Modified Versus Producer Milk Calibration: Mid-Infrared Analyzer Performance Validation¹

K. E. Kaylegian,* J. M. Lynch,* G. E. Houghton,† J. R. Fleming,‡ and D. M. Barbano*²

*Northeast Dairy Foods Research Center, Department of Food Science, Cornell University, Ithaca, NY 14853

†Kestrel Software Consulting, Berkshire, NY 13736

‡USDA, Agricultural Marketing Service, Texas Milk Marketing Area, P.O. Box 110939, Carrollton, TX 75011

ABSTRACT

Our objective was to determine the validation performance of mid-infrared (MIR) milk analyzers, using the traditional fixed-filter approach, when the instruments were calibrated with producer milk calibration samples vs. modified milk calibration samples. Ten MIR analyzers were calibrated using producer milk calibration sample sets, and 9 MIR milk analyzers were calibrated using modified milk sample sets. Three sets of 12 validation milk samples with all-laboratory mean chemistry reference values were tested during a 3-mo period. Calibration of MIR milk analyzers using modified milk increased the accuracy (i.e., better agreement with chemistry) and improved agreement between laboratories on validation milk samples compared with MIR analyzers calibrated with producer milk samples. Calibration of MIR analyzers using modified milk samples reduced overall mean Euclidian distance for all components for all 3 validation sets by at least 24% compared with MIR analyzers calibrated with producer milk sets. Calibration with modified milk sets reduced the average Euclidian distance from all-laboratory mean reference chemistry on validation samples by 40, 25, 36, and 27%, respectively for fat, anhydrous lactose, true protein, and total solids. Between-laboratory agreement was evaluated using reproducibility standard deviation (s_R). The number of single Grubbs statistical outliers in the validation data was much higher (53 vs. 7) for the instruments calibrated with producer milk than for instruments calibrated with modified milk sets. The s_R for instruments calibrated with producer milks (with statistical outliers removed) was similar to data collected in recent proficiency studies, whereas the s_R for instruments calibrated with modified milks was lower than those calibrated with producer

milks by 46, 52, 61, and 55%, respectively for fat, anhydrous lactose, true protein, and total solids.

Key words: infrared milk analysis, validation, calibration

INTRODUCTION

Mid-infrared (MIR) milk analysis is an indirect method that requires instrument calibration with milk samples that have reference values established by reference chemistry methods. Traditional milk analysis has been done using fixed-filter instruments, with specific wavelengths for fat, protein, and lactose determination. Fourier transform infrared (FTIR) instruments for milk analysis have options for using a fixed-filter wavelength mode or a full spectral calibration mode. The principles underlying the MIR analysis of milk are presented elsewhere (Biggs et al., 1987). In the present study, all instruments were operated in the fixed-filter wavelength mode. Use of partial least squares calibration was outside the scope and objectives of the present study.

Accuracy of MIR analysis of milk is affected by instrument factors, quality of reference chemistry, characteristics of the calibration sample set, and individual milk sample composition factors (Biggs et al., 1987; Barbano and Clark, 1989; Kaylegian et al., 2006). Characteristics of the calibration sample set that affect calibration performance include the number of samples, the range of component concentration, and distribution within the range (presence of high leverage samples), correlation of fat and protein concentrations, and changes in these characteristics between consecutive sample sets (Kaylegian et al., 2006).

Traditional calibration sample sets are made from preserved raw individual producer milk samples ($n = 8$ to 12) obtained locally and generally analyzed by a single laboratory using reference chemistry methods. Producer milk calibration sets often have a narrow range of component concentrations, high leverage samples, and a positive correlation between fat and protein concentrations. These factors can cause reduced accuracy of the calibration, which is indicated by a larger

Received January 3, 2006.

Accepted March 16, 2006.

¹Use of names, names of ingredients, and identification of specific models of equipment is for scientific clarity and does not constitute any endorsement of product by authors, Cornell University or the Northeast Dairy Foods Research Center.

²Corresponding author: dmb37@cornell.edu

confidence interval around the calibration linear regression line (Kaylegian et al., 2006). An approach to overcoming these limitations is the use of preserved pasteurized modified milk calibration samples. The modified milk calibration approach eliminated high leverage samples and substantially decreased uncertainty of instrument calibration by reducing the size of the 95% confidence interval around the linear regression calibration line for each component (Kaylegian et al., 2006). The objective of this research was to compare the accuracy of testing (using independent sets of validation samples) of 2 groups of MIR analyzers: one group using traditional individual producer milk samples vs. another group using modified milk samples for calibration.

MATERIALS AND METHODS

Experimental Design

A total of 19 MIR analyzers in 13 laboratories (Cornell University, USDA Federal Milk Market, and commercial) were used. One group of MIR analyzers ($n = 10$) was calibrated using producer milk calibration samples and calibration procedures used in those laboratories for milk payment testing. Another group of MIR analyzers ($n = 9$) was calibrated using modified milk samples manufactured at Cornell University as previously described (Kaylegian et al., 2006).

The validation sample sets were assembled from local individual producer milks by a USDA Federal Milk Market laboratory known to have calibration samples with a wide component range and a good distribution of samples within the component ranges (Kaylegian et al., 2006). The validation milk samples used to evaluate calibration performance were not part of the calibration samples in either group of instruments being evaluated in this study.

Manufacture of the modified milk calibration samples began on Monday of the first week. These pasteurized, dichromate-preserved samples were shipped on wet ice by overnight delivery and arrived at each laboratory on Thursday morning. Each laboratory was sent several sets of samples so that a fresh set of unopened samples could be used for each chemical analysis method and each MIR milk analyzer. All chemical analyses of the modified milk samples were completed by Tuesday of the second week and the data were sent to Cornell for calculation of the all-laboratory mean reference values. The 9 instruments in the study using modified milk calibration were calibrated (i.e., adjustment of slope and intercept) with these samples. The remaining 10 instruments in this study were calibrated using producer milk calibration samples each laboratory normally used as the calibration for their payment testing.

On Monday of the second week, the validation samples ($n = 12$ from individual farms) were assembled and shipped by overnight delivery to all laboratories. On Wednesday of the second week, the validation milk samples were tested on all 19 instruments in the study. Each laboratory was instructed to calibrate their MIR analyzer with the appropriate calibration sample set (either modified milk or their own producer milk calibration set), make any necessary adjustments to the slope and intercept of their calibration, and then immediately test the validation samples. The MIR results were immediately returned to Cornell. After testing the validation samples by MIR, additional sets of the validation samples were analyzed by all laboratories using reference chemistry methods and those results were returned to Cornell by Tuesday of the third week for calculation of the all-laboratory mean chemical reference values for the validation samples. This was repeated 3 times.

