}$, are meant to fulfil the condition $L_{\Delta e/\Delta L} \leq r/\Delta L_{range}$ for the upper acceptable repeatability, with r being the repeatability limit and ΔL_{range} being the difference between the maximum and the minimum concentration in the concentration range of interest. For components where repeatability is not constant over the measuring range, the relative linearity bias limits are set based on the levels of largest repeatability (e.g. somatic cell counting). Common limits for $\Delta e/\Delta L_{range}$ are in the range 0,01 to 0,02.

NOTE 2 The number of replicates, N_L, needed to ensure significance of the Δe/ΔL test can be estimated by the conditions:

$N_L \geq 8\sigma_r^2/(L^2_{\Delta e/\Delta L}\,\Delta L^2_{test})$ or $N_L \geq r^2/(L^2_{\Delta e/\Delta L}\,\Delta L^2_{test})$

NOTE 3 Concentration ranges, ΔL_{test}, larger than ΔL_{range} allow the measurement of larger linearity bias, Δ_e, with a similar relative linearity bias and increased significance for the same maximum repeatability value. The minimum concentration range can be estimated by the conditions: $\Delta L_{test} \geq 2\sqrt{2}\sigma_r/(L_{\Delta e/\Delta L}\sqrt{N_L})$ or $\Delta L_{test} \geq r/(L_{\Delta e/\Delta L}\sqrt{N_L})$.

A one-way ANOVA can be carried out to confirm the statistical significance of nonlinearity. Statistical tests for comparison of variances can be applied to confirm the significance of difference between residual variances.

Furthermore, if needed, nonlinear trends can be evaluated by second- and third-degree polynomial and statistical tests.

5.2.2.1.5 Measurement limits

Limits of a measurement with an instrumental method exist at both extremities of the analytical range, e.g. a lower limit and an upper limit.

It is not required to determine these limits when natural concentration ranges for the respective components and species are normally located far from zero (which is generally the case for biochemical components, i.e. fat, protein, lactose, urea), and within the linearity range of the method.

The assessment of the measurement limits can be carried out in combination with the evaluation of the linearity. If linearity is not achieved throughout the whole concentration range, determine the actual range of application for the method concerned.

5.2.2.1.6 Lower limits

5.2.2.1.6.1 General

Where, routinely, only single determinations are carried out, σ is the standard deviation of random error of the measurement. In the best case, that is the repeatability standard deviation at the proximity of zero content. Standard deviation of repeatability for the blank or standard deviation of repeatability estimated at concentrations close to zero are to be used. At least 20 samples are required as independent replicates.

5.2.2.1.6.2 Detection limit

The detection limit defines the lowest results, which differ significantly from zero. To determine the method detection limit (LOD) it is necessary to analyse a milk sample with a low content of the parameter. Run 10 replicates, calculate the mean and standard deviation. Calculate the LOD as three times the standard deviation.

5.2.2.1.6.3 Quantification limit

The quantification limit, which is the smallest amount of measurand that can be measured and quantified with a defined coefficient of variation. The quantification limit is calculated as 10 times the standard deviation.

5.2.2.1.6.4 Upper limit

Upper limit corresponds to the threshold where the signal or the measurement deviates from linearity. (see 5.2.2.1.4 and Annex B).

5.2.2.2 Evaluation of the overall accuracy

5.2.2.2.1 General

The overall accuracy depends on the repeatability, the accuracy and the applied calibration model.

With raw milk, each part of the overall accuracy is measured through the analysis of individual milk samples and herd bulk milk samples of the animal species specified. Herd bulk milk samples shall be collected in addition to individual milk samples in order to measure more accurately that part of the variance related to herd effects.

The evaluation is to be performed under conditions equivalent to the intended operation in routine (working parameters, speed and calibration).

5.2.2.2.2 Samples

Good quality milk samples should be used. Individual milk samples should cover the maximum concentration range of the component, according to the specifications of Annex B.

The minimum number required for the measurand is related to the statistical significance in comparisons and the matrix variability of the product.

Generally, a minimum number of 100 individual animal milk samples (N_a = 100) from different herds (N_h > 5) and 60 herd milk samples ($N_{sample/h}$ = 60) are required with regard to the need for sample representativeness (see Annex B). These samples should be measured in duplicate with reference and alternative methods.

The time between the two methods should not exceed 2 h.

The data obtained from these duplicates will be used to calculate and assess the repeatability.

5.2.2.2.3 Assessment of repeatability

Repeatability is the main criterion indicating whether a method produces stable results according to user requirements. It is a major element of internal quality control. Therefore, every new instrument has to fulfil a maximum limit for repeatability value, stated in the relevant International Standard, in order to satisfy the criteria for approval.

The standard deviation of repeatability is calculated from duplicate results obtained from the whole set of data and, for criteria covering a wide range of concentrations, that is more than one log scale (part by part after splitting of the whole concentration range into different parts (minimum three parts of maximum one log unit width each, i.e. low, medium and high).

For q samples analysed in duplicate, the standard deviation of repeatability is calculated from Formula (10) [see also Formula (C.5)]:

$$s_r = \left(\frac{1}{2q} \sum_{i=1}^{q} w_i^2 \right)^{1/2} \tag{10}$$

where w_i is the modulus of the difference between duplicates of sample i ($w_i = |x_{1i} - x_{2i}|$).

The value of s_r obtained should be compared with the limit value for the repeatability values, σ_r, as defined for the relevant measurand and application (see Annex B). It is expected that $s_r \leq \sigma_r$.

For the calculation of the repeatability, it is possible to duplicate data obtained during the accuracy test as reported in 5.2.2.2.2 and the data obtained during Phase II where the routine samples are analysed in duplicate.

5.2.2.2.4 Accuracy

5.2.2.2.4.1 General

According to ISO 8196-1 | IDF 128-1, the overall error of trueness comprises the combination of the error of the calibration model and the error of accuracy.

The aforementioned parameters are obtained from a simple linear regression, calculated in accordance with ISO 8196-1 | IDF 128-1 and ISO 8196-2 | IDF 128-2 using means of duplicate instrumental results, *y*, and so-called reference results, *x*, obtained by the reference method or a comparison method in duplicate.

For measurands covering a wide concentration range, that is more than one log scale (i.e. with somatic cell count), an accuracy evaluation should be performed for the whole range and for successive parts of the range after splitting the whole concentration range into different parts (at minimum three parts of maximum one log unit width each, i.e. low, medium and high).

NOTE In cases where the precision error in the result of the alternative method is significantly lower than the precision error with the reference method, *x* and *y* can be swapped.

