Analytical precision performance in ICAR proficiency testing programmes

O. Leray

Actilait, BP 70129, 39800 Poligny, France

Abstract

The increased development of international trade in animal genetic has made worldwide comparability and equivalence recognition forf animal performance measurement a topical issue. This is especially the case for milk recording where milk analysis is a major issue.

To assure a harmonised analytical quality among member organisations, ICAR has built up since 1996 a international reference system based on an international network of dairy reference laboratories. Beside a number of recommendation set in guidelines, ICAR organises annually international proficiency testing schemes to help laboratory members of the network to evaluate their analytical performance and, through continuous improving, upgrade the overall analytical precision in the ICAR world and lead to form a consistent group of expert laboratories capable to establish trusty reference values.

Proficiency testing studies are organised twice a year for cow milk and involve reference methods for fat, protein, any methods lactose, urea and somatic cell counting.

A first review of laboratory performances and precision figures for fat, protein and somatic cell counting between 1996 and 2003 were presented in ICAR Session 2004 (Sousse, Tunisia) and showed unsatisfactory precision performances for the group as not conforming to IDF-ISO standard precision values, with inequal individual laboratory performances.

The compared review for the six last year (until 2009) illustrates a significant improvement with precision figures of the group converging onto or below standard precision values. The number of regular good performing laboratories appears significantly increasing compared to the picture made in 2004.

Nevertheless particular care must maintain and especially be given to somatic cell counting where last trials show a trend to higher precision values.

This work undertaken within ICAR serves as a basis for further undergoing development in the joint IDF-ICAR project "Reference system for somatic cell counting" where qualifying and selecting expert laboratories has become a major issue.

Keywords: milk analysis, dairy laboratories, laboratory network, proficiency.

1.0 Introduction

For the two last decades, the increased development of international trade in animal genetic has made worldwide comparability and equivalence of animal performance measurement a topical issue. This is especially the case for milk recording, being for genetic trade (animals, semen, embryo) or the international evaluation of animal genetic index (Interbull).

To cope with that issue, ICAR set up in 1996 an international network of dairy reference laboratories so as to implement progressively a harmonised international quality assurance system for milk recording analysis worldwide. The reference laboratories are expected to acquire a high expertise in the analytical methods, either standardised or validated, used in milk recording so that they can provide routine laboratories with good reference values through adequate monitoring, anchoring (reference material) and good analytical practices. Harmonising analytical performances within the network is the first major step.

From 1996, ICAR has organised twice a year international proficiency studies for the benefit of the laboratory network members. Those studies were carried out mainly for cow milk and the reference methods for fat, protein, lactose, urea and routine methods for somatic cell counting.

A first review of laboratory performances and the precision figures shown for the methods by the group of laboratories was made in ICAR Session 2004 in Sousse (Tunisia) covering the period from 1996 to 2003 (13 trials). It was the occasion to illustrate how collaborative trials like proficiency studies could be used to strengthen analytical system through measuring analytical performance quality.

Six years later a new review appears necessary to update the knowledge on laboratory performance precision. Indeed, from 2004 and the experience acquired in ICAR, the concept of reference system, including the various uses of laboratory networks, was developed and evolved into the joint IDF-ICAR project "Reference system for somatic cell counting". Qualifying and selecting expert laboratories to provide suitable reference values for somatic cell counting has become a major issue whereas need is still to evaluate the current state of the Art for the main milk components, fat and protein.

2.0 Material and methods

2.1 Protocol of the PT scheme organisation

At every end of year the programme of ICAR proficiency testing scheme for the forthcoming year is addressed to the members of the ICAR Reference Laboratory Network and national member organisations of ICAR. The yearly scheme is organised in two rounds, the first in March and the second in September and is applied on cow milk for the component of interest for milk recording and dairy herd management. They are fat and protein by the relevant reference methods, somatic cell counting, lactose and urea by any validated methods excluding infrared. Indeed infrared is marked by significant interferences related to milk composition that makes so-obtained results irrelevant to assess lab performance quality.

Only results for fat, protein and somatic cell counting are reported here as main components used for the genetic evaluation.

