THE GLOBAL STANDARD
FOR LIVESTOCK DATA

Protocol for the evaluation of milk analysers for ICAR approval

Version n°3 – 04/01/2002

1 Foreword

The present protocol has been produced by the Working Group on Milk Testing Laboratories.

Though various standards or normative documents already treat the subject of the evaluation of instrumental or indirect or alternative methods, there are as yet no documents with sufficient practical indications on the way to execute, and on the specific technical requirements to fulfil, in the evaluation of analytical routine methods for the particular aspect of the approval for milk recording by an (official) international body such as ICAR..

Therefore, it is the aim of the present document to define an overall procedure starting from the request for the approval, the procedure for the approval, the description of the technical evaluation needed, providing at the end the elements for a decision on approval.

The present document complies with ISO Standard 8196 (equivalent of IDF standard 128) and will concern milk of various species within the scope of ICAR (cows, goats, ewes, buffaloes) and the various components of interest for milk recording (fat, protein, lactose, somatic cell count, urea).

Contents :

1. Introduction :

Before being used for milk recording, a new analytical method or new equipment is to be submitted to an evaluation and must be approved for use by a competent body. At present, evaluations are carried out individually with, as a consequence, possible multiplication of evaluations in numerous countries. Moreover, the absence of a common protocol for such evaluations can result in incomplete and inaccurate technical information and numerous reports with non-comparable or partly comparable results.

The objective of this protocol is to define all relevant analytical parameters to be evaluated, providing respective limits to comply with in the relevant ranges for various animal species.

On the basis of this protocol, a limited number of evaluations should suffice to decide about an international approval on common ICAR rules for the application of analytical methods and/or equipment in milk recording.

1. Rules of the approval :

2.1. Stages of the evaluation and general principles :

Phase I: Every new instrument will be evaluated in specific conditions of test bed, within the period of time necessary to assess all the technical requirements prescribed in the present protocol. This part of the evaluation must be carried out by an expert laboratory specialised in analytical evaluations as well as experienced in (the) reference method(s) required. This laboratory should be accredited for this activity or be recognised as competent for this task by a competent body (national milk recording organisation and/or ICAR.

Phase II: The second phase of the evaluation starts after having succeeded with the first one. At least two new instruments will be used for a two-month period of observation in routine conditions in two different milk recording laboratories. They should fulfil the day-to-day quality control and satisfactorily respond to general convenience needs.

National approval: Request for an evaluation should be brought by manufacturers (or suppliers) to an official organisation (i.e. national milk recording, ministry, etc) who should appoint the laboratories to be involved in the evaluation and would give them an assignment for the work.
Reports of both phases I and II will be examined by an official committee. Then, on the basis of technical reports produced by laboratories, a national approval can be pronounced.

International approval: For an international approval by ICAR, the total evaluation should be renewed successfully in three ICAR countries and on similar bases

as defined in the protocol. Collation of reports and the request for ICAR approval should be made by manufacturers to ICAR. Milk analyser files will be submitted to the relevant ICAR Working Group(Milk Testing Laboratories) for a technical advice to the Board. Then the ICAR board will pronounce itself about the request for approval.

2.2. Field of validity of the approval :

An approval is given only :

a) for the field of application where the instruments has been evaluated (component, concentration range, animal species, etc) :

- In case milks of different animal species are to be analysed, specific evaluations for every species concerned have to be carried out to assess that the instrument is appropriate for the expected use.
- In case of breed with unusual milk fat and protein contents (i.e. Jersey breed with high fat and protein contents), the evaluation should be carried out within the same component range with milk of the specific breed.

b) for the specific instrument configuration used during the evaluation :

- In case of configuration changes, the proof should be brought that it does not affect the precision and the accuracy beyond acceptable limits.

Animal species and particularities of configuration(s) assessed should be carefully noted in the evaluation report.

Table 1 : Indicative milk component ranges at least to be covered by an evaluation

	cows	goats	ewes	buffaloes	units
Fat	2.0 – 6.0	2.0 –5.5	5.0 – 10.0	5.0 – 14.0	g/ 100 g
Protein	2.5 – 4.5	2.5 –5.0	4.0 – 7.0	4.0 – 7.0	g/ 100 g
Lactose	4.0 – 5.5	4.0 – 5.5	4.0 – 5.5	4.0 – 5.5	g/ 100 g
Urea	10.0 – 70.0	10.0 – 70.0	10.0 – 70.0	10.0 – 70.0	mg / 100 g
Cells	0 – 2000	0 – 2000	0 – 2000	0 – 2000	10^3 cells/ml

3. Course of operations of a technical evaluation :

➢ **Introduction to the principle of the evaluation (⇨explanatory note) :**

Whatever the indirect method is, a standard measurement processing can be presented by the scheme in figure 1. Each step does not necessarily exist in every instrument. This depends on manufacturers choices in relation to the principle of the measurement and the component measured – for example little or negligible effect (for instance step 3 in somatic cell count in cow's milk) - or in some case can be merged (for instance, steps 2, 3 and 4 in particular infra red devices) . Nevertheless, in theory the different steps of the signal process can be set up in the instrument and remain available to be activated or not, through active or neutral mathematical matrices. On the other hand interactions of major components or carry over effect can be eliminated by the method or the physical device (physical treatment, chemical reagents, tube length) and therefore no longer need numerical corrections.

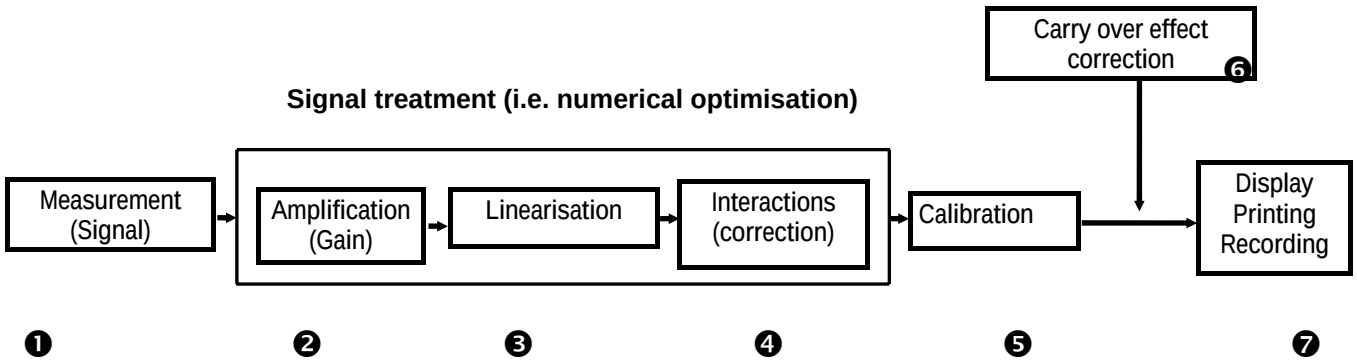


Figure 1 : Example of a theoretical measurement process in conventional analysers.
Every step of the measurement process corresponds to an element of the breakdown of overall accuracy of the method. Minimising the overall error is achieved through minimising every component thereby optimising every step 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.

❶	Measurement	zero/blank, repeatability, stability, reproducibility.
❷	Amplification	sensitivity, measurement lower limit ; repeatability.
❸	Linearisation	linearity range ; upper limit; accuracy.
❹	Interactions	effect of other milk components ; accuracy.
❺	Calibration	suitability of manufacturer calibration system ; accuracy.
❻	Carry over effect	effect of previous milk intake ; repeatability, accuracy.

Every step of the evaluation described in the following paragraphs can be required to fulfil appropriate limits for each analytical criteria (component) before starting up the next step.

3.1. Minimum necessary assessments for an evaluation :

This part defines and describes the elements of the evaluation which are compulsory to evaluate.
Whatever the method and precision element assessed, an evaluation is to be carried out from test results displayed expressed in standardised units and no prior data transformation should be performed (e.g. log or square root for somatic cell counting). Evaluation results should comply with specifications stated in the following paragraphs.

3.1.1. Assessment of preliminary instrumental fittings :

Before starting any further assessment, one has to verify basic criteria that indicate a proper functioning of the method or the instrument. These criteria are daily precision (including repeatability and short-term stability), carry-over and linearity.

3.1.1.1. Daily precision (repeatability and short-term stability) :

Basically, a milk analyser should present a signal stability which complies with the precision requirements. If not, the analyser is either in dysfunction (and should not be used) or its precision is not suitable for the objective of the analysis. Therefore, the instantaneous stability (repeatability) and the signal level stability have to be assessed prior to any other parameters.

Along a whole day period and every 15-20 minutes, analyse a same milk sample in triplicate by the instrument without any change in the adjustment of the calibration in order to obtain a minimum of 20 check test series. It should be preferably operated in as close as possible conditions as routine. Therefore sufficient number of samples should be planned to keep the instrument running between the periodical checks.

The precision will be evaluated at three different concentrations of each component, low, medium and high. To achieve this three different milk samples can be split in as many identical sub-samples as necessary for the analyses.

Using a one-way ANOVA, calculate the estimate of the standard deviation of repeatability (Sr), the standard deviation between check series (Sc) and the standard deviation of daily reproducibility (SR), referring to Annex A : $SR = (Sc^2 + Sr^2)^{1/2}$

The values Sr and SR obtained should comply with the limits stated for milk recording analysis (see Tables 2 and 3).

One can check the significance of the non-stability using a F-test. Alternatively, a one-way analysis of variance can be carried out to confirm the non-stability of signal.

3.1.1.2.Carry-over effect :

Strong differences in component contents between two successive milk samples analysed may influence the result of the latter one. It can happen because of an incomplete rinsing of the flow system and the measuring cell by milk circulation and/or a contamination of the former sample by the stirring device. The overall carry-over effect (including both sources of error) will be evaluated on the one hand and the rinsing efficiency of the flow system on the other hand.

Automated analysers often allow to apply on-line corrections to compensate the overall carry-over effect when necessary, therefore:
- rinsing efficiency of the flow system must be assessed by running tests without any correction (correction factor fit to zero) in manual mode (bypass the automated stirrer). Rinsing efficiency should not be less than 99 % or the internal carry-over should not exceed 1 %.
- overall carry-over effect will 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 for the component for milk recording purposes.

