International Dairy Journal 97 (2019) 185–190

Contents lists available at ScienceDirect

International Dairy Journal

journal homepage: www.elsevier.com/locate/idairyj

Official control and self-monitoring: Data agreement report in the integrated food safety system of an Italian dairy chain

Francesco Martelli [a, *], Claudia Giacomozzi [a], Roberto Dragone [b], Carlo Boselli [c], Simonetta Amatiste [c], Giovanni Brajon [c], Gerardo Grasso [b], Alberto Mantovani [d], Chiara Frazzoli [a]

[a] Istituto Superiore di Sanità, Department of Cardiovascular, Dysmetabolic and Aging-associated Diseases, Viale Regina Elena 299, 00161, Rome, Italy
[b] National Research Council - Institute of Nanostructured Materials (CNR-ISMN), c/o Università di Roma "La Sapienza", P.le Aldo Moro, 5, 00185, Rome, Italy
[c] Istituto Zooprofilattico Sperimentale del Lazio e della Toscana "M. Aleandri", via Appia Nuova, 1411, 00178, Rome, Italy
[d] Istituto Superiore di Sanità, Department of Food Safety, Nutrition and Veterinary Public Health, Viale Regina Elena 299, 00161, Rome, Italy

ARTICLE INFO

Article history:
Received 26 February 2019
Received in revised form
17 May 2019
Accepted 22 May 2019
Available online 4 June 2019

ABSTRACT

The dairy industry's silos is a critical point in the safety and quality control system. However, limited scientific evidence is available on measurement agreement between the milk analyses done by official control bodies and the self-monitoring analyses done by milk processing industries. Milk production data from a milk processing plant were collected for four months and analyzed by an official control body and the dairy company for freezing point, total bacterial count, somatic cell count, and for fat, lactose and protein percentages. Correlation and Bland-Altman analysis showed a good agreement between the two determinations for most of the variables (Spearman's rho > 0.82 for Somatic cell count, Fat% and Protein %), while low agreement was found for total bacterial counts (Spearman's rho =0.78). It was found that the difference between total bacterial counts was influenced by collecting route, time between sampling and analysis, and milk temperature inside the truck tank.

1. Introduction

Bovine milk is a relevant component in the diet of many industrialised and developing countries (Muehlhoff, Bennett, & McMahon, 2013). The nutritional value and the physicochemical properties of raw milk are directly dependent on milk composition which, in turn, directly affects the economic value of production (Dong, Hennessy, & Jensen, 2012). Moreover, milk composition gives valuable information on herd nutritional status and animals' general health (Auldist, Coats, Rogers, & McDowell, 1995; Hamann & Krömker, 1997). Daily measurement of milk components, both at the individual and herd level, is becoming a common tool to assess the safety and economic value of milk production.

Following EU Regulation 178/2002 (EU, 2002), Italian milk processing companies follow a strict self-monitoring process as per hazard analysis and critical control points (HACCP), and are subject to official control by the Italian Competent Authority (Ministry of Health, MoH). Planning, coordination and control activities are devolved to regional authorities, which in turn rely on local health units for operational implementation. In addition, a network of 10 Experimental Zooprophylaxis Institutes (Istituti Zooprofilattici Sperimentali; IZSs), supervised by the Italian Regions and the MoH, provide laboratory services for official analyses of raw milk samples. IZS labs are all accredited for such tests, and some of them act as National Reference Centres for certain milk parameters.

Internal testing methods are usually designed for a balanced compromise between ease of execution, accuracy, and speed. For these reasons, those methods may have different limits of detection and quantification, and different levels of measurement uncertainty. In certain cases, measurement principles may be different, according to the specific features of each process. To date, limited scientific evidence is available regarding measurement agreement between milk analyses carried on by official control bodies and self-monitoring analyses done by milk processors.

A key difference between the two kinds of methods lays in the measurements procedures; while IZSs are national reference laboratories with validated ISO/IEC 17025:2005 accredited procedures

* Corresponding author. Tel.: +39 6 49903271.
E-mail address: francesco.martelli@iss.it (F. Martelli).

https://doi.org/10.1016/j.idairyj.2019.05.006

and measurement protocols, HACCP routines are internally validated, and the two measurement procedures may make use of different instrumentation.