2.1.1 Samples

Sets of samples used are made of 10 samples preserved with bronopol at a concentration of 0.02% in milk, covering evenly the range of concentration usually met in routine testing that is

10 whole milk samples regularly ranging from 1.5 % to 4.9 % fat.

10 whole milk samples regularly ranging from 2.5 % to 4.0 % crude protein.

10 whole milk samples regularly ranging from 4.6 % to 5.1 % lactose.

10 whole milk samples regularly ranging from 50 to 1600 $x10^3$ cells/ml.

10 whole milk samples regularly ranging from 10 to 70 mg urea /100 ml.

Sample containers are 65 ml or 35 ml polyethylene screw-capped vials with airtight joints to prevent breaking and leakage, and sample temperature before and during shipment to laboratories is +4°C. Possible storage prior analysis is required to be + 4°C whereas analysis is to be performed within 5 days for somatic cell counting and 10 days for chemical analysis after the dispatch date.

2.1.2 Milk testing, statistical analysis and assessment parameters

Milk testing is required to be performed in duplicate and according to the current version of the relevant international standard. Cautions for sample preparation before analysis are remined in an advisory technical note appended to samples. The order of analysis is to be better that one indicated in the numbering in order to avoid errors and reporting is made through adequate tables.

Statistical analysis is performed according to the model developed by the Institut de l'Elevage then used by Cecalait as described in the IDF Bulletin n°342:1999, annexe 3. Assigned values used as reference are calculated according to ISO 13528.

Each sample corresponds to a different concentration level. Assessment is made through dedicated tables allowing the evaluation of lab performance in lines and group performance in columns, for repeatability (ranges of duplicates and standard deviation of labs or samples, accuracy (means of duplicates, assigned reference per level, differences to assigned reference values, lab scores made of the mean, d, and the standard deviation of the differences, sd. A synthesis table with lab ranking according to the Euclidian

distance D (as the geometric mean of the two latter parameters, $D = (d^2+sd^2)$) provides indications of the range of analytical performances and a figures illustrates the related location of each lab vs an indicative conformity target (Figure 1).

For somatic cell counting an table for calibration equation estimates is given as an additional information.

2.2 Meta-analysis for a global evaluation

2.2.1 Meta-analysis and group performance evaluation

Meta-analysis consists in gathering individual lab performances in a single large table per criterion - repeatability, mean of bias, standard deviation of difference, distance D – and to present such results in figures in a form of control chart with values in ordinate and trial number in abscissis.

This is done at first with raw data as including all the methods and abnormal scores for a first visual scrutiny and evaluation of the evolution of performance throughout time, then after discarding not expected methods (e.g. Gerber, infrared methods) and outlier laboratories.

If the source of outliers are evident and reflects only a basic error of unit or disorder than can be repaired, correction and recalculation of scores are made since they would be detected in lab situation when calibrating routine analysers. Otherwise outlier are detected then discarded through a Cochran test applied on the distance D with a risk of error of 1% and in the limit of 20% max which was attained or passed but occasionally.

As well the geometric averaging per trials of all the participant values serve to evaluate the precision figures for each trial and the similar integration of the individual trial precision figures for a defined period of time allow to measure the overall precision improvement of the group of laboratories (Table 1).

2.2.2 Meta-analysis and lab performance evaluation

Individual lab control chart can be built up to assess each lab over a period of time and the calculation of an average score for the defined period (for instance the four last trials) becomes possible for a defined period of time to apply a suitable selection of laboratories for a defined purpose according to their performances (Table 2). Additional selection criterion can be the frequency of each lab participation.

Similarly as well as shown in Sousse (2004), such individual score merging allow to calculate a robust true individual uncertainty for the measurement applied for a given representative period.

3.0 Results

3.1 Overall scrutiny of individual scores

Compared to the period of 1996 to 2003, the control charts have shown in general a lower frequency of outliers with lower upper values for the period of 2004 to 2009, can this be for repeatability standard deviation, mean bias, standard deviation of bias or Euclidian distance D.

The outlier discarding suppressed less scores and rarely reached the percentage limit for deletion of 20%.