Method:

a) Analyses: Replicate as many times (n) as necessary the analytical sequence (L_L,L_L,L_H,L_H) where L_L is a low component concentration sample and L_H is a high component concentration sample.

b) Samples :
- Sufficient number of sub-samples of each sample L_L and L_H must be prepared prior to analysis in order to analyse each sub-sample only once.
- L_L and L_H should preferably be milks or liquids of similar viscosity as milk.
- Respective component concentrations must differ considerably. Depending on the component and the method, this can be achieved by using natural separation (creaming for fat), artificial separation (ultra-filtration for protein, micro-filtration for somatic cells)or addition (lactose and urea).
- For biochemical component determinations, concentrations of L_L and L_H should better be extreme values in the measuring range. At the contrary, for somatic cell count, one will assess the carry-over for three different high cell contents (500, 1000, 1500 10^3 cells/ml) and a single low cell content, preferably a zero cell milk.

c) Calculation:

- calculate the mean and the standard deviations of the differences $d_{Li} = L_{1i}-L_{2i}$ and $d_{Hi} = H_{2i}-H_{1i}$, respectively $\bar{d}_L$, Sd_L, $\bar{d}_H$, Sd_H.
- calculate the mean difference of concentration $\bar{d}_C = \bar{H}_2 - \bar{L}_2$

Then carry-over ratios C.O.R. and their standard deviations $S_{C.O.R.}$ are obtained using the following formulas:

C.O.R. (H/L) = $\bar{d}_L \cdot 100 / \bar{d}_C$ and $S_{C.O.R.(H/L)} = Sd_L \cdot 100 / (\bar{d}_C . \sqrt{n})$
C.O.R. (L/H) = $\bar{d}_H \cdot 100 / \bar{d}_C$ and $S_{C.O.R.(L/H)} = Sd_H \cdot 100 / (\bar{d}_C . \sqrt{n})$

As well, C.O.R. can be obtained by the equivalent formulas :

C.O.R. (H/L) = $(\sum L_1 - \sum L_2) \cdot 100 / (\sum H_2 - \sum L_2) = (\bar{L}_1 - \bar{L}_2) \cdot 100 / (\bar{H}_2 - \bar{L}_2)$
C.O.R. (L/H) = $(\sum H_2 - \sum H_1) \cdot 100 / (\sum H_2 - \sum L_2) = (\bar{H}_2 - \bar{H}_1) \cdot 100 / (\bar{H}_2 - \bar{L}_2)$

The two values obtained should not significantly differ from each other and should not exceed the limit ($L_{c.o.r.}$) stated for the component.

Note : *1) Acceptable limit for conformity: At the worst, the carry-over effect should not produce in the extreme case of lowest and highest concentration of the measuring range (ΔC) an error higher than the repeatability admitted for the method* $r=2.\sqrt{2}.Sr$

Therefore, the limit of c.o.r. can be defined as $L_{c.o.r.} = (r / \Delta C)x100$

A 1-2 % limit is generally recommended in standards.

2) Number (n) of analytical sequences: It can be defined in order to allow to estimate C.O.R. values with a ±20 % maximum relative confidence interval (i.e. 1±0,2 %).

Thus: $2.S_{C.O.R.} \leq 0{,}20 . (C.O.R.)$

⇨ $2.Sd . 100 / (\bar{d}_C . \sqrt{n}) \leq 0{,}20 . (\bar{d} . 100 / \bar{d}_C)$

⇨ $n \geq 100. (Sd / \bar{d})^2$

Between 10 and 20 analytical sequences are generally recommended in standards.

3.1.1.3.Linearity :

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

Nevertheless, newly developed indirect methods can be based on much less specific signal, still providing consistent results from multiple signals through multivariate statistical approaches. For these latter analysers linearity is no longer an absolute requirement in every case (though it must be in some specific utilisation of dairy industry, i.e. on processed milk with progressive contents stemming from concentration or dilution). Since then, for those methods and depending of analytical objectives, the step of linearity assessment can be discarded.The quality of the relationship with reference will be assessed in evaluating overall accuracy. In such a case, any routine measurement outside the calibration concentration range should be considered of doubtful quality and preferably not be used.

Linearity expresses the constancy of the ratio between the increase of milk component measured and the corresponding increase of the instrument measurement. Therefore linearity of the instrument signal is in most cases essential to maintain a constant sensitivity along the measuring range and to allow easy handling of calibration and fittings. Moreover, it allows in routine (to some extent) measurements beyond the concentration range of calibration through a linear extrapolation of calibration within the assessment range. Since then it can help to cope with possible particular limitations of reference methods (e.g. somatic cell count for goat's milk).

It can be assessed using sets of (n=8 to 15) samples with component concentrations regularly distributed all over the measuring range:

- samples should preferably be milks or liquids of similar physical characteristics (i.e. density, viscosity) as milk obtained by accurate dilution (weighing) of a high content sample by a low content one.
- concentrations should vary in regular intervals. Depending on the component, this can be obtained using various ways such as natural separation (creaming for fat), artificial

separation (ultra-filtration for protein, micro-filtration for somatic cells)and pure solutions (lactose and urea).
- assessment concentration range should be at least the ones stated in Table 1, §2.2. Nevertheless, it is up to the evaluator to extend linearity assessment range in order to determine the upper limit for acceptable measurements.
- reference for linearity will be either the volume mixing ratio (volume/volume or mass/volume) or theoretical concentrations calculated from the concentrations of the initial samples (one can refer to Annex A of IDF Standard 141).

Note : Independently of expression units, reference for linearity should be according to the intake measurement principle, that is volumetric in all milk analysers developed till today, at the opposite of milk weighting quite impracticable. Since then the theory would require volume/volume or mass/volume ratio.
Nevertheless, using mass/mass ratios provides identical figures when mixing liquids with the same density.
Analyse each sample in triplicate, first in the order of increasing concentrations, second in the order of decreasing concentrations and calculate the linear regression equation y=bx+a (y=instrument, x=dilution ratio) and the residuals ei (ei=yi-(bxi+a)) from the means of replicates and dilution rations. Plot the residuals ei (y axis) versus the dilution ratio (x axis) on a graph. A visual inspection of the data points will usually yield sufficient information about the linearity of the signal.

Calculate the ratio of the residual range to the signal values range :

$$De/DC = (e_{max} - e_{min}) / (C_{max} - C_{min})$$

where:

e_{max} and e_{min} = the upper and lower residuals, respectively
C_{max} and C_{min} = the upper and lower signal values, respectively

DE/DC should not exceed the limit stated for the component (generally 1-2 %) :

Criteria	F	P	L	Urea	SCC
Limits for De/DC	0.01	0.01	0.02	0.02	0.02

Alternatively, a one-way analysis of variance can be carried out to confirm the statistical significance of non-linearity and statistical tests of comparison of variances can be applied to confirm the significance of difference between residual variances (see Annex).

One way is to calculate polynomial regressions with a progressive increase of the degree to determine the most appropriate adjustment of the signal that is, providing minimum standard deviation Sy,x^k (the degree of the polynomial should better not exceed 3 with significant coefficient) and to compare the estimate Sy,x^k with Sy,x of linear regression on the basis of significant ratio or F-test.

The final judgement on linearity adjustment of instrument is :
- good if the value $Sy,x \leq Sy,x^k$
- correct if $Sy,x > Sy,x^k$ and $DE/DC \leq$ limit%
- incorrect if $Sy,x > Sy,x^k$ and DE/DC > limit%

Using the statistical test for comparison of residual variances or standard deviations (see Annex A).

3.1.1.4. Measurement limits :

Limits of an instrumental method measurement exist at both extremities of the analytical range, lower limit and upper limit.
It is not required to determine these limits in case where natural concentration ranges for the respective components and species are normally located far from zero (general case for biochemical components, i.e. fat, protein, lactose, urea) and within the range of linearity of the method. Determination and assessment of measurement limits are carried out with the evaluation of linearity.

3.1.1.4.1 Lower limits :

Lower limits evaluation is not treated by ISO 8196 (equivalent IDF 128) therefore reference can be made to standard EN ISO 16140:2000, which is dedicated to alternative microbiological methods, for definition and general principles.

At the date of the redaction of this document, only somatic cell counting is concerned by a lower limit evaluation for milk recording.

a) Definition :

Lower limits are defined in three ways depending on the risk of error accepted and a priori precision requirements :

- Critical level (CL) or decision limit which is the smallest amount which can be detected (not null), but not quantified as an exact value (risk β=50 %). Below it cannot be assumed that the value is not null :

$$CL = u_{1-\alpha} . \sigma \quad \text{or} \quad CL = 1.645 . \sigma \quad \text{with } \alpha = 5\% \qquad (1)$$

- Detection limit (DL) for which the second type of error is minimised up to a defined level, generally equal to the level of risk α (5 %). It consists in the lowest result, which differs significantly from zero (first type error α), that can be produced with a sufficiently low probability (second type error β) of including the blank value (zero) and with a sufficient confidence interval :

$$DL = (u_{1-\alpha} + u_{1-\beta}) . \sigma \quad \text{or} \quad DL = 3.29 . \sigma \quad \text{with } \alpha = \beta = 5\% \qquad (2)$$

- Quantification limit (QL) or determination limit which is the smallest amount of analyte which can be measured and quantified with a defined relative standard deviation SD% (or coefficient of variation CV%) :

$$QL = k_q . \sigma \quad \text{and} \quad SD\% = \sigma / QL \Rightarrow k_q = 1 / SD\%$$
$$QL = DL \Rightarrow k_q = 3.29 \Rightarrow SD\% = 30\% \qquad (3)$$

b) Limit values to fulfil : In somatic cell counting, DL of cell milk counters should not be higher than 5000 cells/ml and SD% (CV%) at the lower level (close to zero) should not exceed 30 %, with QL equal to DL.

c) Standard deviation σ :

In milk recording analysis, where only single determinations are carried out in routine, σ is the standard deviation of random error of the measurement that is, in the best case, the repeatability standard deviation at the proximity of zero content.

Standard deviation σ can be estimated in different ways :

- repeatability is dependent on concentration levels : standard deviation of repeatability (Sr) of the blank (zero) or estimated standard deviation at concentration values close to zero ;

- repeatability is not dependent on concentration levels : standard deviation of repeatability (Sr) estimated by taking benefit of replications at different levels in linearity assessment,

- repeatability and sample variance are not dependent on concentration levels : standard deviation $Sy_{(0)}$ of the single estimate $y_{(0)}$ for x=0 using linear regression equation calculated in a linearity assessment in a linear part close to zero :

$$Sy_{(0)} = S_{y,x}. (1 + 1/q + \bar{x}^2 / SCE_X)^{1/2}$$

Note: In that case, $Sy_{(0)}$ slightly overestimates σ as it takes into account sample errors and line estimation error in addition to repeatability.