A second difference between the two measurement procedures is that, while the industrial monitoring routines are carried out immediately after the raw milk arrival at the processing premises, in official monitoring procedures the milk samples have to be dispatched to the accredited laboratory. While this procedure is validated and refrigerated transportation is routinely used, effects on raw milk sample biological parameters, especially during the hot season, are plausible but yet to be completely characterised. Even if time between sampling and analysis and milk temperature are well known promoting factors for bacterial growth, there is less evidence on their influence on bacterial growth or other milk parameters when combined with other managerial or environmental factors.

To address such topics, this study presents a comparison that was carried out between the self-monitoring measurements at a dairy plant in the Rome area (Centrale del Latte di Roma SpA (CLR)) and the official control measurements performed by the Istituto Zooprofilattico Sperimentale del Lazio e della Toscana "M. Aleandri" (IZS). The main goal of the study was to verify the agreement between the official control and self-control measurements at IZS and CLR, respectively, interpreting possible differences or biases and identifying the combination of factors that mostly influenced measurement discordances.

2. Materials and methods

2.1. Milk sample collection

Milk production and component measurement data were obtained during a monitoring program involving the milk Industry Centrale del Latte di Roma SpA, a FSSC 22000:2010 – certified milk processing plant with an average 350,000 L day^{-1} raw milk throughput, and the Istituto Zooprofilattico Sperimentale del Lazio e della Toscana "M. Aleandri", the IZS with territorial competence on the Lazio Region, where CLR runs its business. Data collection was performed between March, 2014 and June, 2014 within the framework of the Alert (http://www.alert2015.it/) project funded by the "Industria 2015″ program of the Italian Ministry of Economic Development and addressing the monitoring of wholesomeness and quality in the bovine milk chain from primary (dairy farm) to secondary production (manufacturing industry). Data collection regarded the raw milk collection was continued on by CLR on a regional basis (40% of the daily collection), consisting of a daily average of 137,000 kg raw milk collected from local farmers using a fleet of 15-metric ton refrigerated trucks. Traceability of bulk milk was limited to groups of individual farmers, identified by the collecting route (CR). Transportation was traced as per the internal CLR procedures. Milk transportation temperature was monitored and checked on arrival of the trucks to the CLR premises (average 6.4 °C ± 1.6 °C). Milk samples were collected on truck arrival, before loading milk into CLR silos, dipping the milk directly from the truck's tank into a separate container and splitting two milk samples from the container itself into plastic vials.

All samples were taken following rules matching both CLR internal procedures and IZS procedures. A total of 114 paired samples were collected, mainly in the first days of the week (12.3% on Sundays, 34.2% on Mondays, 50% on Tuesdays and 3.5% on Thursdays).

2.2. Sample analysis

The first sample was immediately sent to CLR internal laboratory and analysed within 5 h for freezing point (FP), total bacterial count (TBC), somatic cell count (SCC), fat (Fat), lactose (Lactose) and protein (Protein) percentages. The second sample was kept refrigerated and sent to the IZS labs the same day of collection, except for the Sunday samples, which departed from CLR on Mondays. Milk samples arriving at the IZS labs were analysed with accredited methods as per EN ISO/IEC 17025:2005 (Accreditation Certificate 0201).

Most IZS samples were analysed within 2 days from the sampling (69 samples, 60.5%, within 1 day and 36 samples, 31.6%, within the second day). A minor (9 samples, 7.9%) fraction was analysed within 4 days from collection (7 samples, 6.1%, within 3 days, 2 samples, 1.7%, within 4 days).

The resulting dataset (114 twin samples, 6 twinned variables) consisted of 1323 observations, as a total of 45 (3.3%) observations were incomplete due to technical issues and/or data capture errors (4 IZS somatic cell count observations, 17 CLR somatic cell count observations, 24 CLR total bacterial count observations).

2.3. Statistical and data analysis

A preliminary statistical analysis was performed to check normality of each parameter, for heteroscedasticity and normality in measurement differences. An outlier removal procedure (Tukey's interquartile for outer fences) was applied to remove extreme outliers. Each test was carried on with completely paired data sets. When a measurement was not available in both the paired sets, it was discarded. Descriptive statistics were reported for each parameter, and groups comparison was performed using the paired sample Wilcoxon Signed Rank test and paired sample t-test, as appropriate.