5.2.2.2.4.2 Assessment of accuracy

For raw milk, accuracy is assessed for individual animal milks and herd bulk milks separately. It is based on the residual standard deviation, s_{yx}, of the simple linear regression of instrumental results, *x*, and reference results, *y*. It is assumed that the differences from the regression line are normally distributed and the variance is homoschedastic. If not, the logarithmic data transformation is recommended. Any outlying results (e.g. Grubbs test) should be carefully scrutinized. For outlying results, further test samples drawn from the same calibration sample should preferably be reanalysed by both the reference method and the alternative method.

When outlying figures remain, the report should present s_{yx} estimates and graphs including all data (with the outliers identified, their number and respective biases) as well as s_{yx} estimates after discarding outliers. Statistical methods used to identify outliers should be specified in the evaluation report. The proportion of outliers should not exceed 5 %.

The estimated value of s_{yx} should fulfil the limits σ_{yx} as defined for the parameter and the matrix concerned. It should respect the condition $s_{yx} \leq \sigma_{yx}$. Limits for individual animal and herd bulk milk samples are given in Annex B.

In the case of assessing the accuracy of an instrument by comparing it with another already validated instrument, the same number and type of samples should be analysed with both instruments. In addition, it is required to analyse with the reference method a limited number of these samples (e.g. 20 individual milk samples and 10 herd milk samples) that cover the expected range.

When comparing with an already validated instrument, results with this instrument are to be put on the x-axis and results with the instrument under validation on the y-axis.

5.2.2.2.4.3 Assessment of a revised calibration model

In case of a new version of the calibration model being provided by the manufacturer, analyse 20 milk samples that cover the working range, or use available reference materials, before and after installing the new calibration model.

Apply a *t*-test to the two sets of results obtained to evaluate if there is a significant difference.

5.2.3 Additional informative investigations

5.2.3.1 General

The following items are not compulsory elements for an evaluation even though they are of interest as possible contributors to the overall accuracy of the method. Moreover, knowledge gained about the method can have implications in milk sample handling (sampling, preservation, shipment, etc).

5.2.3.2 Ruggedness

5.2.3.2.1 General

Ruggedness is the ability of a method not to be influenced by external elements other than the component measured itself. Possible effects can come from concentration variation of major milk components or interactions, biochemical changes of milk components related to preservation (lipolysis, proteolysis, lactic souring) or chemicals added to the milk such as preservatives.

The principle of a ruggedness assessment is to produce a significant change in the concentration of each interacting component separately and measure the corresponding change in the measurement result of the influenced component.

Then the ratio of the difference observed and the change introduced is calculated and expressed in the relevant units.

5.2.3.2.2 Effect of major milk components (interference)

5.2.3.2.2.1 General

To determine if there are any interactions of different milk components (fat, protein, lactose), appropriate guidance is provided in ISO 9622 | IDF 141 for sample preparation and calculation procedures. Although first developed for mid-infrared methods, the approach is also applicable with other methods.

The effect of lactose or urea on other component measurements can be evaluated by the addition of lactose or urea to milk.

The effect of high fat and protein content on the somatic cell count in milk (sheep, goat and buffalo milk) can be evaluated by recombining cream (natural creaming) and milk retentate in a similar way to that specified in ISO 9622 | IDF 141.

The effect should preferably be measured at three relevant levels within the range of the measurand, i.e. low, medium and high, for the animal species.

5.2.3.2.2.2 Effect of biochemical changes in components

Biological changes in milk usually result in breakdown of milk components that can be induced by bacterial growth or enzymatic activity. Deterioration of milk samples may go unnoticed. Therefore, it is relevant to check the susceptibility of an alternative method for such deterioration, in particular in order to evaluate the quality of the sample preservation and the suitability of sampling and shipment conditions.

Clotting, churning and oiling are generally clearly visible defects of raw milk that can affect analytical results. In those cases, samples should be discarded.

5.2.3.2.2.3 Lipolysis

The possible effect of lipolysis can be monitored through artificial induction (i.e. repeated cooling, heating and vigorous mixing), through activation of native lipase or through addition of bacterial lipase (e.g. from *Pseudomonas* spp.). The level of free fatty acids should be raised up to at least 5 meq/100 g of fat.

At least five levels are required. The effect exists if the slope of a linear regression equation of measurement result, y, versus free fatty acid concentration, x, is significantly different from 0,00.

5.2.3.2.2.4 Proteolysis

The possible effect of proteolysis can be monitored through artificial induction (i.e. using microflora proteases). A minimum range of 0,8 % soluble nitrogen in milk should be obtained.

At least five levels are required. The effect exists if the slope of a linear regression equation of measurement result, *y*, versus percentage mass fraction of soluble nitrogen, *x*, is significantly different from 0,00.

5.2.3.2.2.5 Lactic souring, pH

The possible effect of souring can be monitored through the addition of lactic acid.

At least five levels are required. Check that at the higher levels the milk does not clot at the water-bath temperature in order to prevent blockage or damage to the liquid flow systems.

The effect exists if the slope of a linear regression equation of measurement result, *y*, versus lactic acid concentration, *x*, is significantly different from 0,00.

5.2.3.2.3 Effect of interferent

5.2.3.2.3.1 General

The combination of cooling and storage at 0 °C to 6 °C with a preservative such as bronopol (2-bromo-2-nitropropan-1,3-diol) appropriately preserves clean (uncontaminated) milk samples. These conditions generally apply to calibration and control milk samples. In practice, sample conditions can differ (different type of chemical preservatives, transport, storage time and temperatures).

Therefore, it is of interest to determine the effect of preservation conditions on alternative method results. Based on this, adequate advice on proper sample handling and preservation can be provided to the workers involved.

This is achieved by analysing two identical sample sets with the method in one run, one set having the usual preservation and having undergone the usual handling conditions, the other set having been subject to optimal conditions, and then comparing the obtained results.

For each item, component concentrations should cover the usual concentration range in practice. Sample numbers of 30 to 40 are generally sufficient. Statistics to be used are the same as in the assessment of overall accuracy, using optimal conditions as reference. The mean and the standard deviation of differences provide information on the average error due to sample conditions tested and possible significance. Additionally, the effect on repeatability can be evaluated by analysing duplicates and comparing s obtained in both preservation/handling conditions using the *F*-test.

5.2.3.2.3.2 Effect of added chemicals (preservatives)

A possible effect on analytical results can be monitored through comparisons of identical parallel series of milk samples preserved with different chemical preservatives. Other preservation parameters shall be maintained equal in order not to influence the results. Evaluate the effect of both the nature and concentration of the preservative. The effect of the preservative shall be evaluated also in comparison with the instrument already certified and/or installed in the laboratory.