3.2 Repeatability and reproducibility figures (precision)

3.2.1 Fat measurement

Repeatability standard deviation values, sr, were higher than 0.10 g/kg from 1996 to 1999 then decreased to keep almost stable from 2000 to 2006 just above the standard limit of 0.07 g/kg of IDF 1 / ISO 1211. From 2007, in conjunction with the implementation of ICAR Quality Certificate, the values dropped below the limit. Referring to new standard value 0.15 g/kg implemented with the recent revision of IDF 1 / ISO 1211 the group shows good compliance from 1996.

A similar trend is observed for reproducibility standard deviation, sR, although the group never passed through the standard reproducibility limit of 0. 14 g/kg. Referring to new standard value 0. 20 g/ kg

implemented with the recent revision of IDF 1 / ISO 1211 the group shows however frequent compliance from 2000 (Figure 2).

3.2.2 Protein measurement

From 1996 to 1999 repeatability standard deviation values are mostly lower than the standard limit of 0.14 g/kg of IDF 20 / ISO 8968 with only two trials outside. However stabilization around 0.10 g/kg from 2006 is observed.

Reproducibility standard deviation was significantly higher than the standard limit of 0.18 g/kg of IDF 20 / ISO 8968 but from 2003 have decreased to a regular fitting onto the limit (Figure 3).

3.2.3 Somatic cell counting

From 1996 to 1999 repeat ability standard deviation va lues a re mostly s ignificantly lo wer than the standard limit of 20 000 cells/ml of IDF 148-2 / ISO 13366-2 with only one trial outside in 2000. However after a optimal performance period between 2005 and 2007 irregular discrepancy is observed.

Besides, whereas reproducibility standard deviation was at the level or higher than the standard limit of 45 000 cells/ml of IDF 148-2 / ISO 13366-2 until 2003, they ha ve been reduced significantly around a level of 30 000 cells/ml from 2004 to 2008. Since a deterioration of the gl obal performance is observed and the limit is passed through in 2009 (Figure 4).

3.3 Individual lab performances and reference lab group selection

Individual performances whatever the component - fat, protein or somatic cells – have shown significant progress as illustrate through the improvement of the overall precision of the group of participants.

Selection of group of reference lab for a defined purpose such as assigning reference values for reference material should be made from the more recent lab performances measured hence to define the last necessary period. Then to rank laboratories according to the overall precision shown by a significant statistical parameter, for instance sRL which covers all the sources of errors of the laboratory.

Table shows a case example dealing with fat for the four last trials (2008-2009) and yellow highlighting indicates prior defined limits are passed. On such a basis, as an example, one could assumed to retain 19 labs of 21 for fat, 17 of 20 for protein and 12 of 18 in somatic cell counting.

Nevertheless not all participated at the same time in the same trials and the observed participation frequency (re Total% in Table 2) indicates regularity of the feedback information and possible corrective actions. So ranking according to the frequency can be associated to the ranking on scores thus introduced a weighting.

4. Conclusion

Findings of the first review and their presentation in Sousse 2004 permitted to inform laboratory network for the need to improve testing practices and invited them to review own ways of work. This has resulted in a effective improvement at individual lab levels and consequently at the level of the whole group of participants. The efficiency of the ICAR reference system is demonstrated there.

Year 2006 saw the i mplementation of the ICA R Qual ity Certifi cate to repl ace the S pecial Stamps and broaden the scope of ICAR quality assurance system to all the parts of its expertise and among them milk analysis. Correlatively tighter regularity and compliance took place for fat a nd protein, at a l ower extent for somatic cell count, demonstrating the efficiency of ICAR quality policy.