Correlations (Spearman's rank correlation) between each paired variables was assessed. A Bland–Altman plot (Bland & Altman, 1999) was used to assess measurement agreement between CLR and IZS determinations. Normality of measurement differences was also checked using the Shapiro-Francia test before computing mean confidence intervals for Bland–Altman analysis (Giavarina, 2015). The absence of heteroscedasticity in measurement differences against their means (Atkinson & Nevill, 1998) was assessed by computing their Pearson's coefficient of correlation.

A Linear model ($Y = TIME + CLRTBC + WT + MT + CR + err$) was used to model the effect of the time between the milk sampling at the CLR premises (at truck arrival) and IZS analysis (TIME, in days), CLR total bacterial count (CLRTBC), weather temperature (WT), milk temperature at truck arrival (MT), and milk collecting route (CR) on the logarithmic difference between CLR and IZS total bacterial count (Y). The model was built on a subset of data (54 twin samples) that had a confirmed complete traceability from the IZS analysis master record to the truck's license plate. Independent variables (predictors) were chosen between those that could realistically be involved in bacterial growth and were standardised. An ANOVA test was performed to analyse the contribution of the coefficients to the model.

For all statistical analyses, the significance threshold was set at 0.05. All calculations were made using Matlab (The MathWorks, Inc., Natick, MA, US) and the R environment (R Core Team, 2017).

3. Results

3.1. Preliminary statistical analysis

The Shapiro–Wilk normality test reported a significant result (non-normal distribution) for almost all variables, except for Lactose (CLR), Protein (CLR) and Protein (IZS).

All the measurement differences showed a leptokurtic distribution and the Shapiro-Francia test showed all non-normal distributions. For this reason, the Bland–Altman limits of agreement were estimated using a nonparametric method, usually entailing wider limits than those estimated by parametric methods. The correlation between measurement differences and their means was significant at least for SCC, Fat and TBC (Table 1).

Preliminary data inspection suggested an outlier identification and removal procedure. Because of the lack of normality, the Tukey's interquartile method for outer fences (Tukey, 1977) was used, i.e., a measurement was considered an outlier if external to the interval [Q1-3·IQR Q3+3·IQR] (Q1 first quartile, Q3 third quartile, IQR interquartile range). For most parameters (Somatic cell count, Freezing Point, Lactose, Protein) no outliers were identified. For Fat, one single outlier was identified both in IZS and CLR measurements. For total bacterial count, 12 outliers were identified in IZS measurements. All outliers in IZS measurements and their corresponding twins in CLR measurements were removed. Final database consistency after outlier removal is reported in Table 1. Heteroscedasticity and normality of variables and their differences were assessed again after the outlier removal procedure, which entailed no change at all in the statistical tests results. Descriptive statistics of the final distributions are reported in Table 2 for each parameter.

3.2. Comparison of measurements

All CLR and IZS corresponding parameters were compared using paired tests. Protein, normally distributed both in CLR and IZS dataset, were compared through the Student's paired t-test and showed no statistically significant difference ($t\ (113) = 1.61$; $p = 0.11$). SCC, FP, Fat, Lactose and TBC were compared through the Wilcoxon signed ranks test, and FP, Lactose, SCC and TBC showed a significant difference in the median of the paired samples. Specifically, CLR measurements were numerically higher for SCC and Lactose content, while IZS measurements exceeded those of CLR for FP and TBC. Fat measurements did not show any statistically significant difference in their medians.

Correlation analysis (Table 3) showed a good correlation for SCC, Protein and Fat, a moderate correlation for total bacterial count and Lactose, and poor correlation for Freezing point ($p < 0.001$).