5.2.3.2.3.3 Effect of sample intake temperature

Analytical instruments can be sensitive to environmental conditions (i.e. humidity, temperature, vibrations). In particular, sample temperature can be a critical point with respect to the internal instrument temperature. A comparison at two extremes (specified lower and upper limit as advised by the manufacturer) on identical sets of different milk samples provides sufficient information.

5.2.3.2.3.4 Effect of storage conditions (time and temperature)

Sample temperature can affect the physical characteristics of milk components (i.e. crystallization of fat, solubility of casein and the mineral fraction).

Besides, storage time can determine the ability of milk to recover its native physical and chemical characteristics before being analysed. For instance, cream separated from skim milk becomes so firm that obtaining a homogenous sample is hampered when applying normal mixing conditions.

5.3 Method confirmation study (Phase II)

5.3.1 General

This part of the evaluation consists of various elements that determine the ability of the laboratory to produce analytical results within the time expected and at the cost expected or needed.

5.3.2 Verification of precision in routine conditions

During Phase II, two or three of the same type of instruments should be evaluated in one or more laboratories for at least two months. The instruments should be routinely used for at least 4 h per day during 5 consecutive days. In this test period, the following protocol should be applied.

Set of pilot samples at 3 levels of measurand should be measured each 15 min to 20 min in duplicate (n = 2) during 5 consecutive days. In the time between the pilot samples, routine milk samples should be measured in duplicate.

All samples should be representative for the milk under validation (single cow's milk, bulk cow's milk, milk from other species if required).

Before starting the validation analysis, the routine start up procedure of the instrument should be performed and give satisfactory results. If there is any deviation in the results with the daily start up then the procedure should be repeated or obtained results should not be contained in the statistical evaluation. Ensure that problems in the daily start up are noted and reported. Estimate, for each pilot level per day and for the entire Phase II, the performance as reported in 5.2.2.1.2.

The values obtained for s_r and s_{Rintra} should conform to the limits stated in Annex B.

The stability of the method response during the analyses of the pilot samples can be visualized by plotting the mean results of the three different pilot samples versus the time. In the report of the method, confirmation study opinions should be provided on criteria such as compliance with performance limits, simplicity, ruggedness, testing rate, user friendliness and safety. This can be done by grading (e.g. poor, medium, good) and including specific remarks. When the method or instrument is routinely applied in enough laboratories willing to join in an interlaboratory study, in accordance with ISO 5725-2 and IDF Bulletin 453/2012, this can be organized as part of Phase II in order to calculate the instrument precision figures, repeatability and reproducibility. In an interlaboratory study, for instruments that need a calibration adjustment/check it is required to supply the same set of calibration samples for slope and bias evaluation to each of the participants with clear instructions on their use. The instrument setting should be adjusted according to the manufacturer's instructions if necessary.

In addition to the interlaboratory study, to complete the verification, it is required to analyse the data from the pilot sample(s) from the past 12 months from two routine laboratories.

Additionally, the downtime of the instruments should be evaluated.

The alternative method should also be assessed for general convenience aspects such as speed, consumables, user-friendliness, safety in use and robustness.

For the validation of the manual method or instrument, Phase II will consist of measuring at least 200 routine samples in duplicate with at least two different instruments/apparatus and batch reagents.

5.3.3 Data collection

Via the application of internal quality control, results of quality checks, as recommended in ISO 8196-2 | IDF 128-2, should be collected and reported in order to consolidate the results of Phase I. They can be summarized in tables, indicating for each type of check the successive check parameter values collected during the test period, their means, minima and maxima, total check numbers and degree of conformity.

5.3.4 Pilot samples

A raw milk with three different levels of the measurands can be used as pilot samples. The different levels of the measurands can be obtained via additions of the components of interest to the raw milk. In addition, frozen pilot samples can be used for Phase I and Phase II after the stability of the parameters tested have been proven. When a new batch of pilot samples is prepared, the test results should be compared to the test result of the batch in use. For this, the current and the new batch should be analysed in parallel during at least one day of Phase II.

5.4 Report and approval delivery

5.4.1 General

Data and experience gathered in both Phases I and II shall be duly described in the validation report with all the necessary information on the evaluation course, tables with results on analytical performances, discussion, conclusions and summaries, and be made available for an eventual certification.

5.4.2 National validation

The overall report should comprise four documents:

a) a report of the test bed evaluation (Phase I), including raw results in annexes;

b) a report of the routine verification (Phase II);

c) a summary of conclusions of the Phases I and II;

d) a general conclusion in regard to the intended use of the method.

5.4.3 International validation

The reports of Phase I and Phase II obtained according to 4.1.5 are collected for transmission to the approval body by the requesting organization. If the three single national approvals (see 4.1.4) have been obtained, these can be considered as well. The formal request should be made according to the procedure defined by the international organization granting the international validation or approval. It should comprise the elements proving prior successful validations in the required number. Examples of this are:

a) the formal request for the validation/approval with appropriate forms, where available;

b) the technical documentation relating to the method/instrument (i.e. principle, device and capabilities, calibration version) as supplied by the manufacturer;

c) a report of the test bed evaluation (Phase I), including raw results in annexes;

d) reports of routine verifications (Phase II).

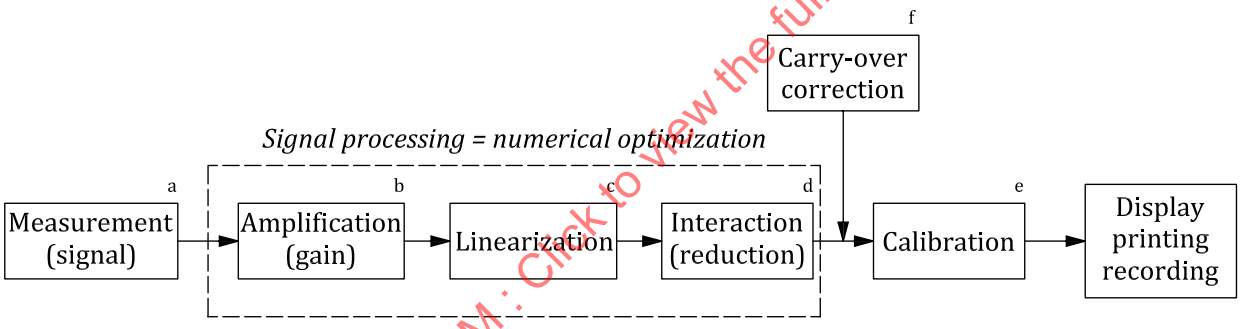
Annex A
(informative)

Measurement process and overall accuracy

Regardless of the indirect method used, Figure A.1 represents the measurement process schematically. Not all steps necessarily exist in every method or instrument. This depends on manufacturer choice in relation to the principle of the measurement and the component measured. For instance, only a small or negligible effect of fat and protein is to be expected in the interaction (reduction) step in somatic cell counting by fluoro-opto-electronic methods in milk, as interactions are normally overcome by dispersion through treatment with chemical reagents before the measurement.