The MIR-predicted value for each of the validation samples was compared with the all-laboratory mean reference chemistry value. The mean difference (**MD**) and standard deviation of the difference (**SDD**) for each milk component of each validation set were determined for each instrument. Validation performance was evaluated by plotting the SDD as a function of MD for each component for the modified milk and producer milk calibrated instruments and calculating the Euclidian distance (**ED**) for each component, on each instrument, as an index of the impact of type of calibration set on testing accuracy. Reproducibility standard deviation (**S_R**) was used as an index to evaluate the closeness of agreement of results between laboratories within each type of calibration set.

Producer Calibration Samples Used to Calibrate MIR Analyzers

Each laboratory that calibrated their MIR analyzer with producer milk calibration samples used the procedures that they routinely used for milk payment testing. The general process used to assemble producer milk calibration samples (USDA Federal Milk Market and commercial) was described by Kaylegian et al. (2006). Producer calibration sample sets consisted of 10 to 12 milk samples, depending on the laboratory. The raw milk samples were preserved, split into vials and refrigerated (4°C). The samples were analyzed using reference chemistry methods by 1 to 4 laboratories; mean reference chemistry values were used when more than one laboratory determined the chemistry.

Modified Milk Calibration Samples Used to Calibrate MIR Analyzers

The manufacture of the 14 sample modified milk calibration set was done in the Cornell University pilot plant described by Kaylegian et al. (2006). Pasteurized milk was gravity separated overnight at 4°C. The gravity skim layer (90% by weight) was drained from the bottom of the tank and the cream was removed in several layers. The cream layers were analyzed for fat content and selected layers were blended to create a cream ingredient with a fat content of 22 to 27%. The gravity skim layer was further separated by centrifugal separation to reduce the fat content to <0.07%. The centrifugally separated skim milk was ultrafiltered (2×) to obtain retentate and permeate. The cream ingredient, skim UF permeate, skim UF retentate, reagent grade α -lactose monohydrate (MultiPharm, EM Science, Gibbstown, NJ), and laboratory-grade water were blended to create 14 calibration samples with a broad range and an orthogonal matrix of component concentrations (Kaylegian et al., 2006). Samples were preserved with potassium dichromate, split into vials, and refrigerated (4°C).

The modified milk calibration sets used in this study were produced at Cornell University and shipped with wet ice overnight to each of the participating laboratories for analysis. Samples were analyzed using reference chemistry methods by at least 7 laboratories for fat, true protein, and total solids, and by at least 4 laboratories for lactose using an enzymatic method. The all-laboratory mean chemistry values were used as the reference chemistry values for calibration of 9 MIR instruments. Over the course of the study, 3 different batches of modified milk calibration samples were produced and used with all 9 MIR analyzers.

Validation Samples

Validation sets were obtained from a USDA Federal Milk Market Laboratory and consisted of 12 individual farm raw milk samples. The chemical analyses of the validation samples were performed by the same group of laboratories that determined the reference chemistry values for the modified milk calibration sets. The all-laboratory mean chemistry values (Table 1) were used as the reference chemistry values for validation. All laboratories analyzed validation samples from the same batch regardless of the type of samples they used to calibrate an MIR analyzer.

Chemical Analyses

Chemical analyses of all samples were conducted using the following AOACI (2000) methods: fat by modi-

fied Mojonnier ether extraction (method 989.05; 33.2.26), true protein by Kjeldahl analysis (method 991.22; 33.2.13), total solids by oven drying (method 990.20; 33.2.44), and lactose determined by enzyme analysis (method 984.15; 33.2.24) modified to measure lactose by weight instead of volume with the results expressed as anhydrous lactose. For instruments calibrated with producer milks, lactose by difference [lactose = total solids – (fat + true protein + ash + 0.19)] was used as reference. Ash was estimated using an updated version of the equation described by Lynch et al. (1990): ash = (0.0596 × true protein) + 0.5379.

MIR Analysis

A total of 19 instruments in 13 laboratories were used. No laboratory had more than 2 instruments. Both fixed-filter and FTIR instruments were included in this study (Table 2). All of the laboratories participated in the monthly precalibration of instruments according to the procedures of the USDA Federal Milk Markets. Precalibration procedures ensure that the instruments perform within the specified mechanical and electronic tolerances as described by Lynch et al. (2006). All instruments, including FTIR instruments, used fixed fat B, lactose, protein, and fat A filter wavelengths with corresponding reference wavelengths, as described by Kaylegian et al. (2006). All instruments in this study were operated in the traditional fixed-filter mode. The non-FTIR instruments (e.g., Milkoscan 134, 255, 300, and 605) used the classical sample filter wavelengths of 3.48, 9.61, 6.46, and 5.73 μm and reference filter wavelengths of 3.60, 7.70, 6.70, and 5.60 μm for fat B, lactose, protein, and fat A, respectively. The traditional fixed-filter wavelengths used in the Foss FT 6000 were not disclosed by the manufacturer (Foss Electric, Hillerød, Denmark) but presumably are similar to those recommended in their model FT 120. Several FTIR instruments and a Foss FT 120 participated in the study and were operated in a fixed-filter mode. The sample filter wavelengths used on the Delta FTIR instruments were 3.51, 9.54, 6.60, and 5.79 μm and the reference wavelengths were 3.56, 7.79, 6.77, 5.62 μm for fat B, lactose, protein, and fat A, respectively. These wavelengths are slightly different than for the fixed-filter instruments because it is necessary to use a narrower bandwidth than the fixed-filter mode on the FTIR instruments. Instruments used fixed intercorrection factors that were established as part of the precalibration procedures, except in the case of FT 6000 instruments, in which the fixed intercorrection factors were established by Foss Electric (Hillerød, Denmark). Instruments calibrated with producer milk samples followed the procedures that each laboratory normally used to