3.1.1.4.2. Upper limit :

Upper limit corresponds to the threshold where the signal or the measurement deviates significantly from linearity (cf. linearity assessment).

An upper limit met on the range of concentration concerned by the evaluation will produce a ratio De/DC exceeding accepted limits (see linearity). Plotting linearity assessment results on a graph will provide necessary information on the shape of the curve response.

One can check if measured upper values deviating from linearity y_U differ significantly from $y(x_U)$ which should be obtained with the linear equation (prediction) calculated on the linear range without taking into account that result :

$t_{obs} = | y_U - y(x_U) | / Sy(x_U)$ with $S\, y(x_U) = Sy,x\, .\, (1 + 1/q + (x_U - \bar{x})^2 / SCE_X)^{1/2}$ and q-2 d.f. and $\alpha = 0,05$

- if $t_{obs} \leq t_{1-\alpha/2}$ => no deviation from linearity at that point
- if $t_{obs} > t_{1-\alpha/2}$ => significant deviation from linearity at that point

3.1.2. <u>Evaluation of the overall accuracy</u> :

One can refer to IDF standard 128 for a general information of this part of the evaluation.

The overall accuracy is composed by the sum of repeatability error, accuracy (or error of estimates versus reference) and error of calibration which occur in routine analytical conditions.
Each part of the overall accuracy is measured through the analysis of individual milk samples and herd milk samples of the specified animal species. Herd milk samples are to be collected in addition to individual milk samples in order to measure more accurately the part of variance related to herd effects.

The evaluation is to be performed on the instrument in the same state (working parameters, speed, calibration) the manufacturer intend to provide customers (users) with.
In case different analytical speed are available, parts of the overall accuracy will be assessed for the higher and the lower ones.

- **Calibration :** A preliminary calibration (or pre-calibration) is required and should be set in the instrument (or supplied with it) by the manufacturer with a detailed calibration procedure appropriate to the instrument.
  In case the instrument is to be used directly without any local calibration (set-in), instrumental analyses of the evaluation will be directly performed on appropriate (representative) milk samples.
  In case local calibration is necessary, prior calibration will be performed according to manufacturer recommendations and using instrument facilities, before starting up the evaluation.

- **Samples :** Milks have to be sampled and collected in optimum conditions such as no damages should occur and could produce erroneous repeatability estimate. Individual milks should cover the maximum concentration range of the component according to specification of § 2.2.
  - Calibration samples : They will be samples prepared according to recommend-ations of relevant standards for the criteria or, if no standardised procedure exits, in a similar way as prediction samples (half part for calibration and the other part for prediction).
  - Prediction samples : Minimum numbers of 100 individual milk samples collected in 4-6 different herds and 50 herd milk samples should be used.

- **Reference methods :** Reference methods should be standardised methods and, in all cases, the method used should be in a close agreement with one or more of the international reference methods (ISO, IDF, AOAC).

3.1.2.1. Assessment of repeatability :

Repeatability is the main criteria which indicates whether an instrument allows suitable results according user requirements or not and 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 to approval criteria.

Milk samples are to be analysed on the instrument calibrated according the manufacturer recommendations, preferably in duplicate. Indeed this minimum replicate number keeps

closer to true conditions of repeatability and prevents from possible damage on fat. Series of 15-20 milk samples are successively analysed twice after recovering initial analytical condition (i.e. temperature by heating) when necessary.

Then standard deviation of repeatability will be calculated from duplicate results obtained from the whole set of data and, for criteria covering a wide range of concentration –that is more than 1 log scale- (case of somatic cell count), part-by-part after splitting of the whole concentration range in different parts, three parts for the minimum (i.e. low, medium and high).

The standard deviation of repeatability will be calculated with the formula of IDF Standard 128 (see Annex I) : $Sr = (\sum w_i^{\,2} / 2q)^{1/2}$

where w_j is the difference between duplicates of sample i ($w_i = x_{1i} - x_{2i}$) and q the sample number.

Compare the values obtained (Sr) with the standardised repeatability values (σr) defined for the criteria and the application in Tables 2 and 3. It is expected that $Sr \leq \sigma r.(X^2_{1-\alpha} / q)^{1/2}$.

3.1.2.2. Assessment of accuracy of the mean :

According to IDF Standard 128 , the error of accuracy of the mean is broken down in the error of exactness of calibration and the error of accuracy (accuracy of estimates).
Statistical parameters to be used are those indicated in IDF Standard 128 and summed up in Annex I : $\bar{d}$; Sd ; Sy,x ; slope (b) ; student t test for $\bar{d}$ and b.
They are obtained from a simple linear regression calculated using means of duplicate instrumental results (x) and so-called reference results (y) obtained with a reference method recognised by ICAR (analyse in duplicates).

3.1.2.2.1. Assessment of accuracy :

Accuracy is assessed for individual animal milks and herd milks separately.
It is measured through the residual standard deviation Sy,x of the simple linear regression of instrumental results (x) and reference results (y).

It is expected that the differences to the regression line are normally distributed, therefore any outlying result should be carefully scrutinised. In case of outlying results, an other split sample of the same milk should be reanalysed by reference and the analyser when possible. When not or if outlying figure remains, reporting should present Sy,x estimates and graphs including all data –with the outliers identified, their number and respective biases- and the same Sy,x calculation 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 estimate value of Sy,x should fulfil respective limits $\sigma_{y,x}$ defined for individual milk samples and herd milk samples in Tables 2 and 3. It is expected that $Sy,x \leq \sigma_{y,x} .[X^2_{1-\alpha} / (q-2)]^{1/2}$.

For criteria covering a wide range of concentration –that is more than 1 log scale- (case of somatic cell count), accuracy evaluation should be performed for the whole range and for successive parts of the range after splitting of the whole concentration range in different parts, three parts for the minimum (i.e. low, medium and high).

Table 2 : Precision values for medium content milk samples (cows, goats).

	ICAR limits				
Criteria (units)	**F (g/ 100 g)**	**P (g/ 100 g)**	**L (g/ 100 g)**	**Urea (mg/ 100 g)**	**SCC (%)**
Repeatability					
Average Sr L / M / H	0.014 (1)	0.014 (1)	0.014 (1)	1.4 (2)	4 % (1) 8 % / 4 % / 2 %
Reproducibility					
Average SR (SR%) L / M / H	0.028 (1)	0.028 (1)	0.028 (1)	2.8 (2)	5 % (1) 10 % / 5 % / 2.5 %
Accuracy					
Animals Sy,x	0.10 (1)	0.10 (1)	0.15 (2)	6.0 (2)	10 % (2)
Herds Sy,x	0,07 (1)	0,07 (1)	0,07 (2)	4.0 (2)	10 % (2)

(1) Limits in accordance with IDF Standard 141C and 148A.
(2) Limits derived from experimental results and IDF 141C (SR≈2.Sr).

Note : For lactose IDF Standard 141C recommends the same limits for Sy,x as for fat and protein which are difficult to fulfil with regards to poor chemical method available as reference at that date.

Table 3 : Precision values for high content milk samples (ewes, buffaloes, particular cow/goat species).
(Derived from medium levels limits by applying relevant level ratios : 2 for F and P ; 1 for L and urea)

	ICAR limits				
Criteria (units)	**F (g/ 100 g)**	**P (g/ 100 g)**	**L (g/ 100 g)**	**Urea (mg/ 100 g)**	**SCC (%)**
Repeatability					
Average Sr (Sr%) L / M / H	0.028 (0.35 %)	0.028 (0.4 %)	0.014 (0.3 %)	1.4 (2)	4 % (1) 8 % / 4 % / 2 %
Reproducibility					
Average SR (SR%) L / M / H	0.056 (1) (0.7 %)	0.056 (1) (0.8 %)	0.028 (1) (0.6 %)	2.8 (2)	5 % (1) 10 % / 5 % / 2.5 %
Accuracy					
Animals Sy,x (Sy,x%)	0.20 (2.5 %)	0.20 (3.0 %)	0.15	6.0	10 %
Herds Sy,x (Sy,x%)	0,14 (1.75 %)	0,14 (2.0 %)	0,07	4.0	10 %

3.1.2.2.2. Assessment of exactness of calibration :

Prior to analyses, the instrument is calibrated according to the procedure recommended by the manufacturer and expressed in the same units as reference method used for the evaluation. Since then raw signals are not concerned and further statistical comparisons can be made at a same scale for both instrumental and reference values, allowing classical tests of identity and assessments against standardised target values. For this purpose, individual animal and herd samples will be analysed to provide the relevant information on the quality of the adjustment.

Depending on the principle of the method, quality of calibration can be more or less influenced by the representativeness of calibration samples in addition to calibration technique applied (i.e. mathematical model, experimental design, process). Therefore, sources of error of representativeness shall be reduced at the maximum for instance by sampling calibration samples in close or identical condition as for prediction milk samples.

Exactness of calibration is to be assessed using the parameters of the regression $y=b.x+a$: the mean bias $\bar{d}$ and the slope b (see Annex I and IDF Standard 128) taking care of eventual outlying results as in 3.1.2.2.1. Estimates $\bar{d}$ and b should normally fulfil the limits in Table 4. Failing that goal should normally imply further investigations or explanations.

Table 4 : Tentative indicative ICAR limits for exactness of calibration assessment.

a) Medium level (cows, goats) :

Criteria	F	P	L	Urea	SCC
Mean bias $\bar{d}$	±0.05 (1)	±0.05 (1)	±0.05 (1)	±2.5 (2)	±5 % (2)
Slope b	1±0.05 (1)	1±0.05 (1)	1±0.05 (1)	1±0.05 (2)	1±0.05 (2)

b) High level (ewes, buffaloes, goats) :

Criteria	F	P	L	Urea	SCC
Mean bias $\bar{d}$	±0.10 (1)	±0.10 (1)	±0.10 (1)	±2.5 (2)	±7 % (2)
Slope b	1±0.05 (1)	1±0.05 (1)	1±0.05 (1)	1±0.05 (2)	1±0.07 (2)

(1) Limits in accordance with IDF Standard 141C.
(2) Limits derived from experimental results.

3.2. Additional informative investigations :

The following items are not compulsory elements to evaluate even though they are of interest as possible parts of the overall accuracy of the method and the knowledge one can get about the method may have implications in milk sample handling (sampling, preservation, shipping, etc). Therefore, they can be considered as only informative for a proper use of the method if it obtains ICAR approval thanks to the former part. Nevertheless, for ICAR approval and the common knowledge, it would be very useful that they are evaluated once when the information is not available from manufacturers.