In some cases, several steps can be combined, e.g. those lying within the dashed line box in Figure A.1 in particular with infrared devices. Nevertheless, in theory, the different steps of the signal processing can be set up in the instrument and remain available to be activated or not, through active or neutral mathematical matrices.

Interactions of major components or carry-over effects can be accommodated for adapting the principle of the method and/or the physical device (physical treatment, chemical reagents, tube length) and therefore no longer needing numerical corrections.



a Zero/blank, repeatability, stability, reproducibility.

b Sensitivity, measurement lower limit, repeatability.

c Linearity range, upper limit, accuracy.

d Effect of other milk components, accuracy.

e Suitability of manufacturer calibration system, accuracy.

f Effect of previous milk intake, repeatability, accuracy.

Every step of the measurement process corresponds to an element of the breakdown of overall accuracy of the method. Minimizing the overall error is achieved through minimizing every component, thereby optimizing every step of the measurement process. Then the experimental design for the evaluation of a milk analyser is defined in order to assess that every measurement step is correctly adjusted.

Figure A.1 — Example of a theoretical measurement process in conventional analysers

Annex B
(informative)

Limits for the performance characteristics with raw milk

B.1 Limits for milk with medium fat and protein content

The limits listed in Table B.1 apply to cow and goat milk samples with medium content in the ranges mentioned.

Table B.1 — Limit for a milk with medium content of fat and protein

Measurand (units)	Criteria limits						
	Fat g/100 g	**Protein**[a] g/100 g	**Lactose** g/100 g	**Urea** mg/100 g	**Freezing point** m °C	**pH**	**SCC** × 1 000 cells/ml
Range, ΔL_{range} — Whole	2,0 to 6,0	2,5 to 4,5	4,0 to 5,5	10,0 to 70,0	480 to 530	6 to 7,5	0 to 2 000
— Low (L)							0 to 100
— Medium (M)							100 to 1 000
— High (H)							> 1 000
Carry-over ratio limit, L_c	1	1	1				2 [b]
Sequence number, N_c	20	20	20	20			20
Minimum range, ΔL_{test}	4	3	1,5	45			500
Linearity: ratio limit, $\Delta e/\Delta L$	0,01	0,01	0,01	0,02			0,02
Replicate number for linearity, N_L	6	6	6	6			8
Maximum range, ΔL_{test}	4	4	4	100			2 000
Repeatability							
Average standard deviation of repeatability, s_r (filter instrument)	0,014	0,014	0,014				
Average standard deviation of repeatability, s_r (FT instrument)	0,008	0,008	0,008	1,4	1,1	0,02	
Average repeatability, r (filter instrument)	0,04	0,04	0,04				
Average repeatability, r (FT instrument)	0,02	0,02	0,02	3,92	3,08	0,056	
Relative standard deviation of repeatability, s_r, %							
s_r — Whole							4 %
s_r — Low (L)							6 %
s_r — Medium (M)							4 %
s_r — High (H)							3 %

[a] Same criteria limits for protein apply to crude protein, CP ($w_{CP} = w_{N,tot} \times 6,38$), true protein, TP [$w_{TP} = (w_{N,tot} - w_{N,non\text{-}P}) \times 6,38$] and casein, Cas [$w_{Cas} = (w_{N,tot} - w_{N,non\text{-}C}) \times 6,38$] where $w_{N,tot}$ is total nitrogen content, $w_{N,non\text{-}P}$ is non-protein nitrogen content and $w_{N,non\text{-}C}$ is non-casein nitrogen content.

[b] Limit for $C_{H/L}$.

Table B.1 *(continued)*

Measurand (units)	**Criteria limits**						
	Fat g/100 g	**Protein**[a] g/100 g	**Lactose** g/100 g	**Urea** mg/100 g	**Freezing point** m °C	**pH**	**SCC** × 1 000 cells/ml
	Intralaboratory reproducibility						
Average standard deviation, s_{Rintra} (filter instrument)	0,02	0,02	0,02				
Average standard deviation, s_{Rintra} (FT instrument)	0,014	0,014	0,014	2	18	0,025	
Average reproducibility, R_{intra} (filter instrument)	0,06	0,06	0,06				
Average reproducibility, R_{intra} (FT instrument)	0,04	0,04	0,04	5,6	5,04	0,070	
Relative standard deviation of reproducibility intra, s_{Rintra}, % s_{Rintra} — Whole s_{Rintra} — Low (L) s_{Rintra} — Medium (M) s_{Rintra} — High (H)							 5 % 7 % 5 % 4 %
	Calibration						
Mean bias, $\bar{d}$	±0,05	±0,05	±0,05	±1,2			
Relative mean bias, $\bar{d}_{rel}$							5 %
Slope, b	1 ± 0,05	1 ± 0,05	1 ± 0,10	1 ± 0,10			1 ± 0,05
	Accuracy **Comparison of alternative against reference method**						
	Individual animal milk samples						
Standard deviation, σ_{yx}	0,06	0,06	0,06	6	4	0,04	
Relative standard deviation of σ_{yx}							10 %
Number of individual animal milk samples, N_a	100	100	100	100	100	100	100
	Herd bulk milk samples						
Standard deviation, σ_{yx}	0,05	0,05	0,05	4	2	0,04	
Relative standard deviation of σ_{yx}							10 %
Number of herds, N_{h1}	5	5	5	5	5	5	5
Number of herd bulk milk samples	60	60	60	60	60	60	60

[a] Same criteria limits for protein apply to crude protein, CP ($w_{CP} = w_{N,tot} \times 6,38$), true protein, TP [$w_{TP} = (w_{N,tot} - w_{N,non\text{-}P}) \times 6,38$] and casein, Cas [$w_{Cas} = (w_{N,tot} - w_{N,non\text{-}C}) \times 6,38$] where $w_{N,tot}$ is total nitrogen content, $w_{N,non\text{-}P}$ is non-protein nitrogen content and $w_{N,non\text{-}C}$ is non-casein nitrogen content.