For each parameter, a logarithmic Bland–Altman analysis was performed and reported as a scatter plot of the differences of log measurements (y axis) versus the mean of the measurement differences (x-axis) (Fig. 1). Due to the non-normality of the data, the limits of agreement (LOAs) in the plots were set at ±1.5 IQR. LOAs were also represented in Table 3. Specifically, good agreement (relative difference of IZS and CLR measurement less than 5%) was found as for FP, Fat, Lactose and Protein, a moderate to good agreement for SCC, and poor agreement for TBC. Presence of a clear bias was detected in SCC and TBC plots (Fig. 1A,C): the value at the intercept (Y (0), bold horizontal line) represents the geometric

Table 3
Correlation table of the twinned variables, and their expected range of agreement.[a]

Parameter	Spearman's rho	p	Bland and Altman's LOA (%)	
			Upper	Lower
Somatic cell count	0.88	<0.001	113.8	80.3
Freezing point	0.30	0.002	102.3	98.9
Fat	0.83	<0.001	103.9	97.2
Lactose	0.66	<0.001	100.4	97.9
Protein	0.90	<0.001	102.8	96.6
Total bacterial count	0.78	<0.001	723.7	74.2

[a] LOA, limits of agreement.

Table 1
Database characteristics: number of coupled measurements before and after outlier removal, correlation and normality of differences.[a]

Parameter	Original measurements	Number of outliers		Final measurements	Difference correlation		Normality of differences	
		CLR data	IZS data		Pearson's r	p	W (Shapiro − Francia)	p
Somatic cell count	93	0	0	93	0.31	0.003[a]	0.85	<0.001[a]
Freezing point	110	0	0	110	0.15	0.129	0.97	0.012[a]
Fat	114	0	1	113	−0.19	0.046[a]	0.80	<0.001[a]
Lactose	114	0	0	114	0.11	0.245	0.94	<0.001[a]
Protein	114	0	0	114	−0.12	0.219	0.90	<0.001[a]
Total bacterial count	90	1	12	78	0.99	<0.001[a]	0.49	<0.001[a]

[a] Difference correlation is the correlation between the absolute differences against their means; the superscript letter a marks significance of correlation ($p < 0.05$).

Table 2
Descriptive statistics of the dataset.[a]

Parameter	Analyst	N	Median	IQR	Mean	SD	Shapiro–Wilk normality test	
							W	p
Somatic cell count (counts mL^{-1} 1000)	CLR	93	320	141.75			0.971	0.036
	IZS	93	289	126			0.950	0.001
Freezing point (m °C)	CLR	110	−522	4			0.972	0.019
	IZS	110	−525	6			0.969	0.012
Fat %	CLR	114	3.67	0.15			0.969	0.009
	IZS	114	3.69	0.19			0.951	<0.001
Lactose %	CLR	114	4.76	0.1	4.75	0.08	0.981	0.107
	IZS	114	4.72	0.1			0.946	<0.001
Protein %	CLR	114	3.23	0.14	3.23	0.09	0.981	0.111
	IZS	114	3.22	0.17	3.22	0.12	0.989	0.522
Total bacterial count (cfu mL^{-1} 1000)	CLR	90	55	53			0.383	<0.001
	IZS	90	184	233			0.500	<0.001

[a] CLR, Centrale del Latte di Roma SpA; IZS, Istituto Zooprofilattico Sperimentale del Lazio e della Toscana “M. Aleandri”. The mean and standard deviation (SD) are only for normally distributed variables.