Table 1. All-laboratory mean chemistry values for 3 sets of validation milks

Sample	Validation set 1				Validation set 2				Validation set 3			
	Fat	Lactose ¹	True protein	Total solids	Fat	Lactose	True protein	Total solids	Fat	Lactose	True protein	Total solids
	%											
1	2.7827	4.5601	3.3409	11.8970	3.2152	4.6703	2.9610	11.9421	3.2999	4.6825	2.9554	12.0415
2	3.7680	4.2849	3.1807	12.3185	3.6620	4.5131	3.0624	12.2923	2.6683	4.5654	3.1446	11.6225
3	3.5057	4.6206	3.0099	12.2729	3.4490	4.6489	3.0186	12.2922	3.2968	4.4886	3.0243	11.8994
4	3.5671	4.6365	2.9534	12.2985	3.8902	4.6437	3.0798	12.7287	3.5755	4.6555	2.9816	12.3534
5	4.1978	4.6016	3.3222	13.2548	2.5159	4.5538	3.0830	11.3836	3.9482	4.5992	2.9382	12.5763
6	4.2975	4.4904	3.3992	13.3195	4.0685	4.6194	3.2549	13.0449	4.1563	4.5102	3.2377	13.0156
7	4.5256	4.2959	3.5040	13.5983	4.1455	4.4887	3.2926	13.0442	4.0325	4.6045	3.1713	12.9259
8	3.4107	4.2943	3.0657	11.8862	3.3256	4.3606	2.9599	11.7588	3.5144	4.3180	2.9623	11.9281
9	4.3811	4.6201	3.1670	13.3056	4.7299	4.6159	3.3962	13.8816	4.7223	4.4765	3.5244	13.9109
10	4.1987	4.5438	3.3722	13.2720	4.2841	4.7024	3.1197	13.1940	5.5352	4.5482	3.8211	15.0876
11	4.5451	4.5585	3.1284	13.3694	4.8575	4.4765	3.5925	14.1034	4.3867	4.7030	3.0185	13.2292
12	5.8269	4.4873	3.8973	15.4178	5.6056	4.5365	3.9394	15.2520	4.3141	4.6001	3.5178	13.5742
Mean	4.0839	4.5103	3.2784	13.0175	3.9799	4.5692	3.2300	12.9098	3.9542	4.5626	3.1914	12.8470

¹Lactose reported as anhydrous lactose in all validation sets.

calibrate (i.e., adjust slope and bias) their instrument for milk payment testing. Instruments calibrated with modified milk samples followed the calibration procedures described by Kaylegian et al. (2006). The validation samples were analyzed for fat, true protein, anhydrous lactose, and total solids on each instrument in duplicate. Data were analyzed as the mean of duplicates. The results were sent electronically to Cornell University for evaluation of validation performance.

Validation Performance of Modified and Producer Milk Calibrations

MD and SDD. The MD and SDD were determined for fat, anhydrous lactose, true protein, and total solids for each instrument for each validation set. The difference value was calculated for each sample in the set by subtracting the reference chemistry value from the MIR predicted value. The mean of the individual sample differences (MD) and the standard deviation of these differences (SDD) were calculated for the entire validation set (12 samples) for each instrument. The MD and

SDD values for all instruments were compared using a Euclidian distance plot by calibration set type (modified milk or producer milk) and component (fat, true protein, anhydrous lactose, and total solids).

ED. The ED is a statistical measure of similarity that is the distance from an individual data point to the center point of a cluster of similar data (Massart et al., 1988). In this study, ED was used as a measure of the distance of each instrument's MD and SDD for each milk component from the mean reference chemical value for the validation samples and reflects the accuracy of the MIR method. The center point (mean chemistry value for the validation set) in this study was set at (0, 0). The ED was calculated as follows:

$$ED = \sqrt{(MD)^2 + (SDD)^2}.$$

The ED values were determined for each instrument for each validation set for fat, true protein, anhydrous lactose, and total solids on MIR instruments calibrated with modified milk sets or producer milk sets. A mean

Table 2. Type of calibration, makes, and models of mid-infrared analyzers used in this study

Calibration set type	No. of instruments	Instrument type	Instrument manufacturer ¹	Instrument model
Modified milk	5	Filter FTIR ²	Foss	MS 134, MS 255, MS 303, MS 605
	4		Foss	FT 120
Producer milk	4	Filter	Delta	LactoScope FTIR
			Foss	MS 134, MS 4000
	6	FTIR	Bentley	2000
			Foss	FT 6000
			Delta	LactoScope FTIR

¹Foss, Hillerød, Denmark; Delta, Drachter, The Netherlands; and Bentley, Chaska, MN.

²FTIR = Fourier transform infrared.

ED for each milk component for the group of instruments for each calibration method for each of the 3 validation sets was calculated. The data were analyzed using the GLM procedure in SAS (Version 8e, 2001; SAS Institute, Cary, NC) to determine if the mean ED for the modified milk calibration vs. producer milk calibration were different. The ANOVA model was as follows: calibration set type and validation set were defined as class variables, and the model was $ED = \text{calibration set type} + \text{validation set} + \text{calibration set type} \times \text{validation set} + \text{error}$. If the model was significant ($P < 0.05$), the mean ED values were compared using a t -test at $P < 0.05$.

Repeatability and Reproducibility Standard Deviation. The statistical metric for within-laboratory variation of a method is the repeatability standard deviation (s_r) and the metric for between-laboratory variation is the reproducibility standard deviation (s_R). These values are commonly calculated as part of the process of method performance validation. The calculation and practical use of these metrics in the laboratory were described by Lynch (1998). In the present study, the impact of the type of calibration set on agreement between laboratories was determined by comparison of the s_R on the validation samples.

Our validation data were analyzed by the statistical procedures of the AOACI (2000, Appendix D: Guidelines for collaborative study procedures to validate characteristics of a method of analysis) to determine the s_R for instruments calibrated with producer milk samples vs. modified milk samples. The outlier identification procedures of the AOACI were used to remove individual laboratory outlier data points ($\alpha = 0.025$).

RESULTS AND DISCUSSION

Validation of MIR Analyzers Calibrated with Modified and Producer Milk Samples

Comparison of ED Plots. There was no more than 1 outlier instrument in any given validation set for the modified milk or the producer milk calibration sets. When removal of the MD or SDD data from a single instrument reduced the range of MD or SDD by at least 50% across instruments within a calibration set type it was considered an outlier. The data presented in Figures 1 to 4 and Table 3 are shown with these outliers removed.

The scatter of MD and SDD for fat, lactose, true protein, and total solids across all 3 validation sets (shown in Euclidian distance plots) was reduced by MIR calibration with modified milk sets (Figures 1a, 2a, 3a, and 4a) compared with calibration using producer milk sets (Figures 1b, 2b, 3b, and 4b). Generally, the modified milk calibration sets produced smaller MD and SDD

Figure 1. Plot of mean difference (MD) and standard deviation of the difference (SDD) for fat (g/100 g) from mid-infrared analyzers calibrated with (a) modified milk sets ($n = 26$) and (b) producer milk sets ($n = 30$). Outlier point removed (1 point, Figure 1b).

from the all-laboratory mean reference chemistry than producer milk calibration sets. In a validation study of 50 fixed-filter MIR milk analyzers calibrated with producer samples conducted by Ginn and Packard (1989) over a 6-mo period, the MD and SDD for the group of instruments. In general, the MD and SDD values for the validation of modified milk calibration

Figure 2. Plot of mean difference (MD) and standard deviation of the difference (SDD) for lactose (g/100 g) from mid-infrared analyzers calibrated with (a) modified milk sets ($n = 26$) and (b) producer milk sets ($n = 30$). Outlier point removed (1 point, Figure 2a, 3 points, Figure 2b).