[b] Limit for $C_{H/L}$.

B.2 Limit for milk with high fat and protein content

The limits listed in Table B.2 apply to samples of sheep milk, buffalo milk and to milk of particular breeds of cow and goat with high fat and protein content in the ranges mentioned.

The indicative limits reported have been obtained from proficiency tests and validated studies if available.

Table B.2 — Limit for a milk with high content of fat and protein

Measurand (units)	Criteria limits						
	Fat g/100 g	**Protein**[a] g/100 g	**Lactose** g/100 g	**Urea** mg/100 g	**Freezing point** m °C	**pH**	**SCC** × 1 000 cells/ml
Range, ΔL_{range} — Whole	5,0 to 14,0	4,0 to 7,0	4,0 to 5,5	10,0 to 70,0			0 to 2 000
— Low (L)							0 to 100
— Medium (M)							100 to 1 000
— High (H)							> 1 000
Carry-over ratio limit, L_c	1	1	1				2 [b]
Sequence number, N_c	20	20	20	20			20
Minimum range, ΔL_{test}	4	3	1,5	45			500
Linearity: ratio limit, $\Delta e/\Delta L$	0,01	0,01	0,01	0,02			0,02
Replicate number for linearity, N_L	6	6	6	6			8
Maximum range, ΔL_{test}	4	4	4	100			2 000
	Repeatability						
Average standard deviation of repeatability, s_r (filter instrument)	0,014	0,014	0,014				
Average standard deviation of repeatability, s_r (FT instrument)	0,008	0,008	0,008	1,4			
Average repeatability, r (filter instrument)	0,04	0,04	0,04				
Average repeatability, r (FT instrument)	0,02	0,02	0,02	3,92			
Relative standard deviation of repeatability, s_r, %							
s_r — Whole							4 %
s_r — Low (L)							6 %
s_r — Medium (M)							4 %
s_r — High (H)							3 %
	Intralaboratory reproducibility						
Average standard deviation, s_{Rintra} (filter instrument)	0,025	0,025	0,025				
Average standard deviation, s_{Rintra} (FT instrument)	0,02	0,02	0,02				
Average reproducibility, R_{intra} (filter instrument)	0,07	0,07	0,07				

[a] The same criteria limits for protein apply to crude protein, CP ($w_{CP} = w_{N,tot} \times 6{,}38$), true protein, TP [$w_{TP} = (w_{N,tot} - w_{N,non\text{-}P}) \times 6{,}38$] and casein, Cas [$w_{Cas} = (w_{N,tot} - w_{N,non\text{-}C}) \times 6{,}38$] where $w_{N,tot}$ is total nitrogen content, $w_{N,non\text{-}P}$ is non-protein nitrogen content and $w_{N,non\text{-}C}$ is non-casein nitrogen content.

[b] Limit for $C_{H/L}$.

Table B.2 *(continued)*

Measurand (units)	Criteria limits						
	Fat g/100 g	**Protein**[a] g/100 g	**Lactose** g/100 g	**Urea** mg/100 g	**Freezing point** m °C	**pH**	**SCC** × 1 000 cells/ml
Average reproducibility, R_{intra} (FT instrument)	0,06	0,06	0,06				
Relative standard deviation of reproducibility intra, s_{Rintra}, %							
s_{Rintra} — Whole							5 %
s_{Rintra} — Low (L)							7 %
s_{Rintra} — Medium (M)							5 %
s_{Rintra} — High (H)							4 %
	Calibration						
Mean bias, $\overline{d}$	±0,10	±0,05	±0,05	±1,2			
Relative mean bias, $\overline{d}_{rel}$	±1,25	±1,5					±5 %
Slope, b	1 ± 0,05	1 ± 0,05	1 ± 0,10	1 ± 0,10			1 ± 0,05
	Accuracy Comparison of alternative against reference method						
	Individual animal milk samples						
Standard deviation, σ_{yx}	0,06	0,06	0,06	6			
Relative standard deviation of σ_{yx}							10 %
Number of individual animal milk samples, N_a	100	100	100	100			100
	Herd bulk milk samples						
Standard deviation, σ_{yx}	0,05	0,05	0,05	4			
Relative standard deviation of σ_{yx}							10 %
Number of herds, N_{h1}	5	5	5	5			5
Number of herd bulk milk samples	60	60	60	60			60

[a] The same criteria limits for protein apply to crude protein, CP ($w_{CP} = w_{N,tot} \times 6,38$), true protein, TP [$w_{TP} = (w_{N,tot} - w_{N,non\text{-}P}) \times 6,38$] and casein, Cas [$w_{Cas} = (w_{N,tot} - w_{N,non\text{-}C}) \times 6,38$] where $w_{N,tot}$ is total nitrogen content, $w_{N,non\text{-}P}$ is non-protein nitrogen content and $w_{N,non\text{-}C}$ is non-casein nitrogen content.

[b] Limit for $C_{H/L}$.

Table B.3 — Limits for milk with medium fat and protein content in case of comparison against a former validated instrument with same analytical principle and only minor technical changes

Measurand (units)	Criteria limits						
	Fat g/100 g	**Protein**[a] g/100 g	**Lactose** g/100 g	**Urea** mg/100 g	**Freezing point** m°C	**pH**	**SCC** × 1 000 cells/ml
Range, ΔL_{range} — Whole	2,0 to 6,0	2,5 to 4,5	4,0 to 5,5	10,0 to 70,0	480 to 53	6 to 7,5	0 to 2 000
— Low (L)							0 to 100
— Medium (M)							100 to 1 000
— High (H)							> 1 000
Carry-over ratio limit, L_c	1	1	1				2 [b]
Sequence number, N_c	20	20	20	20			20
Minimum range, ΔL_{test}	4	3	1,5	45			500
Linearity: ratio limit, $\Delta e/\Delta L$	0,01	0,01	0,01	0,02			0,02
Replicate number for linearity, N_L	6	6	6	6			8
Maximum range, ΔL_{test}	4	4	4	100			2 000
	Repeatability						
Average standard deviation of repeatability, s_r (filter instrument)	0,014	0,014	0,014				
Average standard deviation of repeatability, s_r (FT instrument)	0,008	0,008	0,008	1,4	1,1	0,02	
Average repeatability, r (filter instrument)	0,04	0,04	0,04				
Average repeatability, r (FT instrument)	0,02	0,02	0,02	3,92	3,08	0,056	
Relative standard deviation of repeatability, s_r, %							
s_r — Whole							4 %
s_r — Low (L)							6 %
s_r — Medium (M)							4 %
s_r — High (H)							3 %
	Intralaboratory reproducibility						
Average standard deviation, s_{Rintra} (filter instrument)	0,02	0,02	0,02				
Average standard deviation, s_{Rintra} (FT instrument)	0,014	0,014	0,014		1,8	0,025	
Average reproducibility, R_{intra} (filter instrument)	0,06	0,06	0,06				
Average reproducibility, R_{intra} (FT instrument)	0,04	0,04	0,04		5,04	0,070	
Relative standard deviation of reproducibility intra, s_{Rintra}, %							