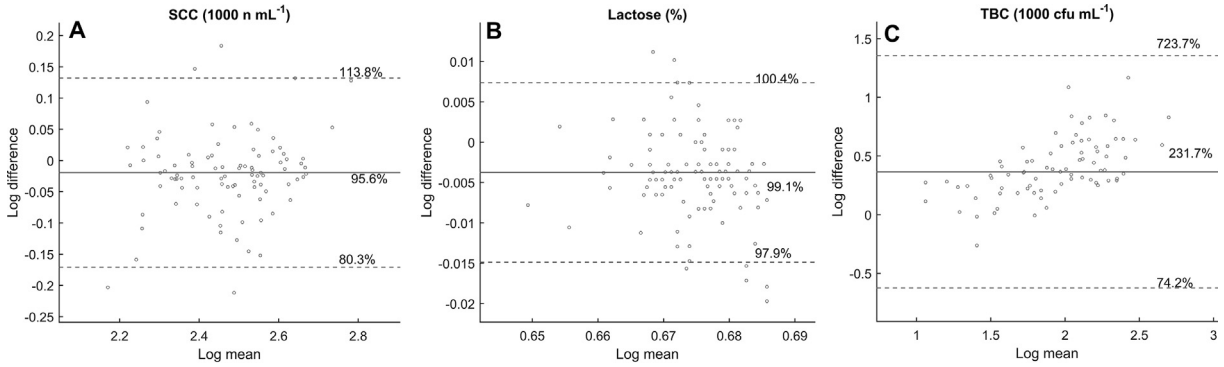

Fig. 1. Logarithmic Bland–Altman plot for (A) somatic cell count (SCC), (B) lactose and (C) total bacterial count (TBC) measurements. Log difference [log (IZS)-log (CLR)] are plotted against their log mean. The solid horizontal line represents the mean difference, while dashed lines represents upper and lower limit of agreement (LOAs). By construction, 95% of the differences lie between upper and lower LOAs. In panel A, it can be seen that IZS somatic cell count measurements are on average 95.6% of those from CLR and IZS SCC measurements are mostly (95%) from 80.3% to 113.8% of the corresponding CLR measurements.

mean ratio of CLR versus IZS measurements; e.g., IZS somatic cell count measurements are on average 95.6% of those from CLR, suggesting the presence of a bias, which was statistically confirmed.

3.3. Linear model performance and validation

The linear model was significant ($F(13,40) = 5.59$; $p < 0.01$, and the model explained 64.5% of the total variance. All the predictors significantly contributed to the model, except for weather temperature WT and CLR total bacterial count (CLRTBC, ANOVA results detailed in Table 4). The model highlighted that collecting route CR and milk temperature MT were the factors that most influenced the difference between CLR and IZS TBC measurements.

The model's predicted CLR total bacterial counts were compared with the observed (measured) CLR total bacterial counts to validate the linear model. More than half (51.9%) of the predicted CLR total bacterial counts had an absolute error less than 20%. Detailed results are shown in Fig. 2.

Choosing a reference value of 100 (1000 cfu mL^{-1}) for total bacterial count (TBC), a comparison was made between predicted (from the model) and observed (measured) IZS TBC exceeding the reference value. Interpreting the measured IZS TBC as ground truth, it was found that the model lead to 6 false positives (predicted IZS TBC > 100, while observed IZS TBC <100), 26 true positives, 18 true negatives and 4 false negatives. Since the model was applied to 54 twin samples, this resolved to an 81.3% positive predictive value and to an 81.8% negative predictive value in detecting samples with total bacterial count higher than the chosen reference value.

Table 4
Linear model analysis of variance.[a]

Factor	Df	Sum Sq	Mean Sq	F value	Pr (>F)
CLRTBC	1	4.0×10^{-5}	4.0×10^{-5}	$1.2 \cdot 10^{-3}$	0.97
TIME	1	0.45	0.45	12.79	<0.001[a]
MT	1	0.20	0.20	5.76	0.02[b]
CR	9	1.90	0.21	5.99	<0.001[a]
WT	1	7.2×10^{-3}	7.2×10^{-3}	0.20	0.65
Residuals	40	1.41	0.04		

[a] CLRTBC: CLR total bacterial count; TIME: number of days between CLR and IZS analysis; MT: milk temperature; CR: collecting route; WT: weather temperature. Significance codes: [a], < 0.001; [b], < 0.05.

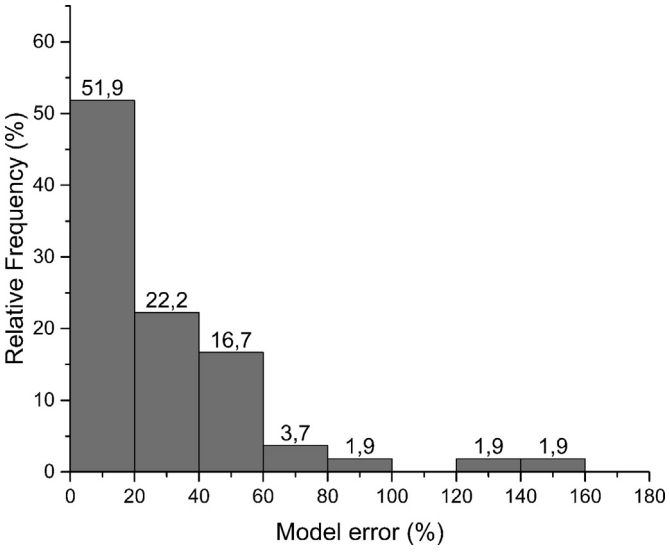

Fig. 2. CLR total bacterial count model error (predicted CLR TBC − observed CLR TBC) distribution; using the model, 51.9% of the predicted CLR total bacterial counts have an absolute error less than 20%.