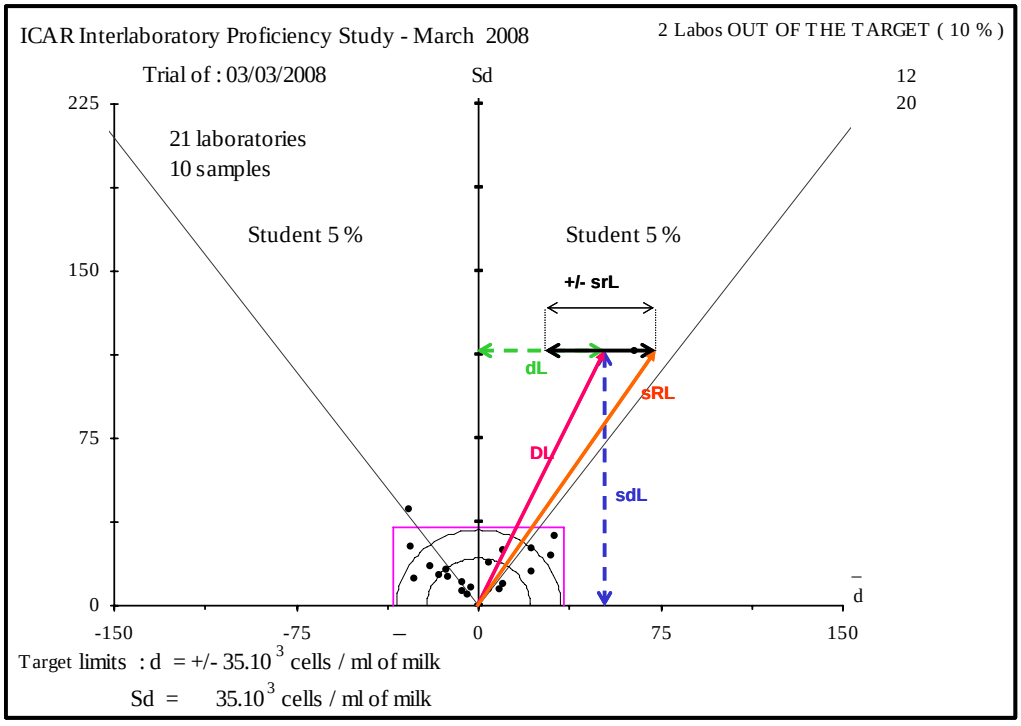


Figure 1. Statistical parameters for individual lab performance evaluation - Example of ICAR trial of March 2008