[a] Same criteria limits for protein apply to crude protein, CP ($w_{CP} = w_{N,tot} \times 6,38$), true protein, TP [$w_{TP} = (w_{N,tot} - w_{N,non\text{-}P}) \times 6,38$] and casein, Cas [$w_{Cas} = (w_{N,tot} - w_{N,non\text{-}C}) \times 6,38$] where $w_{N,tot}$ is total nitrogen content, $w_{N,non\text{-}P}$ is non-protein nitrogen content and $w_{N,non\text{-}C}$ is non-casein nitrogen content.

[b] Limit for $C_{H/L}$.

Table B.3 *(continued)*

Measurand (units)	Criteria limits						
	Fat g/100 g	**Protein**[a] g/100 g	**Lactose** g/100 g	**Urea** mg/100 g	**Freezing point** m°C	**pH**	**SCC** × 1 000 cells/ml
s_{Rintra} — Whole							5 %
s_{Rintra} — Low (L)							7 %
s_{Rintra} — Medium (M)							5 %
s_{Rintra} — High (H)							4 %
	Calibration						
Mean bias, $\overline{d}$	±0,05	±0,05	±0,05	±1,2			
Relative mean bias, $\overline{d}_{rel}$							±5 %
Slope, b	1 ± 0,05	1 ± 0,05	1 ± 0,10	1 ± 0,10			1 ± 0,05
	Accuracy Comparison between two different instrument models						
	Individual animal milk samples						
Standard deviation, σ_{yx}	0,04	0,04	0,04	6			
Relative standard deviation of σ_{yx}							8 %
Number of individual animal milk samples, N_a	100	100	100	100			100
	Herd bulk milk samples						
Standard deviation, σ_{yx}	0,03	0,03	0,03	4			
Relative standard deviation of σ_{yx}							8 %
Number of herds, N_{h1}	5	5	5	5			5
Number of herd bulk milk samples	60	60	60	60			60

[a] Same criteria limits for protein apply to crude protein, CP ($w_{CP} = w_{N,tot} \times 6,38$), true protein, TP [$w_{TP} = (w_{N,tot} - w_{N,non\text{-}P}) \times 6,38$] and casein, Cas [$w_{Cas} = (w_{N,tot} - w_{N,non\text{-}C}) \times 6,38$] where $w_{N,tot}$ is total nitrogen content, $w_{N,non\text{-}P}$ is non-protein nitrogen content and $w_{N,non\text{-}C}$ is non casein nitrogen content.

[b] Limit for $C_{H/L}$.

Annex C
(informative)

Calculation examples

C.1 Assessment of preliminary instrumental fittings

C.1.1 Reproducibility intralaboratory executed during the same working day

Data from fat analysed by infrared spectroscopy (see ISO 9622 | IDF 141) are listed in Table C.1.

Table C.1 — Standard deviation of reproducibility intralaboratory

Check q	Replicate results x	Sum	Mean $\bar{x}\mu$	Mean bias $\bar{d}$	Test number n	Sum of squares S	Variance V	Standard deviation s_{ri}
1	4,00 4,03 4,01	12,04	4,013	0,008	3	0,000 467	0,000 233	0,015
2	4,02 4,03 4,02	12,07	4,023	0,018	3	0,000 067	0,000 033	0,006
3	4,01 4,00 4,00	12,01	4,003	−0,002	3	0,000 067	0,000 033	0,006
4	3,99 4,00 4,02	12,01	4,003	−0,002	3	0,000 467	0,000 233	0,015
5	3,99 4,01 4,01	12,01	4,003	−0,002	3	0,000 267	0,000 133	0,012
6	3,97 3,99 4,00	11,96	3,987	−0,018	3	0,000 467	0,000 233	0,015
7	4,01 4,00 3,98	11,99	3,997	−0,008	3	0,000 467	0,000 233	0,015
8	4,02 4,02 3,99	12,03	4,010	0,005	3	0,000 600	0,000 300	0,017
9	4,01 4,00 4,03	12,04	4,013	0,008	3	0,000 467	0,000 233	0,015

Table C.1 *(continued)*

Check q	Replicate results x	Sum	Mean $\bar{x}\mu$	Mean bias $\bar{d}$	Test number n	Sum of squares S	Variance V	Standard deviation s_{ri}
10	3,99 3,99 4,01	11,99	3,997	−0,008	3	0,000 267	0,000 133	0,012
Sum	120,150	120,150	40,050	0,000	30	0,003 60	0,001 80	
Average	4,005		4,005	0,000		0,000 180	0,000 180	0,013
Standard deviation			0,010 5	0,010				

Check on the homogeneity of variances using the Cochran test, as shown by Formulae (C.1) and (C.2):

$$L_{\text{Coch}}\,(p=0,95;2;10)=0,445$$

$$I_{\text{Coch}}=\frac{V_{\max}}{\Sigma V_i}<L_{\text{Coch}} \qquad \text{(C.1)}$$

$$I_{\text{Coch}}=\frac{0,000\ 30}{0,001\ 80}=0,166\,6 \qquad 0,166\,6<0,445$$

$$L_s=\sqrt{L_{\text{Coch}}\,\Sigma V_i} \qquad \text{(C.2)}$$

$$L_s=\sqrt{0,445\times 0,001\ 80}$$

$$L_s=0,028\ 3$$

where

I_{Coch} is the Cochran index;

L_{Coch} is the Cochran limit;

L_s is the standard deviation limit;

$s_{\text{obs},i}$ are the standard deviation values observed;

$V_{\max}$ is the maximum variance;

ΣV_i is the sum of variances.

As all observed standard deviation values are below the limit for the standard deviation, this implies that variance homogeneity is confirmed.

Standard deviation of intralaboratory reproducibility, as shown by Formula (C.3):

$$s_{R\text{intra}}=\sqrt{s_{\bar{x}}^2+s_r^2\left(1-\frac{1}{n}\right)} \qquad \text{(C.3)}$$

$$s_{R\text{intra}}=\sqrt{0,010\ 5^2+0,013\ 4^2\left(1-\frac{1}{3}\right)}$$

$s_{R\text{intra}} = 0{,}015 < 0{,}028$

This implies that intralaboratory reproducibility complies with the stated limit in Table B.1.