4. Discussion

A notable finding was the lack of normality in most parameter distributions from both data sources. Specifically, a non-normal distribution was found for somatic cell count, freezing point, fat and lactose%, as well as total bacterial count measurements (Table 2). This aspect is worth of special attention, since several Codex Alimentarius guidelines (FAO/WHO, 2006; Horwitz, 1995) strongly rely on normality of measurement distributions. Despite the relevance of this aspect, limited data exists on the statistical distributions of milk parameters, with the notable exception of Napel et al. (2009), who found evidence of non- normal distribution of SCC and that a combination of four normal and one exponential distributions did fit the experimental sample. Litwińczuk, Król, Brodziak, and Barłowska (2011) and O'Connell, McParland, Ruegg, O'Brien, and Gleeson (2015) did recognise and deal with absence of normality in SCC in individual milk, and Jamrozik and Schaeffer

(2011) highlighted that data from healthy and mastitic cows exhibited markedly different distributions. Our study did not focus on the investigation of milk parameters statistical distributions; however, we could confirm non-normality of SCC measurements on bulk tank milk, as well of non-normality of FP, Fat, Lactose and TBC. Conversely, we could assess normality in protein measurements.

With respect to the specific issue of data agreement between self-controls and official measurements, correlation analysis suggested for a very good measurement association for Protein, SCC and Fat, good association for TBC, moderately good association for Lactose, and a poor association between CLR and IZS freezing point determinations (Table 3). The last issue is currently under deeper investigation, although it should be noted that this correlation is strongly influenced by the least significant digits of the measurement (1 m°C), close to the precision limit of the measurement method. However, data correlation shall be well distinguished from data agreement, and even though correlation between two measurements is a long-standing tool to assess their concordance, some authors expressed criticism on this approach (Bland & Altman, 1995), mostly because this approach can mask the presence of a systematic bias between the two methods, but also because, in a number of situations, the difference in measurements is somehow related to the magnitude of the measurement itself. Moreover, correlation may be affected by the range and distribution of the variables (Bland & Altman, 2003). Regarding this aspect, in our study CLR measurements were numerically higher for SCC and Lactose content, while IZS measurements exceeded CLR's ones for FP and TBC. In agreement with the literature on the cited statistical methods (Bland & Altman, 2003; Giavarina, 2015), we showed that a more informative comparison can be made using Bland–Altman plots. For example, while correlation analysis only indicates the agreement between CLR and IZS SCC as a very good agreement ($r = 0.88$; Table 3), the real validity of the agreement itself may be better judged – as either indeed good or to be improved – by looking at the Bland–Altman plot for SCC (Fig. 1A). Examining this plot, it emerges that any IZS SCC determination can be from 80.3% to 113.8% of the CLR SCC. Whether a range of agreement more than 30% is adequate or not it is an issue that can deserve consideration when defining risk management indicators. Conversely, Lactose measurements had a lower correlation ($r = 0.66$; Table 3) while the Bland and Altman limits of agreement (Table 3; Fig. 1B) show that IZS Lactose determination can be from 97,9%–100.4% of the CLR one.

With respect to the specific samples analysed in our study, findings on total bacterial count suggested a deeper investigation. Twinned TBC data showed good or moderately good correlation and both absolute ranges of values, though different, confirmed the adequate quality of hygienic-sanitary conditions of the collected milk. However, as shown in Fig. 1C, a clear dependency of the difference on the mean emerged. Whether this heteroscedasticity can be ascribed to absence of normality, features of the measurement methods, deviation from the expected sample management or to a combination of all those factors is currently under investigation. Within this specific sample, we can only observe that measurements with higher total bacterial count (IZS) are more likely to diverge, spanning through a very wide range of percentages of the corresponding CLR determinations (Table 3).