3.2.1. Ruggedness :

Ruggedness is the ability of an instrument not to be influenced by external elements other than the component measured itself. Possible effects can come from concentration variation of major milk component or interactions (depending on the instrument, they can be compensate by intercorrections), biochemical changes of milk component related to preservation (lipolysis, proteolysis, lactic souring) or chemicals added in milk such as preservatives.
Principle of robustness measurement is to produce a significant change in the concentration of each interacting component separately and measure the corresponding measurement change of the influenced component. Then, one calculates the ratio (difference observed)/(change introduced) expressed in the relevant units.

3.2.1.1. Effect of major milk components (interactions) :

For milk composition (fat, protein, lactose), one will refer to Annex B of IDF Standard 141 for sample preparation and calculation procedures : single variation method or multiple variation method by recombination of non correlated milk sample sets.

Effect of urea on other component measurements will be evaluated by addition of urea in milk as it is proposed for lactose.

Effect of high fat and protein content on somatic cell count in milk (ewes, goats and buffaloes) will be evaluated using cream (natural creaming) and milk retentate according a recombination in a similar way as in IDF 141

Effect should be better measured at three relevant levels for the component under interaction and the species (i.e. low, medium and high). A minimum of two strongly different level are required and better three.

3.2.1.2.Effect of biochemical changes in components :

Biological changes in milk result usually in damages to milk component affected. They can be produced by bacteria growths in milk or enzymes activity which affect directly or not milk components measured in milk. Unless achieved souring turning into milk clotting, there are no quick and easy way to distinguish such milks from well preserved milk samples and they are normally analyses. Then, the sensitivity of the method of measurement of milk components to such ways of deterioration can be of interest, in particular in order to evaluate the suitability of sampling and shipping conditions of routine milk recording of which depends the preservation quality of samples.
Clotting, churning and oiling are more evident defects of milk of which effects on analytical results are drastic for the first (no way for analysis) or depends essentially of the homogeneity of milk and representativeness of intakes. In those cases, defects can be easily identified and samples discarded.

3.2.1.2.1. Lipolysis :

One will relate modifications in the measurement with the most appropriate indicator of lipolysis (milk fat acidity) after an artificial induction of an increased lipolysis (cooling and action of native lipase or addition of bacterial lipase (i.e. Pseudomonas). One will raise lipolysis level up to 5 meq/100 g fat minimum.
At least 5 levels are required. The effect exists if the variation ration calculated (slope of linear regression) is significantly different from 0.00.

3.2.1.2.2. Proteolysis :

One will relate modifications in the measurement with the most appropriate indicator of proteolysis (whey protein or soluble nitrogen SN) after having achieved a proteolysis (i.e. using microflora proteases). One will try to obtain a minimum range of 0.8 % SN in milk.
At least 5 levels are required. The effect exists if the variation ration calculated (slope of linear regression) is significantly different from 0.00.

3.2.1.2.3. Lactic souring :

One will proceed by addition in milk of increasing amount of lactic acid. At least 5 levels are required. Check that the higher level does not clot at the water-bath temperature in order not to damage the instrument liquid system.

One will relate modifications in the measurement with the amount of lactic acid added. The effect exists if the variation ration calculated (slope of linear regression) is significantly different from 0.00.

3.2.1.3. Effect of sample history and handling conditions :

Condition of optimal preservation of milk samples are well known but often not fit at the optimum for economical reasons. For instance, combination of cooling and storage at about 4°C with a preservative such as bronopol (2-bromo 2-nitro 1,3-propandiol) is known to allow a quite good preservation for clean (uncontaminated) milk samples. These optimal conditions are in most cases applied to calibration and control milk samples. Different conditions for sample preservation may exist in one laboratory depending of the origin (different type of chemical preservatives, life-times and temperatures) .
Therefore it is of interest to determine how far differences in preservation conditions can affect analytical results obtained by the instrument and provide the relevant information to milk recording organisation and laboratories for good choices in analytical apparatuses and sample handling methods.
For each item, for practical conclusions, component concentrations should cover the usual range of routine and sample number per series be defined in order to allow to conclude to positive effects through statistically significant differences (30 to 40 is generally sufficient).

3.2.1.3.1. Effect of chemicals added (preservatives):

Differences in analytical results will be measured by comparisons of identical parallel series of milk samples preserved with different chemical preservative used in routine conditions. Other preservation parameters must be maintained equal not to bias the results. The effect of both nature and concentration is to be evaluated.

3.2.1.3.2. Effect of milk intake temperature :

Analytical instrument may be sensitive to environmental conditions related to their analytical principle (i.e. humidity, temperature, vibrations) and dispose of systems to compensate these sources of dysfunction. Indications are given by manufacturers regarding cautions to be taken by users in particular for sample temperature with respect to internal instrument temperature. Then, it is a useful information to know how large is the effect within the range of temperature of milk samples analysed in routine and allow to refine sample preparation before analysis (i.e. heating temperature and time). The comparison of effect of two extremes limits (lower and upper) advised by manufacturers on identical set of different milk sample will provide with a sufficient information.

3.2.1.3.3. Effect of storage conditions (i.e. time and temperature) :

Sample temperature can determine the physical aspect of milk components (i.e. crystallisation of fat glycerides ; solubility of casein and mineral fraction).
Besides, storage time can determine the ability of milk to recover its native physical and chemical aspects before being analysed. It is often the case that cream separated from skim milk becomes so firm that difficulties in reincorporating it uniformly in milk can occur. In

such cases, fat globule clusters can remain and be source of troubles in the instrument (i.e. milk homogenisation in infra red devices). The effect of various couples (time x temperature) can be measured by comparison with an optimal preservation method defined as reference.

3.2.2. Practical conveniences : (Phase II)

It consists in various elements of which depends the laboratory ability to produce analytical results within the time expected and at the cost expected or needed. These practical and economical elements are evaluated during Phase II of the course of the total evaluation, on a period of time and a number of laboratories such as stated in paragraph 2.

3.2.2.1.Speed :

Speed announced by the manufacturer will be verified. Precision performances should be reported with the information on the speed used when different speeds are available and were successfully tested in Phase I.

3.2.2.2.Robustness :

Frequency of troubles and servicing operations will be registered with the nature of incidents happened.

3.2.2.3.Monitoring and servicing facilities :

Convenience for the utilisation of calibration procedure will be noted with user-friendliness of interfaces and software. Easiness for troubleshooting and operating reparations and servicing will be noted as well as weak points of devices in order users to be able aware of them and be able to cope with them.

3.2.2.4.Validation of precision in routine conditions :

Via the application of the internal quality control according to recommendations of relevant guidelines of ICAR routine checks will be applied on instruments during Phase II of the evaluation and results will be registered and reported to complement the report of Phase I.

4. **Report and approval delivery :**

Evaluation reports for both Phases I and II will be duly reported in specific documents with all the necessary information on the evaluation course, tables of results of analytical performances measured, discussion or comments and summaries.

Raw results will be available on paper format and magnetic records for microcomputer (i.e. magnetic diskett, CD) and in a record format compatible with usual data calculation programmes.

The report material will be provided to ICAR by the organisation asking for ICAR approval according to conditions defined in paragraph 2.

Comments to be addressed to :

Olivier Leray, CECALAIT, BP 70129, F-39802 POLIGNY Cédex
Tel. +33/ 3 84 73 63 20 Fax. +33/ 3 84 73 63 29
E-mail : O.Leray@CECALAIT.Fr>

- Annex A -

USUAL STATISTICAL FORMULAS FOR METHOD EVALUATIONS

APPLICATION IN THE ASSESSMENT OF THE PRECISION

Standard deviation of repeatability : (q levels and n replicates)

$$Sr = \left(\sum_{i=1}^{q} \left(\sum_{j=1}^{n} (x_{ij} - \bar{x}_i)^2 \right) / q.(n-1) \right)^{1/2} \quad \Rightarrow \quad Sr = \left(\sum_{i=1}^{q} (x_{1i} - x_{2i})^2 / 2q \right)^{1/2}$$

when n=2

or $$Sr = \left(\sum_{i=1}^{q} w_i^{\,2} / 2q \right)^{1/2}$$

Standard deviation of daily reproducibility: (q check tests and n replicates)

$$SR^2 = S_{\bar{x}}^2 - Sr^2.(1-1/n)$$

and $$SR^2 = Sc^2 + Sr^2$$

Standard deviation between control test checks: $$Sc = (S_{\bar{x}}^2 - Sr^2/n)^{1/2}$$

APPLICATION IN THE ASSESSMENT OF THE ACCURACY

Means:

$$\bar{x} = \sum xi / q$$
$$\bar{y} = \sum yi / q$$
$$\bar{d} = \sum di / q = (\sum xi - \sum yi) / q = \bar{x} - \bar{y}$$

Sum of Squares and of products :

$$SCE_X = \sum (xi - \bar{x})^2 = \sum xi^2 - (\sum xi)^2 / q$$
$$SCE_Y = \sum (yi - \bar{y})^2 = \sum yi^2 - (\sum yi)^2 / q$$
$$SCE_d = \sum (di - \bar{d})^2 = \sum di^2 - (\sum di)^2 / q$$
$$SPE_{XY} = \sum (xi - \bar{x}).(yi - \bar{y}) = \sum xi.yi - \sum xi.\sum yi / q$$

Slope : $$b = SPE_{XY} / SCE_X$$

Intercept : $$a = \bar{y} - b . \bar{x}$$

Estimate for x : $$\underline{y}(x) = bx + a$$

Conditional mean for x : $$\bar{y}(x) = b\bar{x} + a$$

Residual (e) : $$e = y - \bar{y}(x) = y - b.x - a$$

Difference (d) : $$d = x - y$$

Correlation coefficient: $$r = (SPE_{XY}^2 / (SCE_Y . SCE_X))^{1/2}$$

Standard deviations of :

- **differences (d) :** $Sd = (SCE_d / (q-1))^{1/2}$
  $Sd = ((SCE_Y^2 + SCE_X^2 - 2.SPE_{XY}) / (q-1))^{1/2}$
- **residuals (ei):** $Sy,x = (\sum(yi-b.xi-a)/(q-2))^{1/2}$
  $Sy,x = ((SCE_Y^2 - SPE_{XY}^2 / SCE_X) / (q-2))^{1/2}$
  $Sy,x = (SCE_Y.(1-r^2) / (q-2))^{1/2}$
- **slope (b) :** $S_b = Sy,x / SCE_X^{1/2}$
- **intercept (a) :** $S_a = Sy,x . (1/q + \bar{x}^2 / SCE_X)^{1/2}$
- **conditional mean $\bar{y}(x_0)$:** $S\ \bar{y}(x_0) = Sy,x . (1/q + (x_0 - \bar{x})^2 / SCE_X)^{1/2}$
- **estimate $y(x_0)$:** $S\ y(x_0) = Sy,x . (1 + 1/q + (x_0 - \bar{x})^2 / SCE_X)^{1/2}$