Standard deviation between checks, as shown by Formula (C.4):

$$s_c = (s_{\bar{x}}^2 - s_r^2 / n)^{1/2} \quad \text{(C.4)}$$

$$s_c = \sqrt{0{,}010\ 5^2 - \frac{0{,}013\ 4^2}{3}}$$

$s_c = 0{,}007$

Repeatability, as shown by Formula (C.5):

$$s_r = \left[\sum s_{r,i}^2 / q\right]^{1/2} \quad \text{(C.5)}$$

$$s_r = \sqrt{\Sigma \frac{0{,}001\ 80}{10}}$$

$s_r = 0{,}013 < 0{,}014$

This implies that repeatability complies with the stated limit in Table B.1.

Table C.2 — Instrument stability check and F test

Source of variation	Degrees of freedom v	Sum of squares S	Mean square S/v	Standard deviation s	F
Intercheck	9	0,002 950	0,000 328	0,018	1,821
Intracheck	20	0,003 600	0,000 180	0,013	
Total	29	0,006 550	0,000 226	0,015	

Because $F_{obs} = 1{,}82$ is smaller than $F_{0,95} = 2{,}39$, it can be concluded that stability is assessed positively: no significant shift of instrument response observed.

Another conclusion is that, as the residual standard deviation, at 0,013, is smaller than 0,014, instrument functioning is assessed positively: no abnormal individual fluctuation.

C.1.2 Carry-over effect

As an example, carry-over effect data from fat analysed by infrared spectroscopy (see ISO 9622 | IDF 141) are listed in Table C.3.

Table C.3 — Carry-over effect

Sequence number	Concentrations				Differences	
	L_{L1}	L_{L2}	L_{H1}	L_{H2}	$\lvert\Delta L\rvert_L$	$\lvert\Delta L\rvert_H$
1	0,00	−0,01	3,98	3,99	0,010	0,010
2	0,01	−0,01	3,99	4,01	0,020	0,020
3	0,00	−0,02	3,97	3,99	0,020	0,020
4	−0,01	−0,02	3,97	3,98	0,010	0,010
5	−0,01	−0,02	3,96	3,98	0,010	0,020
6	0,01	0,00	3,98	4,00	0,010	0,020
7	0,00	−0,02	3,99	4,01	0,020	0,020
8	0,01	−0,01	3,97	3,99	0,020	0,020
9	−0,01	−0,02	3,98	3,99	0,010	0,010
10	0,01	−0,01	3,99	4,00	0,020	0,010
Mean	0,001	−0,014	3,978	3,994	0,015	0,016
Standard deviation	0,009	0,007	0,010	0,011	0,005	0,005
Number, N	10	10	10	10	10	10
Student, t	—	—	—	—	9,00	9,80
Minimum	−0,01	−0,02	3,96	3,98	0,01	0,01
Maximum	0,01	0,00	3,99	4,01	0,02	0,02
Δ = max. − min.	0,02	0,02	0,03	0,03	0,01	0,01

Mean bias d_{LL} and d_{HL} are significant according to the Student t-test, $t_{0,975}$ = 2,26.

Table C.4 — Carry over confidence limit

	Value	Confidence limit		Conclusion
		Lower	Upper	
$C_{L/H}$ =	0,40	0,31	0,49	COR lower than 1 % ⇒ conformity
$C_{H/L}$ =	0,37	0,28	0,47	COR lower than 1 % ⇒ conformity

C.1.3 Assessment of linearity

C.1.3.1 As a first example, linearity data from fat analysed by infrared spectroscopy (see ISO 9622 | IDF 141) are listed in Table C.5.

Table C.5 — Linearity — Test sample set with progressive dilution of 10 % fat milk by skim milk

Level number	Dilution	Replicates			Mean concentration	Mean residual	Standard deviation
	mass per volume	1	2	3			
	x %				$\bar{y}$	$\bar{e}$	s_r
1	15,500	1,540	1,520	1,530	1,530	−0,023	0,010
2	20,350	2,020	2,020	2,020	2,020	-0,013	0,000
3	25,640	2,550	2,560	2,550	2,553	−0,003	0,006
4	31,180	3,100	3,110	3,120	3,110	0,005	0,010
5	34,800	3,490	3,480	3,490	3,487	0,024	0,006
6	39,800	3,970	3,990	4,000	3,987	0,029	0,015
7	45,150	4,500	4,500	4,510	4,503	0,016	0,006
8	50,500	5,020	5,020	5,010	5,017	0,000	0,006
9	56,650	5,610	5,630	5,620	5,620	−0,006	0,010
10	61,950	6,110	6,130	6,120	6,120	−0,030	0,010
Number, N	10	10	10	10	10	10	
Mean	38,152 0	3,791 0	3,796 0	3,797 0	3,794 7	−0,000 1	0,008 8
Standard deviation	15,488 7	1,529 1	1,537 5	1,532 8	1,533 2	0,019 3	
Minimum	15,500 0	1,540	1,520	1,530	1,530 0	−0,030 0	
Maximum	61,950 0	6,110	6,130	6,120	6,120 0	0,029 0	
Δ = max. − min.	46,450 0	4,570	4,610	4,590	4,590 0	0,059 0	
Slope	0,099 0						
Bias	0,018 5						
N replicates	30						
N mean	10						
Standard deviation of level biases (S_L)	0,013 0						

On performing a statistical test to evaluate the $\Delta e/\Delta L$ ratio, a Δe of 0,059 and ΔL of 4,590 give a value of 0,013. As this is greater than 0,01, this implies that linearity is inadequate.