This lack of agreement, if confirmed, may be inappropriate in the context of assessment of the safety and economic value of milk production and needs further insight. To cope with this additional analysis, we used the traceability information recorded in the study's database and designed a linear model to investigate the influence of several, plausible independent variables (time between sampling and analysis, truck tank temperature, weather temperature, collecting route) on TBC measurement differences between CLR an IZS. The model significantly explained a relevant part of the TBC measurement differences (Table 4). In particular, it highlighted that the most influential factor in the specific TBC increase in the IZS determinations was the collecting route (the fixed and predetermined sequence of farms whose milk is collected by each truck tank), followed by milk temperature in the tank and by the time elapsed between sampling and analysis.

We included in the model CLR total bacterial count to constitute a form of "baseline" measurement but, despite its intuitive expected impact on the measurement difference, its influence on the dependent variable was very small. As for milk temperature and time elapsed between sampling and analysis, their (expected) influence on TBC increase is a direct consequence of the bacterial level in the milk sample. We noted that time or sample conservation before analysis had a significant influence on TBC increase (Table 4). Those results are consistent with those reported by O'Connell, Ruegg, Jordan, O'Brien, and Gleeson (2016), who reported the increase of TBC of milk stored at 6 °C. The major influence of the collecting route on the TBC measurement difference may suggest that milk coming from certain groups of farmers may have had an effect on bacterial growth. All the above observations should be corroborated by stronger evidence, and, if confirmed, might be used to derive a more accurate detection and weighting of factors, for example management factors, which could have represented the ultimate cause of TBC measurement difference between the specific CLR and IZS samples of our study.

Besides the identification of factors influencing the total bacterial count, another practical value of the identified model could be found in its ability of predicting, at milk truck unloading, which milk samples will eventually be found as exceeding a reference TBC value by the Official Control. Even though this result should be corroborated by simulations based on new and extended data, our preliminary analysis showed that, choosing a reference value of 100 (1000 cfu mL^{-1}) for total bacterial count, the model was able, at truck unloading, to correctly identified more than 80% of those samples that will be eventually detected as exceeding the reference value by the Official Control. This result could be of relevant economical interest if applied in the context of quality-based payment programs, as well as in waste-reduction initiatives.

As a further observation, possible limitations of our linear model should be taken into account. Linear models are a well-known tool in dairy science (Borneman, Stiegert, & Ingham, 2015; Ramsahoi, Gao, Fabri, & Odumeru, 2011), and they are frequently used to investigate the effect of management practises on somatic cell count and total bacterial count. To our best knowledge, this study is the first using linear models to investigate dairy measurement differences. As reported in the results section, the model did explain a relevant part of the TBC measurement differences, but other factors, not included in the model and still to be identified, could affect the differences. Furthermore, this model, being linear, could not sufficiently model nonlinear relationships between the observed variables.

5. Conclusions

Self-monitoring has gained a relevant position in the food and dairy industry, as a tool to systematically promote food safety and quality. Moreover, all 189 countries adhering to the Codex Alimentarius Commission do share a common food safety control approach based on internationally agreed standards, guidelines, codes of practise and accreditation schemes for laboratory measurements. This study aimed at contributing to increase the yet scant scientific evidence on the consistency between measurement

results carried out by self-monitoring practise and reference laboratories in real operational settings. This study confirmed a moderate-to-good agreement between paired measurements in raw milk control at the interface between farms and the food industry, in particular for protein, somatic cell count, fat, lactose, and for freezing point determinations. Discordance was found for TBC, which, in principle, could not be attributed to differences in the analytical methodologies. Deeper analysis showed that collecting route and milk transportation temperature may reasonably contribute to this difference.

The results of the study do apply to evaluation and identification of issues arising from practical application of food safety laboratory measurements within the investigated scenario; thus, general considerations should be derived from our findings only with some caution. In the meanwhile, the processing and statistical methodology hereby described and implemented might be reliably and effectively used at a general level.

Acknowledgements

The study was carried out within the framework of the project ALERT ("Integrated system of bio-sensors and sensors (BEST) for the monitoring of wholesomeness and quality, as well as for traceability in the cow milk chain", www.alert2015.it) funded by the Italian Ministry of Economic Development under the Call Industria 2015 "New technology for made in Italy" (grant MI 00195). The Authors gratefully acknowledge Centrale del Latte di Roma SpA and Società Cooperativa Agricola Lattepiù for their assistance and for making milk traceability data available.