<u>Conformity tests :</u>

- **conformity of an estimate :**
  - slope b versus 1,000 : $t_{obs} = | b-1,000 | / S_b \leq t_{1-\alpha/2}$
    with q-2 d.f. and $\alpha = 0,05$
  - slope b versus 0,00 : $t_{obs} = | b | / S_b \leq t_{1-\alpha/2}$
    with q-2 d.f. and $\alpha = 0,05$
  - mean difference $\bar{d}$ versus 0,00 : $t_{obs} = | \bar{d} | / (Sd / \sqrt{q}) \leq t_{1-\alpha/2}$
    with q-1 d.f. and $\alpha = 0,05$

  or (when $b \neq 1,000$) $\bar{x}$ versus $\bar{y}$: $t_{obs} = | \bar{x} - \bar{y} | / (Sy,x / \sqrt{q}) \leq t_{1-\alpha/2}$
  with q-2 d.f. and $\alpha = 0,05$

  - intercept a versus 0,00 : $t_{obs} = | a | / S_a \leq t_{1-\alpha/2}$
    with q-2 d.f. and $\alpha = 0,05$
  - conditional mean $\bar{y}(x_0)$ versus reference value y_o or residual e_o versus 0,00

    $e_o = y_o - \bar{y}(x_0) = y_o - b_{q-1}.x_o + a_{q-1}$

    $S_{\bar{y}(x0)} = Sy,x_{q-1}.(1/(q-1) + (x_o - \bar{x}_{q-1})^2 / SCEx_{q-1})^{1/2}$

    $t_{obs} = | e_o | / S_{\bar{y}(x0)} \leq t_{1-\alpha/2}$
    with q-3 d.f. and $\alpha = 0,05$

  ⇨ for outlier detection or departure from linearity : One checks whether point Mo (x_o,y_o) belongs to the linear curve calculated without that point.

- **conformity of a standard deviation S versus σ :**

  - <u>method 1 (Chi^2)</u> : $\sigma^2 \leq k.S^2 / X^2_{1-\alpha} \Rightarrow S \leq \sigma.(X^2_{1-\alpha} / k)^{1/2}$
    with k d.l. and $\alpha = 0,05$

  - <u>method 2 (error standard)</u> : (can replace method 1 for k > 50)
    $S - u_{1-\alpha} . S / \sqrt{2k'} \leq \sigma \Rightarrow S \leq \sigma / (1 - u_{1-\alpha} / \sqrt{2k'})$
    with k' data and $\alpha = 0,05$

- **Linearity test :**

  - comparison of a line with a k degree polynomial (reduction of residual error by) :
    $F_{obs} = ((n-2).Sy,x^2 - (n-k-1).Sy,x^{k\,2}) / (k-1).Sy,x^{k\,2} < F_{1-\alpha}$
    or $Sy,x / Sy,x^k < ((F_{1-\alpha}.(k-1) + (n-k-1)) / (n-2))^{1/2}$

    with : n samples, k polynomial degrees,
    k1 = k-1, k2 = q-k-1 and α risk of error.

  - Sample or level effect interpreted as linearity compared to repeatability :

    $F_{obs} = (n.Ss^2 + Sr^2)/ Sr^2 = (n.(S_{\bar{d}}^2 - Sr^2/n) + Sr^2)/ Sr^2 = n.S_{\bar{d}}^2 / Sr^2 < F_{1-\alpha}$
    or $S_{\bar{d}} / Sr < (F_{1-\alpha}/n)^{1/2}$

    with : n replicates, $\bar{d}$ = means difference of replicates,
    k1 = q-2, k2 = q.(n-1) and α risk of error.

<u>Note</u> : k1 = q-1 when testing the effect of a source of variation with no regression (1 way-ANOVA)

- ANNEX B -

EXAMPLES OF CALCULATION AND PRESENTATION

ANNEX - EXAMPLES

EVALUATION OF MILK ANALYSERS FOR ICAR APPROVAL

1. Assessment of preliminary instrumental fittings :

1.1. Daily precision : Example of fat analysed by infra red spectroscopy (cf. IDF 141)

Test No q	Replicates	Sum	Mean m	Mean bias d	Test number n	Sum of squares SOS	Variance Var	Within check Sr(i)
1	4,00 4,03 4,01	12,04	4,013	0,008	3	0,000467	0,000233	0,015
2	4,02 4,03 4,02	12,07	4,023	0,018	3	0,000067	0,000033	0,006
3	4,01 4,00 4,00	12,01	4,003	-0,002	3	0,000067	0,000033	0,006
4	3,99 4,00 4,02	12,01	4,003	-0,002	3	0,000467	0,000233	0,015
5	3,99 4,01 4,01	12,01	4,003	-0,002	3	0,000267	0,000133	0,012
6	3,97 3,99 4,00	11,96	3,987	-0,018	3	0,000467	0,000233	0,015
7	4,01 4,00 3,98	11,99	3,997	-0,008	3	0,000467	0,000233	0,015
8	4,02 4,02 3,99	12,03	4,010	0,005	3	0,000600	0,000300	0,017
9	4,01 4,00 4,03	12,04	4,013	0,008	3	0,000467	0,000233	0,015
10	3,99 3,99 4,01	11,99	3,997	-0,008	3	0,000267	0,000133	0,012
Sum **Average** **SD**	120,150 4,005	120,150	40,050 4,005 0,010	0,000 0,000 0,010	30	**0,00360** 0,000180	0,00180 0,000180	**0,013**

Check homogeneity of variances within checks :
Thanks to : Cochran Index = Var(max) / Sum of Var < Cochran limit => SD limit = (Cochran limit x Sum of Var)1/2

=> Cochran limit (P=0,95 ; 2 ; 10) = 0,445 **=> SD limit =** 0,0283 never smaller than SD values observed => variance homogeneity admitted

Daily reproducibility :	SR=(Sm2 - Sr2.(1-1/n))1/2	**SR =**	0,015	< 0,028 => conform to IDF 141
Variation between checks :	Sc = (Sm2 - Sr2/n)1/2	**Sc =**	0,007	
Repeatability :	Sr = (Sum Sr(i)2 / q)1/2	**Sr =**	0,013	< 0,014 => conform to IDF 141

Source of variation	df	Sum of squares	Mean squares	SD	F
Betwen tests	9	0,002950	0,00032778	0,018	1,821
Within tests	20	0,003600	0,00018	0,013	
Total	29	0,00655	0,00022586	0,015	

Conclusions :
1- From Fobs = 1,82 smaller than F0,95 = 2,39, stability is assessed positively : no significant shift of instrument response observed
2- From residual SD =0,013 smaller than Sr=0,014, instrument functioning is assessed positively : no abnormal individual fluctuation

ANNEX - EXAMPLES

EVALUATION OF MILK ANALYSERS FOR ICAR APPROVAL

1. Assessment of preliminary instrumental fittings :

1.2. Carry over effect : Example of fat analysed by infra red spectroscopy (cf. IDF 141)

Sequence	Concentrations				Differences	
N°	LL1	LL2	HL1	HL2	dL	dH
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
Std dev.	0,009	0,007	0,010	0,011	0,005	0,005
N	10	10	10	10	10	10
t-Student					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
D=Max-Min	0,02	0,02	0,03	0,03	0,01	0,01

Mean bias dL and dH are significant according to t-Student test t 0,975 = 2,26

		Value	Conf. Min	Conf. Max	
C.O.R. (H/L)	=	0,37	0,28	0,49	C.O.R. lower than 1 % => conform
C.O.R. (L/H)	=	0,40	0,31	0,47	C.O.R. lower than 1 % => conform

EVALUATION OF MILK ANALYSERS FOR ICAR APPROVAL

1. Assessment of preliminary instrumental fittings :

1.3. Assessment of linearity : Example of fat analysed by infra red spectroscopy (cf. IDF 141)

sample set of progressive dilution of a 10 % fat milk by skim milk

Test No	% dilution (m/v) X	Replicates 1	2	3	Mean concent. C Y	Mean residual e	Std Dev. Sr
1	15,50	1,54	1,52	1,53	1,530	-0,023	0,010
2	20,35	2,02	2,02	2,02	2,020	-0,013	0,000
3	25,64	2,55	2,56	2,55	2,553	-0,003	0,006
4	31,18	3,10	3,11	3,12	3,110	0,005	0,010
5	34,80	3,49	3,48	3,49	3,487	0,024	0,006
6	39,80	3,97	3,99	4,00	3,987	0,029	0,015
7	45,15	4,50	4,50	4,51	4,503	0,016	0,006
8	50,50	5,02	5,02	5,01	5,017	0,000	0,006
9	56,65	5,61	5,63	5,62	5,620	-0,006	0,010
10	61,95	6,11	6,13	6,12	6,120	-0,030	0,010
Level N	10,0	10,0	10,0	10,0	10,0	10,0	
Mean	38,152	3,791	3,796	3,797	3,795	0,000	**0,009**
SD	16,428	1,622	1,631	1,626	1,626	**0,020**	
Minimum	15,500	1,540	1,520	1,530	1,530	-0,030	
Maximum	61,950	6,110	6,130	6,120	6,120	0,029	
D=Max-Min	46,450	4,570	4,610	4,590	4,590	0,059	

Linear regression on :	Replicates	Means
Slope :	0,09898	0,09898
Intercept :	0,01856	0,01856
N	30	10

SD of residual means : Se = 0,0203
SD of repeatability : Sr = 0,0088
SD of level bias : Sl = 0,0197 (calculated by Sl = (Se2-Sr2/n)1/2)

Tests :

a- Ratio : De = 0,059
DC = 4,590

De/DC = 0,013 < 0,01 => Conclusion : Linearity default

b- Bias from linearity test using Sd of residual means : Fobs=(Sr2 + n.Sl2) / Sr2 = n.Se2/Sr2 should be lower than F0,95 = 2,45 with k1=q-2 and k2=q.(n-1)