Figure 3. Plot of mean difference (MD) and standard deviation of the difference (SDD) for protein (g/100 g) from mid-infrared analyzers calibrated with (a) modified milk sets ($n = 26$) and (b) producer milk sets ($n = 30$). Outlier point removed (1 point, Figure 3a).

in our study (Figures 1a, 2a, 3a, and 4a) are smaller than those reported by Ginn and Packard (1989).

Comparison of Mean ED. The mean ED values for all 3 validation sets were consistently lower for all milk components for instruments calibrated with modified milk sets compared with instruments calibrated with

producer milk sets (Table 3). There was a reduction ($P < 0.05$) of overall mean ED values (fat, anhydrous lactose, true protein, and total solids were reduced by 40, 25, 36, and 27%, respectively) for instruments calibrated with modified milk sets compared with producer milk sets for all components (Table 3).

Figure 4. Plot of mean difference (MD) and standard deviation of the difference (SDD) for total solids (g/100 g) from mid-infrared analyzers calibrated with (a) modified milk sets ($n = 26$) and (b) producer milk sets ($n = 30$). Outlier point removed (2 points, Figure 4b).

We used the average MD and SDD values reported by Ginn and Packard (1989) to calculate an ED for their data. The ED for fat and protein were 0.049 and 0.034% for data from Ginn and Packard. These values are similar to the ED values of 0.043 and 0.038% for fat and protein, respectively, for validation samples analyzed

using MIR calibrated with producer milk samples (Table 3). The mean validation ED for fat and protein for the instruments calibrated with modified milks were 0.0257 and 0.0246%, respectively (Table 3), which demonstrates better accuracy of testing than observed by Ginn and Packard (1989). It is clear, based on the comparison of the data of Ginn and Packard (1989) to the data presented in Table 3, that in spite of improvements in the quality of the hardware and software used for infrared milk analysis, the validation performance of instruments when using producer calibration samples has not changed since the report in 1989. The data in Table 3 demonstrate that calibration of fixed-filter wavelength MIR milk analyzers with modified milk calibration samples produced significantly lower ED (i.e., better agreement with reference chemistry) on validation for all milk components than calibration with producer milks.

Current Industry Practice for MIR Milk Analysis

AOACI Method. The official first action method for fat, lactose, protein, and solids in milk by an MIR spectroscopic method using fixed-filter wavelengths (AOACI, 2000; method 972.16, 33.2.31) indicates that the MD between instruments and reference method values should be $\leq 0.05\%$ for fat, protein, and lactose, and $\leq 0.09\%$ for total solids, but gives no estimates of method performance for comparison of results among different instruments that are calibrated with different reference samples. The information on method performance for AOACI method 972.16 was collected before the current guidelines for collaborative studies. The methodology for evaluation of method performance has evolved over the years, particularly in the late 1980s and early 1990s to its current status, where there are specific guidelines that are accepted internationally for conducting collaborative studies and calculation of within- and between-laboratory metrics of method performance (AOACI 2000, Appendix D: Guidelines for collaborative study procedures to validate characteristics of a method of analysis).

State Regulations. The states of New York and Wisconsin have specified procedures for electronic testing of milk components that give method performance limits for MD between instrument and reference chemistry. The New York state regulations (New York State Department of Agriculture and Markets, 2005) specify that 20 samples be used for the calibration of fat over the range of 3.0 to $\geq 4.5\%$, and have no specifications for calibration of other milk components. The Wisconsin regulations (Wisconsin Administrative Code, 2005) specify a calibration set of 12 individual herd samples with a fat range of at least 2.5 to 5.0%, a protein range

Table 3. Mean Euclidian distance (ED) of validation samples (g/100 g) using mid-infrared (MIR) analyzers calibrated with modified milk and producer milk samples¹

Component	Validation set	Modified milk	Producer milk	Percent reduction ²	LSD ³
Fat	1	0.0263	0.0376	30	0.0061
	2	0.0279	0.0504 ⁴	45	
	3	0.0233	0.0416	44	
	Mean	0.0257 ^b	0.0430 ^a	40	
Anhydrous lactose	1	0.0182	0.0278 ⁴	34	0.0057
	2	0.0273 ⁴	0.0317 ⁴	14	
	3	0.0279	0.0371 ⁴	25	
	Mean	0.0242 ^b	0.0322 ^a	25	
True protein	1	0.0208	0.0347	40	0.0052
	2	0.0270 ⁴	0.0415	35	
	3	0.0267	0.0388	31	
	Mean	0.0246 ^b	0.0383 ^a	36	
Total solids	1	0.0472	0.0555 ⁴	15	0.0089
	2	0.0397	0.0566	30	
	3	0.0437	0.0675 ⁴	35	
	Mean	0.0437 ^b	0.0597 ^a	27	

^{a,b}Values in the same row with different superscripts are different ($P < 0.05$).

¹Number of MIR analyzers calibrated with modified milk samples $n = 9$; with producer milk samples $n = 10$. Data with outliers were removed.

²Percent reduction = [(mean ED producer milk – mean ED modified milk)/mean ED producer milk] $\times$ 100.

³LSD = Least significant difference ($P < 0.05$).

⁴Outlier value removed.

of at least 2.7 to 3.4%, and a total solids range of at least 11 to 13%. The New York regulations specify the calibration performance limits at a MD (between reference values and instrument values) of ≤ 0.02 and an SDD of ≤ 0.04 for fat and true protein. The Wisconsin regulations set the performance limits for the average MD (between reference value and instrument values) of triplicate analyses at $\pm 0.044\%$ for fat and protein, and $\pm 0.084\%$ for total solids. No data that provides a basis for these performance limits were provided.

USDA Federal Milk Market Laboratories. The laboratories of the USDA Federal Milk Marketing Orders and their affiliated laboratories have been engaged in a long-term research program to improve the accuracy of milk component measurement for use in producer payment testing. The program initially focused on systematic improvement of the accuracy of chemical reference methods that are used as the basis for calibration of MIR milk analyzers. The Babcock (Barbano et al., 1988; Lynch et al., 1995, 1996, 1997a, 2003), ether extraction (Barbano et al., 1988; Lynch et al., 1996, 1997a, 2003), Kjeldahl (Barbano et al., 1990, 1991; Lynch et al., 1998; Lynch and Barbano, 1999), and total solids methods (Clark et al., 1989a,b) for milk analysis have been optimized to improve their within- (s_r) and between- (s_R) laboratory performance, and the method performance values are included in the AOACI (2000)

method descriptions (Babcock method 989.04, 33.2.27; ether extraction method 989.05, 33.2.26; Kjeldahl CP method 991.20, 33.2.11; Kjeldahl NPN method 992.21, 33.2.12; Kjeldahl true protein method 991.22, 33.2.13; Kjeldahl casein nitrogen method 998.06, 33.2.65; total solids method 990.20, 33.2.44; solids-not-fat method 990.21, 33.2.45). Periodically, the USDA Federal Milk Markets have published the results of proficiency testing for some of these methods (Lynch et al., 1994, 1997b).