References

Atkinson, G., & Nevill, A. M. (1998). Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Medicine, 26*, 217–238.

Auldist, M. J., Coats, S., Rogers, G. L., & McDowell, G. H. (1995). Changes in the composition of milk from healthy and mastitic dairy cows during the lactation cycle. *Australian Journal of Experimental Agriculture, 35*, 427–436.

Bland, J. M., & Altman, D. G. (1995). Comparing methods of measurement: Why plotting difference against standard method is misleading. *Lancet, 346*, 1085–1087.

Bland, J. M., & Altman, D. G. (1999). Measuring agreement in method comparison studies. *Statistical Methods in Medical Research, 8*, 135–160.

Bland, J. M., & Altman, D. G. (2003). Applying the right statistics: Analyses of measurement studies. *Ultrasound in Obstetrics and Gynecology, 22*, 85–93.

Borneman, D. L., Stiegert, K., & Ingham, S. (2015). Limitations of on-site dairy farm regulatory debits as milk quality predictors. *Journal of Dairy Science, 98*, 1526–1538.

Dong, F., Hennessy, D. A., & Jensen, H. H. (2012). Factors determining milk quality and implications for production structure under somatic cell count standard modification. *Journal of Dairy Science, 95*, 6421–6435.

EU. (2002). Regulation (EC) No 178/2002 of the European Parliament and of the council of 28 January 2002 laying down the general principles and requirements of food law, establishing the European Food Safety Authority and laying down procedures in matters of food safety. *Official Journal of the European Union, L31*, 1–24.

FAO/WHO. (2006). *CAC/GL 59-2006. Guidelines on estimation of uncertainty of results*. Rome, Italy: Joint FAO/WHO Codex Alimentarius Commission.

Giavarina, D. (2015). Understanding Bland Altman analysis. *Biochemia Medica, 25*, 141–151.

Hamann, J., & Krömker, V. (1997). Potential of specific milk composition variables for cow health management. *Livestock Production Science, 48*, 201–208.

Horwitz, W. (1995). Protocol for the design, conduct and interpretation of method-performance studies: Revised 1994 (Technical Report). *Pure and Applied Chemistry, 67*, 331–343.

Jamrozik, J., & Schaeffer, L. R. (2011). Test-day somatic cell score, fat-to-protein ratio and milk yield as indicator traits for sub-clinical mastitis in dairy cattle. *Journal of Animal Breeding and Genetics, 129*, 11–19.

Litwińczuk, Z., Król, J., Brodziak, A., & Barłowska, J. (2011). Changes of protein content and its fractions in bovine milk from different breeds subject to somatic cell count. *Journal of Dairy Science, 94*, 684–691.

Muehlhoff, E., Bennett, A., & McMahon, D. (2013). *Milk and dairy products in human nutrition*. Rome, Italy: Food and Agriculture Organisation of the United Nations.

Napel, J. ten, de Haas, Y., de Jong, G., Lam, T. J. G. M., Ouweltjes, W., & Windig, J. J. (2009). Characterization of distributions of somatic cell counts. *Journal of Dairy Science, 92*, 1253–1264.

O'Connell, A., McParland, S., Ruegg, P. L., O'Brien, B., & Gleeson, D. (2015). Seasonal trends in milk quality in Ireland between 2007 and 2011. *Journal of Dairy Science, 98*, 3778–3790.

O'Connell, A., Ruegg, P. L., Jordan, K., O'Brien, B., & Gleeson, D. (2016). The effect of storage temperature and duration on the microbial quality of bulk tank milk. *Journal of Dairy Science, 99*, 3367–3374.

R Core Team. (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Core Team.

Ramsahoi, L., Gao, A., Fabri, M., & Odumeru, J. A. (2011). Assessment of the application of an automated electronic milk analyzer for the enumeration of total bacteria in raw goat milk. *Journal of Dairy Science, 94*, 3279–3287.

Tukey, J. W. (1977). *Exploratory data analysis*. London, UK: Pearson.