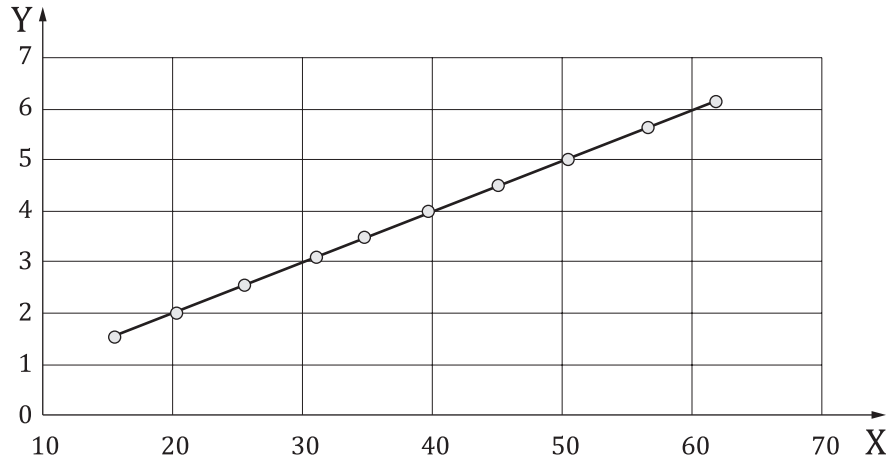
A second statistical test to evaluate bias from the linearity test using the standard deviation of residual means was performed. The value of F_{obs}, given by Formula (C.6):

$$F_{obs} = \left(s_r^2 + ns_L^2\right) / s_r^2 = ns_e^2 / s_r^2 \quad \text{(C.6)}$$

should be lower than $F_{0,95} = 2,45$ with $k_1 = q - 2$ and $k_2 = q(n - 1)$ degrees of freedom.

With $k_1 = 8$ and $k_2 = 20$, $F_{obs} = 16,17 > F_{0,95} = 2,45$, which implies that linearity is inadequate.

See Figures C.1 and C.2.

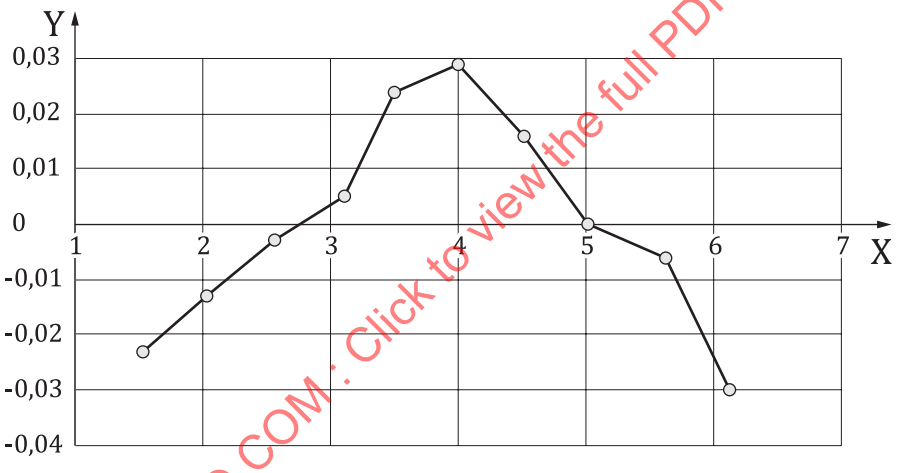


Key

X dilution mass/volume

Y instrument mean concentration

Figure C.1 — Linearity assessment — Instrumental response against dilution



Key

X instrument mean concentration

Y residuals

Figure C.2 — Linearity assessment — Mean residuals against dilution

C.2 Assessment of the overall accuracy

Table C.6 contains an example of fat analysed by infrared spectroscopy (see ISO 9622 | IDF 141) using a set of individual milk samples.

Table C.6 — Example of fat analyses

Sample number	Reference method	Instrumental method				Repeatability	Accuracy	
		Test 1	Test 2	Mean	Corrected results	Range	Bias	Residual
	y	x_1	x_2	$\bar{x}$	$y(x)$	$w = \lvert x_1 - x_2 \rvert$	$d = x - y$	$e = y - y(x)$
1	1,89	1,92	1,94	1,930 0	1,896 4	0,02	0,040	−0,006 4
2	1,98	2,05	2,06	2,055 0	2,025 3	0,01	0,075	−0,045 3
3	2,48	2,55	2,56	2,555 0	2,540 8	0,01	0,075	−0,060 8
4	2,66	2,56	2,56	2,560 0	2,546 0	0,00	−0,100	0,114 0
5	3,1	3,16	3,13	3,145 0	3,149 1	0,03	0,045	−0,049 1
6	3,23	3,2	3,22	3,210 0	3,216 2	0,02	−0,020	0,013 8
7	3,37	3,31	3,34	3,325 0	3,334 7	0,03	−0,045	0,035 3
8	3,57	3,51	3,5	3,505 0	3,520 3	0,01	−0,065	0,049 7
9	3,53	3,51	3,5	3,505 0	3,520 3	0,01	−0,025	0,009 7
10	3,52	3,57	3,57	3,570 0	3,587 3	0,00	0,050	−0,067 3
11	4,02	4	4,01	4,005 0	4,035 9	0,01	−0,015	−0,015 9
12	4,15	4,05	4,09	4,070 0	4,102 9	0,04	−0,080	0,047 1
13	4,59	4,52	4,51	4,515 0	4,561 7	0,01	−0,075	0,028 3
14	4,61	4,59	4,57	4,580 0	4,628 7	0,02	−0,030	−0,018 7
15	5,1	5,06	5,06	5,060 0	5,123 6	0,00	−0,040	−0,023 6
16	5,23	5,18	5,19	5,185 0	5,252 5	0,01	−0,045	−0,022 5
17	5,49	5,44	5,44	5,440 0	5,515 4	0,00	−0,050	−0,025 4
18	5,61	5,48	5,47	5,475 0	5,551 5	0,01	−0,135	0,058 5
19	5,8	5,74	5,76	5,750 0	5,835 0	0,02	−0,050	−0,035 0
20	5,89	5,8	5,78	5,790 0	5,876 3	0,02	−0,100	0,013 7
Number of data, N	20	20	20	20	20	20	20	20
Mean	3,991 0	3,960 0	3,963 0	3,961 5	3,991 0	0,014 0	−0,029 5	0,000 0
Standard deviation	1,260 0	1,223 2	1,219 3	1,221 2	1,259 1	0,011 0	0,059 5	0,045 8
Minimum	1,89	1,92	1,94	1,930 0	1,896 4	0,000 0	−0,135 0	−0,067 3
Maximum	5,89	5,80	5,78	5,790 0	5,876 3	0,040 0	0,075 0	0,114 0
Δ = max. − min.	4,00	3,88	3,84	3,860 0	3,979 9	0,040 0	0,210 0	0,181 4
Slope intercept	1,031 1 −0,093 5							

In conclusion, instrument accuracy conforms to limits defined for the component analysed and the type of milk fat in individual cow milk, by reference to the protocol and to Clause B.1, thus is assessed positively. The mean intercept and slope significantly differ from 0 and 1 and indicate that calibration can still be optimized further if needed for the purpose. See Figures C.3 and C.4.