Over several years, the USDA Federal Milk Market Laboratories have been conducting proficiency tests of MIR milk analyzer performance in their laboratories using 7 unknown milk samples in blind duplicate 6 times each year. The values for s_r and s_R for MIR milk analyzers have been calculated and are summarized in Table 4 for the month of November for the period 1999 through 2003. In general, the within-laboratory repeatability is excellent, as indicated by an s_r that is routinely below 0.01% for fat and protein. However, the between-laboratory agreement (s_R) is usually 2 to 3 times larger than the within-laboratory agreement (s_r). Compared with the performance of chemical reference methods (AOACI, 2000) for fat (ether extraction method 989.05, 33.2.26) and protein (Kjeldahl method 991.20, 33.2.11), the between-laboratory agreement (s_R) is consistently less than 2 times the within-laboratory agreement (s_r).

Table 4. Mean repeatability standard deviation (s_r) and reproducibility standard deviation (s_R) of unknown milk samples ($n = 7$ sample materials in blind duplicate) used in bimonthly laboratory proficiency testing of mid-infrared milk analyzers for selected months between 1999 and 2003¹

Year ²	No. of instruments	Fat		Protein		Total solids	
		s_r	s_R	s_r	s_R	s_r	s_R
1999	15	0.0080	0.0216	0.0068	0.0155	0.0132	0.0400
2000	14	0.0095	0.0244	0.0076	0.0166	0.0157	0.0411
2001	15	0.0075	0.0259	0.0067	0.0207	0.0123	0.0285
2002	15	0.0090	0.0228	0.0086	0.0248	0.0168	0.0504
2003	17	0.0081	0.0209	0.0063	0.0232	0.0130	0.0351

¹Statistical outliers removed using Cochran and Grubbs tests at $\alpha = 0.01$.²Data are for the month of November in each year.

The larger ratio of s_R to s_r for MIR analysis compared with reference chemical analysis methods would indicate that the between-laboratory performance of MIR milk analysis can be improved. The larger difference between within-laboratory repeatability standard deviation (s_r) and between laboratory reproducibility standard deviation (s_R) can originate from laboratory-to-laboratory differences in the reference chemistry for calibration of the MIR milk analyzers and from the characteristics of the calibration sample sets (Kaylegian et al., 2006). However, it is not uncommon for the chemistry results from 2 laboratories to agree very well for fat and protein on a set of unknown samples and at the same time, their MIR instruments do not agree very well on the same samples. In this case, the cause for disagreement in MIR results would appear to be due to differences in the characteristics of the calibration sample sets between the 2 laboratories or other aspects of performance control of the individual MIR milk analyzers.

Between-Laboratory Performance: Producer vs. Modified Milk Calibration

S_R . The s_R values (after outlier removal) for each sample in each of 3 validation sample sets and a grand mean for fat, anhydrous lactose, true protein, and total solids for the modified milk calibration and producer milk calibration approaches are reported in Tables 5 and 6, respectively. A total of 7 single Grubbs outliers were identified and removed from the modified milk validation data (Table 5) and 53 single Grubbs outliers were removed from the producer milk validation data (Table 6). No double Grubbs outliers were detected. The large difference in number of statistical outliers between the 2 types of calibration sets underscores the susceptibility of instruments calibrated with producer milk samples to produce large variations on individual milk samples. The differences between modified milk calibration samples and producer milk calibration sample sets in the number of high leverage samples and

the size of calibration regression confidence intervals was presented previously (Kaylegian et al., 2006), and is the likely cause of the high number of validation outlier values for instruments using producer milk calibration.

Calibration of MIR analyzers with modified milk reduced the mean s_R by 46, 52, 61, and 55% for fat, lactose, true protein, and total solids, respectively, compared with calibration with producer milk (Table 7). Compared with the long term between-laboratory performance (i.e., s_R) of MIR analyzers calibrated with producer milks (Table 4), the modified milk calibration (Table 7) improved performance by reducing s_R from an average of 0.0231, 0.0202, and 0.0390 to 0.0155, 0.0109, and 0.0224 for fat, protein, and total solids, respectively. A useful form of expression of the s_R value for the analyst is to convert the s_R to the reproducibility value (R-value), which is calculated as $s_R \times 2.8$. The R-value indicates that 95% of the time the analysis of an unknown milk sample by 2 laboratories using the method (in this case, MIR) will not differ by more than the R-value, assuming the sample was at the correct temperature and properly mixed at the time of analysis. The overall mean R-values for all validation sets for fat, anhydrous lactose, true protein, and total solids for validation samples analyzed using the modified milk calibration were 0.043, 0.033, 0.030, and 0.063%, respectively, and for producer milk calibrated instruments, the mean R-values were 0.082, 0.071, 0.079, and 0.138%, respectively. These between-laboratory statistical performance values for a method are useful in setting practical guidelines for use in the verification of accuracy on individual validation samples.

Regulatory Verification of the Accuracy of Instrument Performance. The goal of regulatory verification of instrument performance is to detect when a milk payment testing instrument is out of compliance with a standard for accuracy on a set of unknown samples. When is the difference sufficiently large to warrant a required adjustment in instrument calibration? Small differences in milk component tests can have significant

Table 5. Reproducibility standard deviations (s_R) for validation samples (g/100 g) analyzed using mid-infrared instruments calibrated with modified milk samples (statistical outliers removed)