Fobs = 16,17 > F0,95 = 2,45 => Conclusion : Linearity default
k1 = 8
k2 = 20

c- ANOVA from linear regression on the individual data : (equivalent to b-)

Source of variation	df	Sum of squares	Mean squares	SD	F
Regression	1	63,4522972	63,4522972	7,966	827638,66
Between levels	8	0,009916	0,001239516	0,035	16,17
Within levels	20	0,001533	7,66667E-05	0,009	
Total	29	63,46374667	2,188405057	1,479	

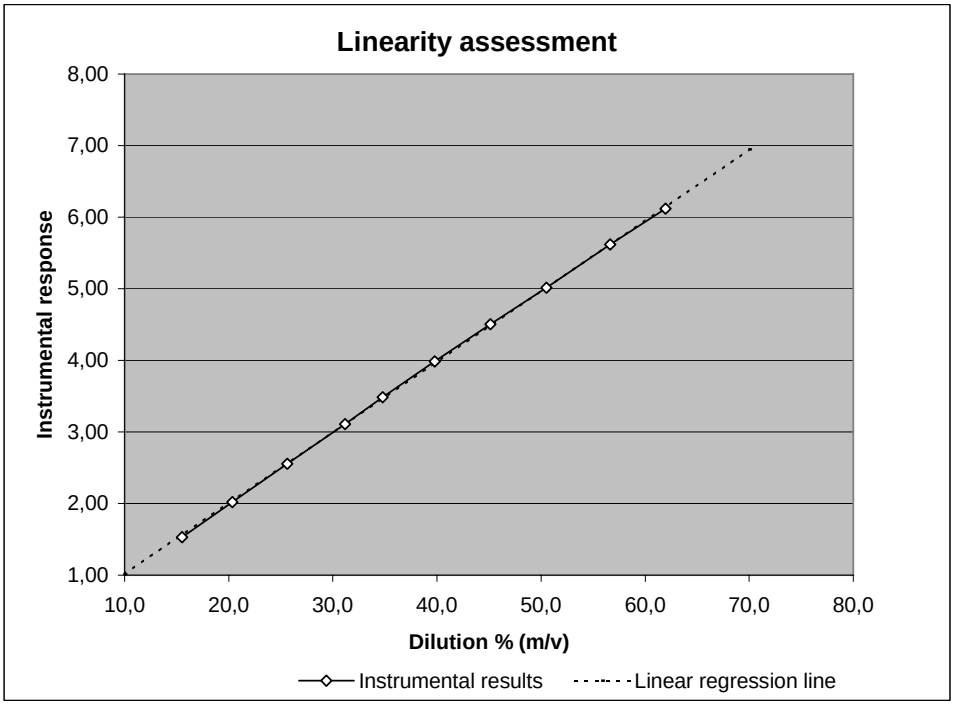
> F0,95 = 2,45 => Conclusion : Linearity default

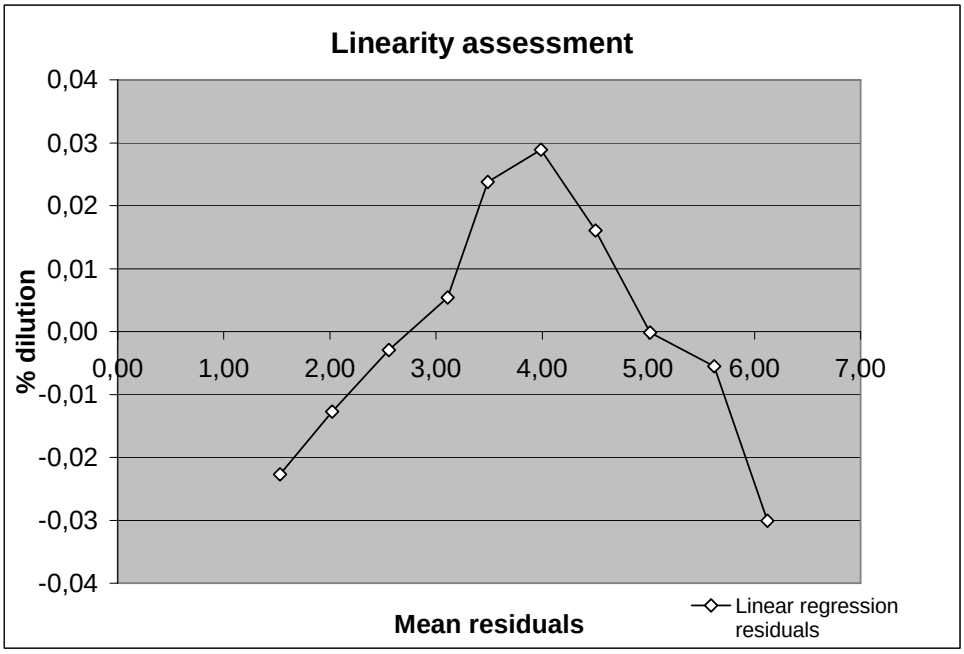
d- Compliance with polynome of 2nd and 3rd degree : Thanks to : Sy,x / Sy,xk < ((F1-a .(k-1) + (n-k-1)) / (n-2))1/2
with : n samples, k polynomial degrees,
k1 = k-1, k2 = q-k-1 and a risk of error.

Polynome	b3	b2	b1	a
degree 3	-0,000001	0,000014	0,102190	-0,056563
degree 2		-0,000087	0,105744	-0,093564
degree 1			0,098975	0,018563

Polynome	Sy,xk	d.f.	Sy,xk/Sy,x3	F0,95	Limit	Sy,xk/Sy,x2	F0,95	Limits
degree 3	0,010	26	1,00					
degree 2	0,010	27	1,01	4,23	1,21	1,00		
degree 1	0,020	28	2,07	3,37	1,26	2,05	4,21	1,18

Conclusions :
Both 2nd and 3rd degrees polynomial adjustement can improve linearity significantly according to the limits defined by F-tests => linearity default.
Nevertheless, a 2nd degree adjustment is sufficient as no significant improvement is noted between the 2nd and 3rd degree polynomial adjustment


Linearity assessment
8,00
7,00
6,00
5,00
4,00
3,00
2,00
1,00
Instrumental response
10,0
20,0
30,0
40,0
50,0
60,0
70,0
80,0
Dilution % (m/v)
Instrumental results
Linear regression line


Linearity assessment
0,04
0,03
0,02
0,01
0,00
-0,01
-0,02
-0,03
-0,04
% dilution
0,00
1,00
2,00
3,00
4,00
5,00
6,00
7,00
Mean residuals
Linear regression residuals

ANNEX - EXAMPLES

EVALUATION OF MILK ANALYSERS FOR ICAR APPROVAL

1. Assessment of preliminary instrumental fittings :

1.3. Assessment of linearity :

Sample set of progressive dilution of a high cell content milk by a low cell content milk

Test No	% dilution (m/v) X	Mean concent. Y	Residuals e regr. 1-21	Residuals e regr. 1-9	Ratio De/DC	Std. dev. prediction Sy,xi	t-test Student from line	
1	0,0	7,2	-25,2	-4,9		5,538	-1,006	
2	5,4	131,2	-18,2	-2,2	0,022	5,255	-0,464	
3	10,1	238,8	-12,4	-0,2	0,021	5,054	-0,039	
4	15,2	356,5	-5,1	3,0	0,023	4,884	0,645	
5	19,7	461,7	2,6	7,1	0,026	4,776	1,559	
6	24,4	564,0	3,1	3,8	0,022	4,704	0,849	
7	30,2	689,7	3,2	-0,7	0,018	4,675	-0,163	
8	35,0	800,2	9,7	2,0	0,015	4,699	0,433	
9	39,9	900,5	3,9	-7,8	0,017	4,770	-1,714	
10	44,9	1013,5	8,6	-7,1	0,015	4,889	-1,541	
11	49,6	1122,8	16,1	-3,4	0,013	5,045	-0,719	
12	55,3	1249,3	19,1	-4,9	0,012	5,292	-1,020	
13	59,8	1348,5	20,8	-6,8	0,011	5,530	-1,379	
14	64,5	1441,7	12,2	-19,1	0,018	5,821	-3,804	**<- upper limit**
15	69,9	1561,0	14,6	-21,1	0,018	6,208	-4,066	t0,975 = 2,365
16	74,6	1653,5	5,3	-34,2	0,025	6,590	-6,389	with P=5 %
17	79,4	1766,5	14,3	-29,0	0,023	7,024	-5,248	and 7 df
18	84,6	1865,2	0,4	-47,1	0,029	7,545	-8,226	
19	89,7	1983,8	8,5	-43,0	0,027	8,105	-7,253	
20	95,5	2074,8	-26,1	-82,3	0,043	8,804	-13,312	
21	100,0	2143,0	-55,4	-115,2	0,057	9,391	-18,038	
Level Numb	21	21	21	9		9	9	
Mean	49,89	1113,02	0,00	0,00				
Std. dev.	30,95	670,56	18,957	4,905				
Minimum	0,00	7,20	-55,39	-7,80	0,02	4,67	-1,71	
Maximum	100,00	2143,00	20,84	7,10	0,03	5,54	1,56	
D=Max-Min	100,00	2135,80	76,23	14,90	0,01	0,86	3,27	

1.4. Assessment of measurement limits : Example of a somatic cell counter (cf. IDF 148)

a- Lower limit : 10 measurements close to zero

Data	3	5	4	3	5
Data	4	5	3	5	4
Mean	4,100				
Std. Dev.	0,876				
CV%	21,4	< 30 % => conform			
DL	2,881	< 5000 => conform			
N	10				

b- Upper limit :

Regression :	**Slope b =**	22,4603	**Intercept a =**	12,1324

> From the figure, identification of the linear part ; calculation of the regression equation y = b.x+a on the linear part (level 1 to 9)
> on the whole range, calculation of :

* residuals : ei = yi -y(xi) = yi - b.xi - a
* t test on residuals : tobs = | ei | / Sy,x .(1/q + (xi -m(x))2 / SCEx)1/2

Conclusion : From level n°14, departure from linearity observed with tobs significant with P=0,95
N°14 corresponds to the increase of the residualrange/concentration range ratio test

1.3. Assessment of linearity : Example of a somatic cell counter (cf. IDF 148)

SD of residual means :	Se =	19,0	(measured)
SD of repeatability :	Sr =	16,4	(measured)
SD of level bias :	Sl =	16,4231	(calculated by Sl = (Se2-Sr2/n)1/2)

Linearity tests :

a- Ratio : (on the whole range i.e. 1 to 21)

De/DC =	0,036	> 0,02	**Conclusion :**	Linearity default

Note : This test is simple to apply - generally advised for quich checks in routine - nevertheless, due to the irregularity of residual scattering with SCC, it is important to confirm by a graph examination of residual plotting.

b- Bias from linearity test : Fobs=(Sr2 + n.Sl2) / Sr2 = n.Se2/Sr2 should be lower than F0,95 = 1,84

(with triplicates on 21 levels)

Fobs = 4,01 > F0,95 = 1,84 => **Conclusion :** Linearity default

Note : This test understands replicates are performed at every level and that the variance of residuals is uniform throughout the range which is rarely observed - therefore is not strictly exact - with SCC due to the very large scale (4 log paths : 10^3 to 10^6) It is more suitable for chemical analyses, nevertheless can be considered as sufficiently informative for SCC.

c- Compliance with polynome of 2nd and 3rd degree : =>Test : Sy,x / Sy,xk < ((F1-a .(k-1) + (n-k-1)) / (n-2))1/2
with : n samples, k polynomial degrees, k1 = k-1, k2 = q-k-1 and a risk of error.