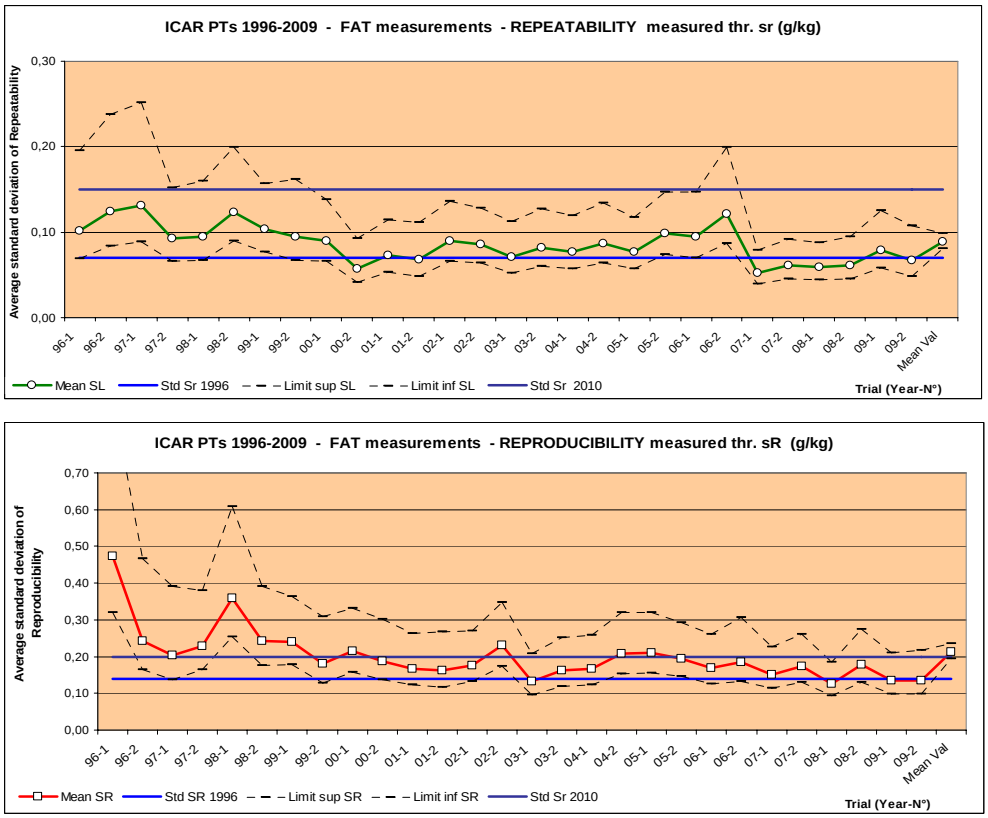


Figure 2. Precision figures in ICAR trials in fat analysis between 1996 and 2009

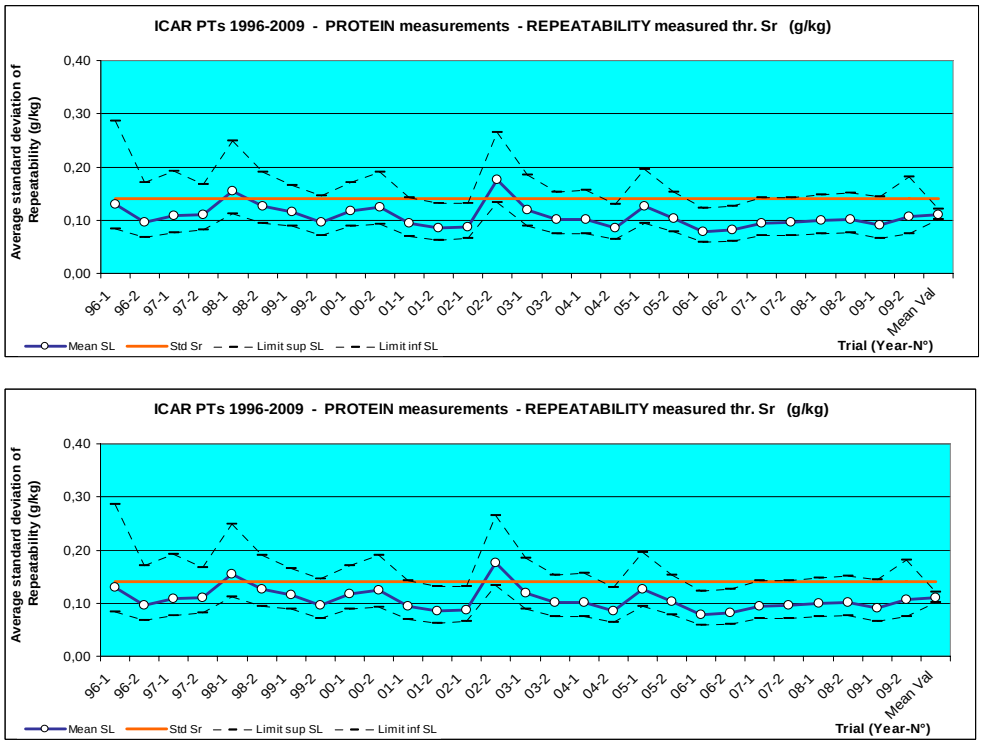
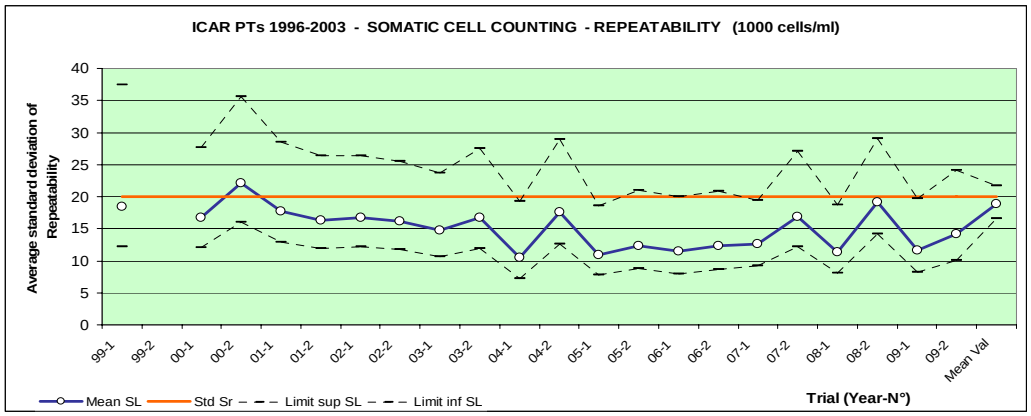