Sample	Validation set 1				Validation set 2				Validation set 3			
	Fat	Lactose ¹	True protein	Total solids	Fat	Lactose	True protein	Total solids	Fat	Lactose	True protein	Total solids
	s_R											
1	0.0187	0.0090	0.0146	0.0277	0.0164	0.0169	0.0086	0.0241	0.0152	0.0078	0.0106	0.0235
2	0.0118	0.0129	0.0117	0.0287	0.0177	0.0096	0.0055	0.0209	0.0129	0.0128	0.0170	0.0227
3	0.0219	0.0082	0.0083	0.0306	0.0063 ^a	0.0206 ^a	0.0064 ^a	0.0233	0.0083	0.0102	0.0125	0.0212
4	0.0110	0.0127	0.0084	0.0201	0.0167	0.0107	0.0064	0.0164	0.0092	0.0097	0.0062	0.0153
5	0.0147	0.0118	0.0182	0.0345	0.0145	0.0106	0.0229	0.0109 ^a	0.0157	0.0097	0.0118	0.0238
6	0.0138	0.0087	0.0070	0.0098	0.0187	0.0127	0.0078	0.0264	0.0126	0.0054	0.0113	0.0164
7	0.0064	0.0075	0.0171	0.0224 ^a	0.0143	0.0095	0.0069	0.0183	0.0132	0.0097	0.0066	0.0148
8	0.0142	0.0075	0.0056	0.0217	0.0114	0.0256	0.0044	0.0247	0.0121	0.0066	0.0095	0.0119
9	0.0227	0.0086	0.0148	0.0345	0.0145	0.0130	0.0093	0.0120	0.0155	0.0081	0.0102	0.0198
10	0.0186	0.0139	0.0134	0.0293	0.0139	0.0120	0.0087	0.0265	0.0280	0.0179	0.0165	0.0131 ^a
11	0.0108	0.0120	0.0083	0.0266	0.0153	0.0131	0.0077	0.0206	0.0284	0.0119	0.0078	0.0289
12	0.0232	0.0180	0.0192	0.0288	0.0260	0.0211	0.0163	0.0356	0.0138	0.0061	0.0127	0.0217
Mean	0.0157	0.0109	0.0122	0.0262	0.0155	0.0146	0.0093	0.0216	0.0154	0.0097	0.0111	0.0194

^aSingle Grubbs outliers removed.¹Lactose reported as anhydrous lactose in all validation sets.

economic impacts in payment testing (Lynch et al., 2004).

The current practice for both state regulatory agencies and USDA Federal Milk Market Laboratories is to have a laboratory test a set of verification samples with their MIR milk analyzer. The set of verification samples has been analyzed, usually by one regulatory laboratory, using chemistry methods. It is clear from the results in Table 8 that there is some degree of uncertainty when only one laboratory's chemistry results are used for reference. The regulatory agency calculates the MD and SDD between their reference chemistry values for the samples and the instrument values. If the MD ex-

ceeds a specified tolerance limit, then the laboratory may be required to adjust its instrument and possibly some of the past test results. Are there practical approaches to MIR milk analyzer calibration and verification that would allow the industry to achieve improved accuracy without excessive cost?

One approach to minimize the impact of laboratory-to-laboratory variation in reference chemistry is to have a set of validation samples with all-laboratory mean reference chemistry for each sample (Table 1). However, if every USDA Federal Milk Market Laboratory or state regulatory laboratory had to produce milk sample sets for instrument accuracy verification that required a

Table 6. Reproducibility standard deviation (s_R) for validation samples (g/100 g) analyzed using mid-infrared instruments calibrated with producer milk samples (statistical outliers removed)

Sample	Validation set 1				Validation set 2				Validation set 3			
	Fat	Lactose ¹	True protein	Total solids	Fat	Lactose	True protein	Total solids	Fat	Lactose	True protein	Total solids
	s_R											
1	0.0275	0.0263 ^a	0.0476	0.0354 ^a	0.0422 ^a	0.0289 ^a	0.0214	0.0463	0.0326	0.0299 ^a	0.0276	0.0500
2	0.0229	0.0165 ^a	0.0249	0.0931	0.0286 ^a	0.0164 ^a	0.0251	0.0383	0.0535	0.0763 ^a	0.0839	0.0498 ^a
3	0.0142	0.0173 ^a	0.0187	0.0335 ^a	0.0350	0.0178 ^a	0.0238	0.0396	0.0314	0.0276 ^a	0.0164	0.0334
4	0.0238	0.0157 ^a	0.0178	0.0396 ^a	0.0311 ^a	0.0206 ^a	0.0238	0.0238 ^a	0.0298	0.0253 ^a	0.0254	0.0470
5	0.0220	0.0175 ^a	0.0271	0.0374 ^a	0.0398	0.0250 ^a	0.0493	0.0438	0.0265 ^a	0.0302 ^a	0.0106	0.0433 ^a
6	0.0258	0.0200 ^a	0.0282	0.0460 ^a	0.0289	0.0237 ^a	0.0315	0.0501	0.0180	0.0200 ^a	0.0152	0.0345
7	0.0216	0.0238 ^a	0.0318	0.0401 ^a	0.0287	0.0170 ^a	0.0334	0.0474	0.0265	0.0217 ^a	0.0116	0.0453
8	0.0209	0.0224 ^a	0.0208	0.0967	0.0320	0.0237 ^a	0.0313	0.0517	0.0336	0.0286 ^a	0.0267	0.0495
9	0.0196	0.0156 ^a	0.0193	0.0365 ^a	0.0284	0.0242 ^a	0.0351	0.0553	0.0319	0.0246 ^a	0.0242	0.0482
10	0.0219	0.0206 ^a	0.0281	0.0386 ^a	0.0267	0.0199 ^a	0.0281	0.0487	0.0301	0.0377 ^a	0.0331	0.0692
11	0.0295	0.0164 ^a	0.0165	0.0414 ^a	0.0278	0.0277 ^a	0.0393	0.0601	0.0218	0.0227 ^a	0.0175	0.0454
12	0.0537	0.0482 ^a	0.0404	0.0951	0.0393	0.0397 ^a	0.0419	0.0688	0.0250	0.0168 ^a	0.0208	0.0514
Mean	0.0253	0.0217	0.0268	0.0528	0.0324	0.0237	0.0320	0.0477	0.0301	0.0301	0.0261	0.0473

^aSingle Grubbs outliers removed.¹Lactose reported as anhydrous lactose in all validation sets.

Table 7. Comparison of reproducibility standard deviations (s_R) for validation samples (g/100 g) analyzed using mid-infrared instruments calibrated with modified milk and producer milk samples

Replicate	Fat	Lactose ¹	True protein	Total solids	Fat	Lactose	True protein	Total solids
	s_R for modified milk calibration				s_R for producer milk calibration			
1	0.0157	0.0109	0.0122	0.0262	0.0253	0.0217	0.0268	0.0528
2	0.0155	0.0146	0.0093	0.0216	0.0324	0.0237	0.0320	0.0477
3	0.0154	0.0097	0.0111	0.0194	0.0301	0.0301	0.0261	0.0473
Mean	0.0155	0.0119	0.0109	0.0226	0.0294	0.0254	0.0284	0.0493
	Reduction of s_R ²				Reduction of s_R , %			
1	-0.0096	-0.0108	-0.0146	-0.0266	38.1	49.8	54.4	50.3
2	-0.0169	-0.0091	-0.0228	-0.0261	52.2	38.3	71.1	54.6
3	-0.0147	-0.0205	-0.0150	-0.0278	48.8	67.9	57.5	58.9
Mean	-0.0137	-0.0135	-0.0175	-0.0268	46.4	52.0	61.0	54.6

¹Lactose reported as anhydrous lactose in all validation sets.²Reduction = s_R for producer milk calibration – s_R for modified milk calibration.