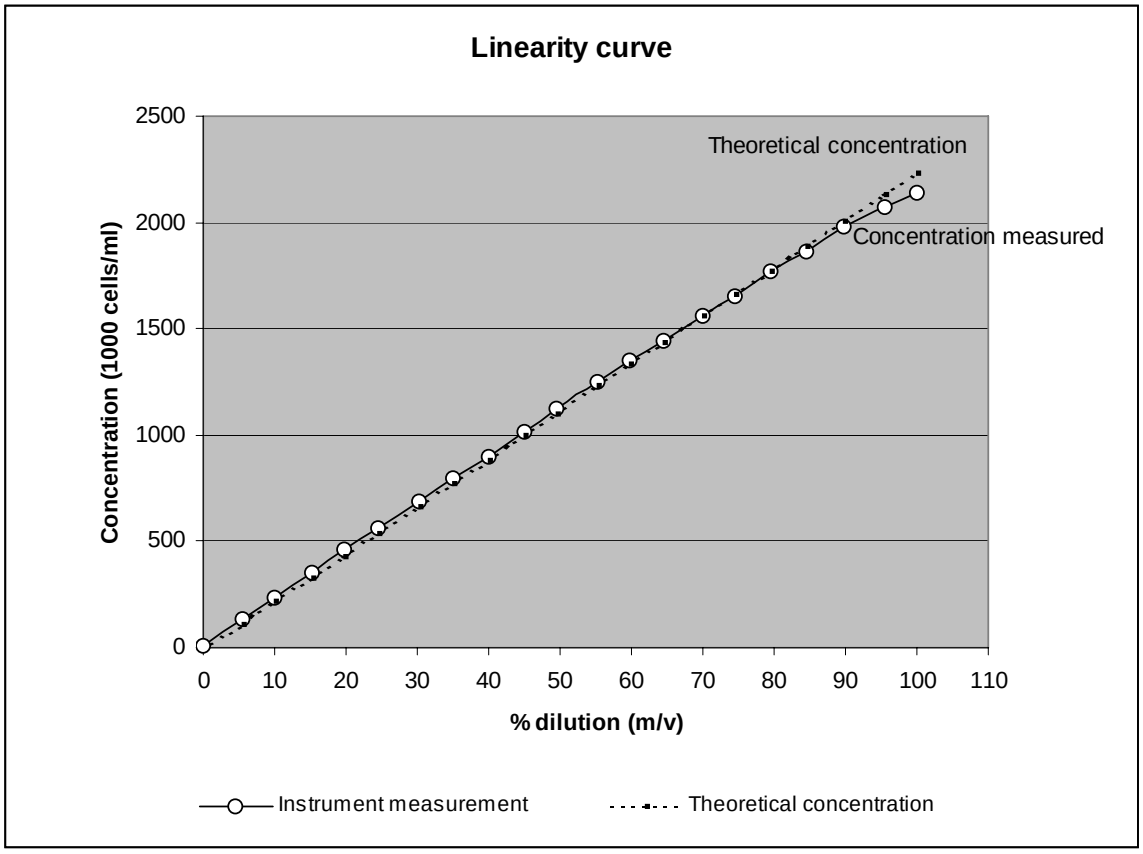
Mean concent. Y	% dilution (m/v) X	% dilution (m/v) X2	% dilution (m/v) X3	Residuals X2	Residuals X3
7,2	0,0	0,0	0	5,4	-5,9
131,2	5,4	29,2	157	2,6	-1,6
238,8	10,1	102,0	1030	0,7	1,1
356,5	15,2	231,0	3512	0,7	4,4
461,7	19,7	388,1	7645	2,8	8,3
564,0	24,4	595,4	14527	-1,8	4,7
689,7	30,2	912,0	27544	-6,8	-0,4
800,2	35,0	1225,0	42875	-3,5	2,0
900,5	39,9	1592,0	63521	-11,7	-7,6
1013,5	44,9	2016,0	90519	-8,4	-6,2
1122,8	49,6	2460,2	122024	-1,4	-1,2
1249,3	55,3	3058,1	169112	2,1	0,0
1348,5	59,8	3576,0	213847	5,2	1,4
1441,7	64,5	4160,3	268336	-1,3	-6,5
1561,0	69,9	4886,0	341532	4,6	-1,7
1653,5	74,6	5565,2	415161	-0,7	-7,2
1766,5	79,4	6304,4	500566	13,3	7,5
1865,2	84,6	7157,2	605496	5,8	1,8
1983,8	89,7	8046,1	721734	21,2	20,4
2074,8	95,5	9120,3	870984	-4,0	0,8
2143,0	100,0	10000,0	1000000	-25,0	-14,3
21	21	21	21	21	21
1113,02	49,89	3401,16	260958	0,00	0
670,56	30,95	3207,87	310802	9,14	7
7,20	0,00	0,00	0	-24,98	-14
2143,00	100,00	10000,00	1000000	21,20	20
2135,80	100,00	10000,00	1000000	46,18	35

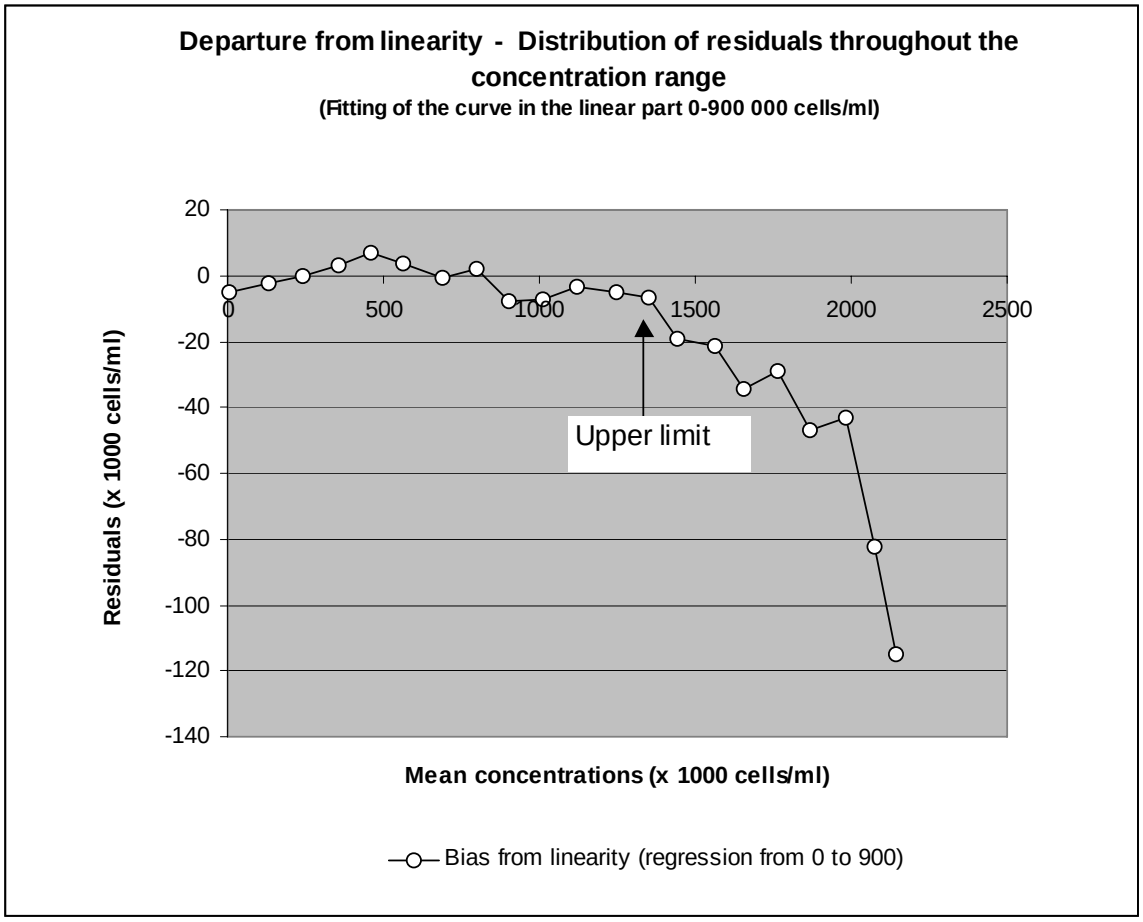
Polynome	b3	b2	b1	a
degree 3	-0,000256	0,019324	22,068420	13,063507
degree 2		-0,019194	23,580701	1,847156
degree 1			21,660009	32,390894

Conclusions : Significant improvement by 2nd and 3rd degree polynomes which confirm a linearity default

Polynome	Sy,xk	d.f.	Sy,xk/Sy,x3	F0,95	Limit	Sy,xk/Sy,x2	F0,95	Limit
degree 3	7,78	17	1,00			0,81		
degree 2	9,63	18	1,24	4,45	1,09	1,00		
degree 1	18,96	19	2,44	3,59	1,15	1,97	4,41	1,09

Note : This test can be run with all data (replicates) or only the mean values (the example) depending on the sensitivity needed. It normally requires the variance of residuals to be uniform throughout the range which is generally not achieved with cell counting. Nevertheless, provided with the residual plotting it can be considered as sufficiently informative for SCC linearity assessment.