Figure 3. Precision figures in ICAR trials in protein analysis between 1996 and 2009



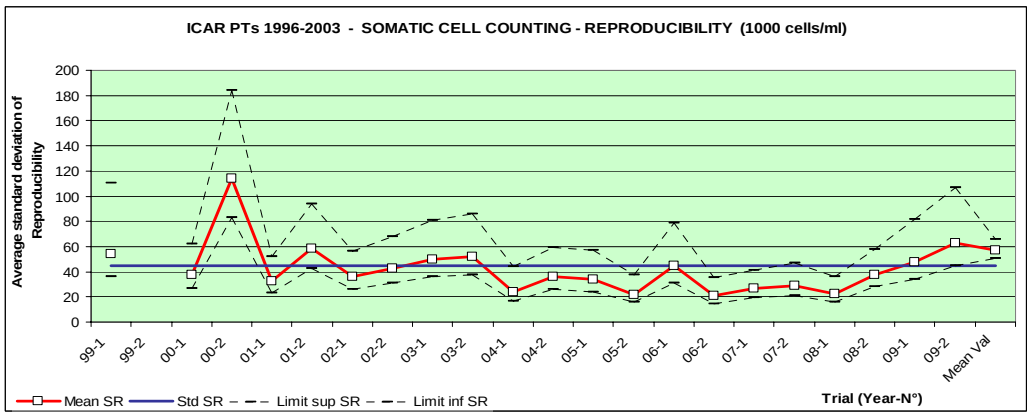


Figure 4. Precision figures in ICAR trials in somatic cell counting between 1996 and 2009

Table 1. Robust estimates of precision and accuracy parameters in fat by meta-analysis for the group of participant in ICAR trials

Fat (g/kg)

N° ICAR	1996 - 2009	1996 - 2003	2004 - 2009	2008 - 2009	Limits	Standard
Mean SL	0,09	0,09	0,08	0,07	0,10	0,07
Mean d	-0,01	-0,01	-0,01	0,00	0,20	
Mean Sd	0,16	0,18	0,13	0,13	0,30	
Mean D	0,20	0,23	0,18	0,18	0,36	
Mean SR	0,21	0,24	0,20	0,19	0,37	0,14
Number N	387	203	184	60		

Table 2. Example of laboratory ranking according to reproducibility (srL) performances in ICAR trials (last four trials in fat)

Fat : Ranking to select lab candidate pool to assign reference value (4 last trials)

RANK	Lab ID	TOTAL %	Mean SL	Mean d	S mean d	Mean Sd	Mean D	Mean SRL
1		100%	0,04	0,02	0,09	0,05	0,06	0,07
2		75%	0,04	0,05	0,06	0,04	0,06	0,07
3		75%	0,04	0,02	0,08	0,06	0,07	0,07
4		100%	0,04	0,04	0,07	0,05	0,07	0,07
5		50%	0,07	0,01	0,12	0,06	0,07	0,08
6		50%	0,04	0,04	0,05	0,07	0,10	0,10
7		100%	0,07	-0,03	0,06	0,07	0,09	0,10
8		100%	0,07	-0,01	0,06	0,06	0,09	0,10
9		100%	0,05	0,07	0,31	0,08	0,12	0,12
10		50%	0,04	0,11	0,05	0,08	0,14	0,14
11		100%	0,07	-0,11	0,18	0,08	0,15	0,15
12		100%	0,06	-0,05	0,61	0,14	0,16	0,17
13		100%	0,06	0,01	0,18	0,09	0,17	0,18
14		100%	0,12	0,01	0,19	0,16	0,17	0,19
15		25%	0,16	0,18	0,14	0,06	0,19	0,22
16		50%	0,07	-0,09	0,08	0,21	0,23	0,23
17		100%	0,07	-0,16	0,10	0,16	0,24	0,24
18		75%	0,09	0,01	0,18	0,28	0,35	0,36
19		50%	0,11	0,04	0,21	0,35	0,36	0,37
20		100%	0,05	-0,13	0,24	0,12	0,38	0,38
21		25%	2,28	-1,16	0,78	1,81	2,15	2,69