network of laboratories to run chemistry on all samples, the cost would be high. The USDA Federal Milk Market Administrator Laboratories maintain and calibrate MIR milk analyzers in their own laboratories for testing. Clearly, as a group of laboratories, the data in Kaylegian et al. (2006) and the validation presented in Figures 1 to 4 and Table 3 indicate that a single common set of modified milk calibration samples with all-laboratory mean chemistry would allow this group of laboratories to achieve closer agreement with each other on MIR milk analysis and better agreement of their MIR milk analysis with reference chemistry methods than the same laboratories can agree with each other's chemistry on these samples. However, these laboratories still have a need to use producer milk samples (with established reference values) in the field to verify the accuracy of performance of instruments in industry laboratories. Currently, most regulatory laboratories would test the validation samples that they make in their

laboratory using their own chemistry. The level of laboratory-to-laboratory uncertainty (with and without outliers removed) in the mean reference chemistry values for 3 different sets of 12 verification samples if each laboratory was running chemistry is shown in Table 8. If only one laboratory was running chemistry on these validation samples, then the means without outliers removed are the values of interest. For example, in validation set 3 for fat (Table 8), a single laboratory running reference chemistry on these samples and using them for validation of instrument performance in the field could have had a mean fat test for the set that was 0.074% lower (3.9542 vs. 3.8802) than the all-laboratory mean and they would not realize they were low. One cost-effective strategy to reduce the risk of this happening in the production and use of validation samples is as follows: use a group (i.e., 8 or more) of MIR milk analyzers that have been calibrated with modified milk samples with all-laboratory mean chem-

Table 8. Comparison of all-laboratory mean chemistry and single laboratory reference chemistry values for the validation sets

	Validation set 1				Validation set 2				Validation set 3			
	Fat	Lactose ¹	True protein	Total solids	Fat	Lactose	True protein	Total solids	Fat	Lactose	True protein	Total solids
	%											
All-lab mean	4.0839	4.5103	3.2784	13.0175	3.9799	4.5692	3.2300	12.9098	3.9542	4.5626	3.1914	12.8470
Single lab with outliers removed ²												
Highest	4.0949	4.5502	3.2982	13.0339	3.9954	— ³	3.2494	12.9320	3.9690	4.5794	3.2096	12.8547
Lowest	4.0782	4.4727	3.2661	13.0007	3.9732	— ³	3.2228	12.8954	3.9409	4.5632	3.1845	12.8316
Range	0.0167	0.0775	0.0321	0.0332	0.0222	— ³	0.0266	0.0366	0.0281	0.0162	0.0251	0.0231
Single lab with no outliers removed												
Highest	4.0949	4.5502	3.2982	13.0339	3.9954	4.6072	3.2494	12.9343	3.9690	4.6104	3.2096	12.8709
Lowest	4.0769	4.4727	3.2661	12.9977	3.9732	4.5363	3.2010	12.8835	3.8802	4.5337	3.1557	12.7981
Range	0.0180	0.0775	0.0321	0.0362	0.0222	0.0709	0.0484	0.0507	0.0888	0.0767	0.0539	0.0728

¹Lactose reported as anhydrous lactose in all validation sets.²Single Grubbs outlier removed.³Only 1 out of 7 laboratories had no outlier values.

Table 9. All-laboratory mean values (g/100 g) for chemistry and instrument (modified milk calibration) prediction of fat, true protein, anhydrous lactose, and total solids of validation samples¹

Component	Validation set	All-lab mean		Difference ²	LSD ³
		Chemistry	Instrument		
		%			
Fat	1	4.0839 ^a	4.0760 ^b	−0.0079	0.0038
	2	3.9799 ^a	3.9634 ^b	−0.0165	0.0042
	3	3.9542 ^b	3.9607 ^a	0.0065	0.0040
	Mean	4.0060	4.0000	−0.0060	
Anhydrous lactose	1	4.5103	4.5159	0.0056	0.0061
	2	4.5692	4.5649	−0.0043	0.0068
	3	4.5626 ^b	4.5797 ^a	0.0171	0.0044
	Mean	4.5474	4.5535	0.0061	
True protein	1	3.2784 ^a	3.2684 ^b	−0.0100	0.0038
	2	3.2300	3.2280	−0.0020	0.0048
	3	3.1914	3.1874	−0.0040	0.0045
	Mean	3.2333	3.2279	−0.0053	
Total solids	1	13.0175 ^a	12.9936 ^b	−0.0239	0.0062
	2	12.9098 ^a	12.8893 ^b	−0.0205	0.0064
	3	12.8470 ^a	12.8393 ^b	−0.0077	0.0051
	Mean	12.9248	12.9074	−0.0174	

^{a,b}Chemistry and instrument means in the same row with different superscripts are different ($P < 0.05$).

¹Instrument mean values determined for all components by 9 instruments.

²Difference = Instrument mean value – chemistry mean value.

³LSD = Least significant difference ($P < 0.05$).

istry reference values to test each validation sample to establish an all-laboratory MIR instrument mean reference value (with statistical outliers removed) for each validation sample instead of chemistry. This approach eliminates the uncertainty of using the chemistry from a single laboratory for validation samples without having all laboratories run chemistry tests (and incurring the chemical analysis cost) on these validation samples. Next, we will explore the feasibility of this unconventional approach using the data we have collected.

All-Laboratory Mean Reference Chemistry vs. Instrument-Predicted Reference Values

The all-laboratory mean reference chemistry values and the all-laboratory instrument mean values (calculated from instruments calibrated with modified milk samples) for the validations sets used in this study are shown in Table 9. The mean difference between all-laboratory mean instrument and reference chemistry values across the 3 validation sets was approximately ± 0.0060 for fat, anhydrous lactose, and true protein, and -0.0174 for total solids (Table 9). The instrument all-laboratory mean was a much better predictor of all-laboratory mean reference chemistry on the validation samples (Table 9) than many individual laboratory ref-

erence chemistry values (Table 8). There were statistically significant differences ($P < 0.05$) between the all-laboratory mean instrument value and the all-laboratory mean reference chemistry value for some components and validation sets (Table 9). However, the least significant difference values were small ($<0.0068\%$), and this is better agreement with the all-laboratory mean chemistry than that achieved by comparison of most individual laboratory's chemistry with the all-laboratory mean chemistry (Table 8). From a practical point of view, the substitution of all-laboratory instrument mean value for individual laboratory reference chemistry on producer milk validation samples could improve regulatory verification of instrument accuracy for fat, lactose, protein, and total solids testing in a more cost-effective manner than obtaining all-laboratory mean chemistry values on producer milk validation sample sets used in various regions of the country. A more costly, but more conventional and acceptable, approach from a regulatory perspective would be to have a common set of calibration samples and a common set of validation samples used by all regulatory laboratories with all-laboratory mean chemistry for values for both sample sets.