Linearity curve
2500
2000
1500
1000
500
0
Concentration (1000 cells/ml)
Theoretical concentration
Concentration measured
0
10
20
30
40
50
60
70
80
90
100
110
% dilution (m/v)
Instrument measurement
Theoretical concentration


Departure from linearity - Distribution of residuals throughout the concentration range
(Fitting of the curve in the linear part 0-900 000 cells/ml)
20
0
-20
-40
-60
-80
-100
-120
-140
Residuals (x 1000 cells/ml)
0
500
1000
1500
2000
2500
Upper limit
Mean concentrations (x 1000 cells/ml)
Bias from linearity (regression from 0 to 900)

ANNEX - EXAMPLES

EVALUATION OF MILK ANALYSERS FOR ICAR APPROVAL

3. Assessment of overall accuracy :

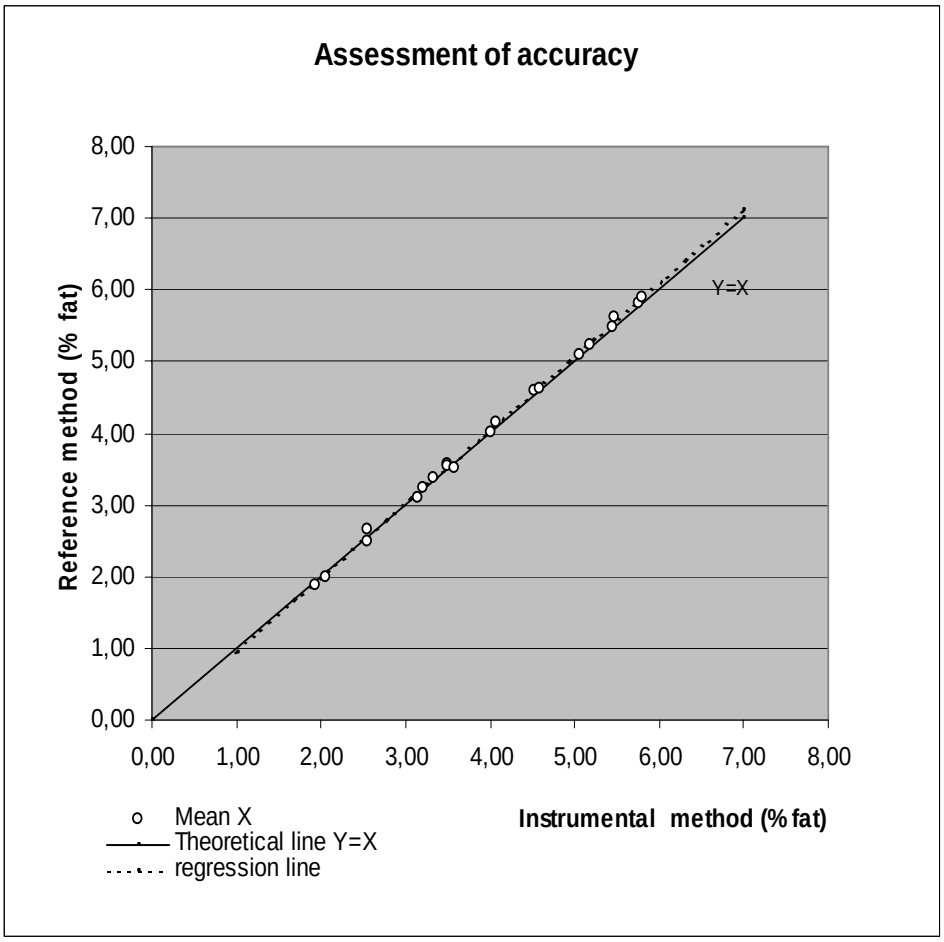
Example of fat analysed by infra red spectroscopy (cf. IDF 141)

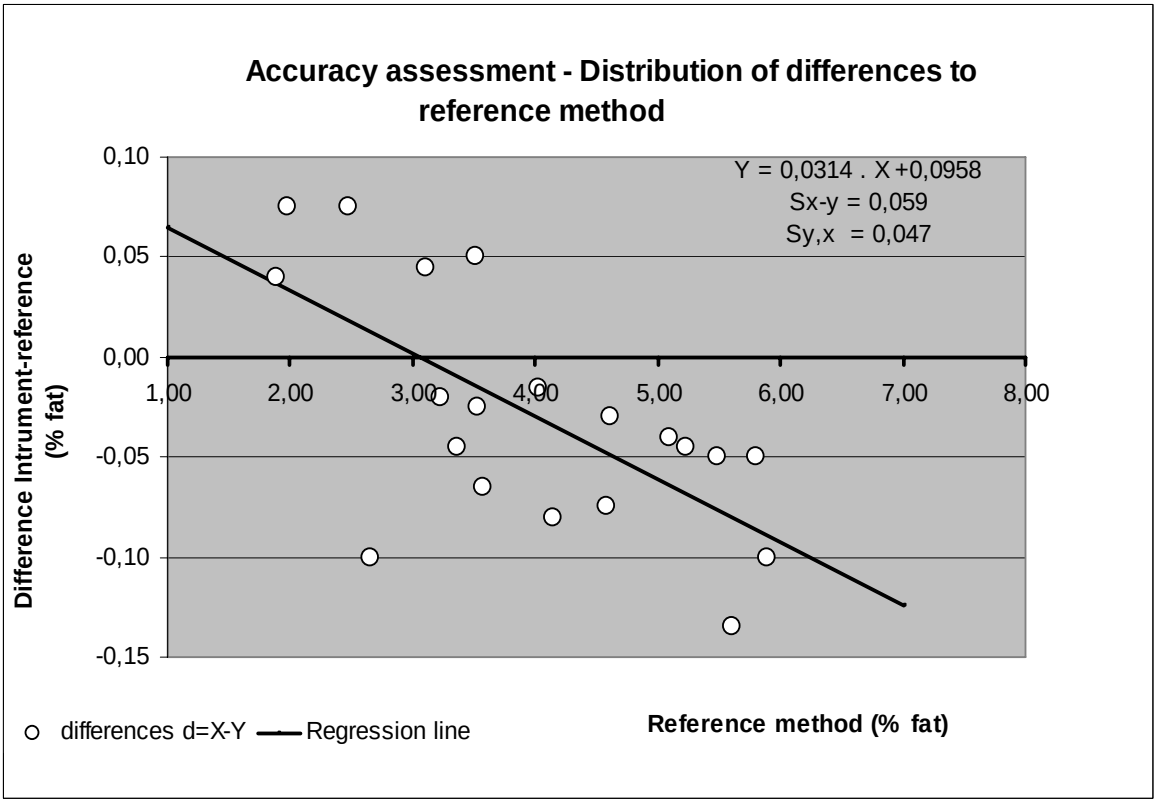
Set of individual cow milk samples

Test No	Reference method	Instrumental method				Repeatability	Accuracy	
		Replic. 1	Replic. 2	Mean	Estimates	Bias	differences	residual
	Y	X1	X2	X	Y/xi	w=|X1-X2|	d=X-Y	e
1	1,89	1,92	1,94	1,930	1,90	0,02	0,04	-0,006
2	1,98	2,05	2,06	2,055	2,03	0,01	0,07	-0,045
3	2,48	2,55	2,56	2,555	2,54	0,01	0,07	-0,061
4	2,66	2,56	2,56	2,560	2,55	0,00	-0,10	0,114
5	3,10	3,16	3,13	3,145	3,15	0,03	0,04	-0,049
6	3,23	3,20	3,22	3,210	3,22	0,02	-0,02	0,014
7	3,37	3,31	3,34	3,325	3,33	0,03	-0,04	0,035
8	3,57	3,51	3,50	3,505	3,52	0,01	-0,06	0,050
9	3,53	3,51	3,50	3,505	3,52	0,01	-0,02	0,010
10	3,52	3,57	3,57	3,570	3,59	0,00	0,05	-0,067
11	4,02	4,00	4,01	4,005	4,04	0,01	-0,01	-0,016
12	4,15	4,05	4,09	4,070	4,10	0,04	-0,08	0,047
13	4,59	4,52	4,51	4,515	4,56	0,01	-0,08	0,028
14	4,61	4,59	4,57	4,580	4,63	0,02	-0,03	-0,019
15	5,10	5,06	5,06	5,060	5,12	0,00	-0,04	-0,024
16	5,23	5,18	5,19	5,185	5,25	0,01	-0,04	-0,022
17	5,49	5,44	5,44	5,440	5,52	0,00	-0,05	-0,025
18	5,61	5,48	5,47	5,475	5,55	0,01	-0,14	0,058
19	5,80	5,74	5,76	5,750	5,84	0,02	-0,05	-0,035
20	5,89	5,80	5,78	5,790	5,88	0,02	-0,10	0,014
N	20	20	20	20	20	20	20	20
Mean	3,991	3,960	3,963	3,962	3,991	0,014	-0,030	0,000
SD	1,260	1,223	1,219	1,221	1,259	0,011	**0,059**	**0,047**
Minimum	1,890	1,920	1,940	1,930	1,896	0,00	-0,14	-0,07
Maximum	5,890	5,800	5,780	5,790	5,876	0,04	0,07	0,11
D=Max-Min	4,000	3,880	3,840	3,860	3,980	0,04	0,21	0,18

	Parameter	Estimate	Limits	Conformity
Repeatability :	**Sr**	0,012	0,014	Yes
Accuracy :	**Mean d**	-0,030	+/-0,050	Yes
	Sd (=Sx-y)	0,059	0,100	Y
	N	20		
	t obs	2,218	t0,975 =2,093	P< 0,05
	df	19		
Regression :	**Slope b**	1,0311	1+/-0,05	Yes
	Sb	0,0088		
	tobs b vs.1	3,511	t0,975 = 2,101	P<0,001
	Intercept a	-0,0935		
	Sa	0,037		
	tobs a vs.0	2,556	t0,975 = 2,101	
	df	18		
	Sy,x	0,047	0,100	Yes

Conclusion : Instrument accuracy complies with limits defined for the component analysed, in the example fat in cow milk.


Assessment of accuracy
Reference method (% fat)
8,00
7,00
6,00
5,00
4,00
3,00
2,00
1,00
0,00
0,00 1,00 2,00 3,00 4,00 5,00 6,00 7,00 8,00
Y=X
Instrumental method (% fat)
Mean X
Theoretical line Y=X
regression line


Accuracy assessment - Distribution of differences to reference method
Difference Intrument-reference (% fat)
0,10
0,05
0,00
-0,05
-0,10
-0,15
1,00 2,00 3,00 4,00 5,00 6,00 7,00 8,00
Y = 0,0314 . X +0,0958
Sx-y = 0,059
Sy,x = 0,047
differences d=X-Y
Regression line
Reference method (% fat)