CONCLUSIONS

Calibration of MIR milk analyzers using modified milk samples increased the accuracy of testing and im-

proved agreement between laboratories on independent sets of validation milk samples compared with MIR analyzers calibrated with producer milk samples. Calibration with modified milk samples reduced the mean ED from all-laboratory mean reference chemistry on validation samples by 40, 25, 36, and 27%, respectively, for fat, anhydrous lactose, true protein, and total solids. The number of single Grubbs statistical outliers in the validation data was much higher for instruments calibrated with producer milk samples, 53 vs. 7 for instruments calibrated with modified milk samples. The s_R for instruments calibrated with producer milk samples (with statistical outliers removed) was similar to data collected in recent proficiency studies conducted by the USDA Federal Milk Market Laboratories, as expected, whereas the s_R for instruments calibrated with modified milk samples was lower than those calibrated with producer milk samples by 46, 52, 61, and 55%, respectively for fat, anhydrous lactose, true protein, and total solids.

ACKNOWLEDGMENTS

The authors would like to thank the staff of all the USDA Federal Milk Market laboratories and affiliated laboratories for their collaboration and sample analysis in this work. The technical assistance in sample preparation by Maureen Chapman, Laura Landolf, Bob Kallaler, Mark Schweisthal, and Pat Wood was important for the success of this project. The authors thank the Test Procedures Committee of the USDA, Dairy Programs, Federal Milk Markets for their financial support of this research.

REFERENCES

- Association of Official Analytical Chemists International (AOAC). 2000. *Official Methods for Analysis*. AOACI, Gaithersburg, MD.
- Barbano, D. M., and J. L. Clark. 1989. Infrared milk analysis – Challenges for the future. *J. Dairy Sci.* 72:1627–1636.
- Barbano, D. M., J. L. Clark, and C. E. Dunham. 1988. Comparison of the Babcock and ether extraction methods for determination of fat content in milk: Collaborative study. *J. AOACI* 71:898–914.
- Barbano, D. M., J. L. Clark, C. E. Dunham, and J. R. Fleming. 1990. Kjeldahl method for determination of total nitrogen content of milk: Collaborative study. *J. AOACI* 73:849–859.
- Barbano, D. M., J. L. Lynch, and J. R. Fleming. 1991. Direct and indirect determination of true protein content of milk by Kjeldahl analysis: Collaborative study. *J. AOACI* 74:281–288.
- Biggs, D. A., G. Johnsson, and L.-O. Sjaunja. 1987. Analysis of fat, protein, lactose, and total solids by infra-red absorption. Pages 21–30 in *Monograph on Rapid Indirect Methods for Measurement of the Major Components of Milk*. Bull. Int. Dairy Fed. No. 208. International Dairy Federation, Brussels, Belgium.
- Clark, J. L., D. M. Barbano, and C. E. Dunham. 1989a. Comparison of two methods for determining total solids content of raw milk: Collaborative study. *J. AOACI* 72:712–718.
- Clark, J. L., D. M. Barbano, and C. E. Dunham. 1989b. Combination of total solids determined by oven drying and fat determined by Mojonnier extraction for measurement of solids-not-fat content of raw milk: Collaborative study. *J. AOACI* 72:719–724.
- Ginn, R. E., and V. S. Packard. 1989. A study of the accuracy of infrared milk component analysis in DHIA laboratories. *Dairy Food Environ. Sanit.* 9:61–64.
- Kaylegian, K. E., G. E. Houghton, J. M. Lynch, J. R. Fleming, and D. M. Barbano. 2006. Calibration of infrared milk analyzers: Modified milk versus producer milk. *J. Dairy Sci.* 89:2817–2832.
- Lynch, J. M. 1998. Use of AOAC International method performance statistics in the laboratory. *J. AOACI* 81:679–684.
- Lynch, J. M., and D. M. Barbano. 1999. Kjeldahl nitrogen analysis as a reference method for protein determination in dairy products. *J. AOACI* 82:1389–1398.
- Lynch, J. M., D. M. Barbano, and J. R. Fleming. 1990. Variation in the ash and nonprotein nitrogen content of milk, and use of milk protein content to predict ash content. *J. Dairy Sci.* 73(Suppl. 1):92. (Abstr.)
- Lynch, J. M., D. M. Barbano, and J. R. Fleming. 1996. Comparison of Babcock and ether extraction methods for determination of fat content of cream: Collaborative study. *J. AOACI* 79:907–916.
- Lynch, J. M., D. M. Barbano, and J. R. Fleming. 1997a. Modification of Babcock method to eliminate fat testing bias between the Babcock and ether extraction methods (modification of AOAC official methods 989.04 and 995.18): Collaborative study. *J. AOACI* 80:845–859.
- Lynch, J. M., D. M. Barbano, and J. R. Fleming. 1998. Indirect and direct determination of the casein content of milk by Kjeldahl nitrogen analysis: Collaborative study. *J. AOACI* 81:763–774.
- Lynch, J. M., D. M. Barbano, J. R. Fleming, and D. Nicholson. 2004. Component testing, the dairy industry, and AOAC International. *Inside Laboratory Management*, a publication of AOACI July/Aug:25–28.
- Lynch, J. M., D. M. Barbano, P. A. Healy, and J. R. Fleming. 1994. Performance evaluation of the Babcock and ether extraction methods: 1989 through 1992. *J. AOACI* 77:976–981.
- Lynch, J. M., D. M. Barbano, P. A. Healy, and J. R. Fleming. 1997b. Performance evaluation of direct forced-air total solids and Kjeldahl total nitrogen methods: 1990 through 1995. *J. AOACI* 80:1038–1043.
- Lynch, J. M., D. M. Barbano, P. A. Healy, and J. R. Fleming. 2003. Effectiveness of temperature modification in decreasing the bias in milk fat test results between the Babcock and ether extraction methods. *J. AOACI* 86:768–774.
- Lynch, J. M., D. M. Barbano, G. E. Houghton, and J. R. Fleming. 1995. Babcock bottle certification apparatus: Performance evaluation. *J. AOACI* 78:463–471.
- Lynch, J. M., D. M. Barbano, M. Schweisthal, and J. R. Fleming. 2006. Precalibration evaluation procedures for mid-infrared milk analyzers. *J. Dairy Sci.* 89:2761–2774.
- Massart, D. L., B. G. M. Vandeginste, S. M. Deming, Y. Michotte, and L. Kaufman. 1988. Clustering techniques. Pages 371–375 in *Chemometrics: A textbook, Data Handling in Science and Technology*. Vol. 2. Elsevier, New York, NY.
- New York State Department of Agriculture and Markets. 2005. Pages 1 to 5 in *Bulletin: Electronic and Other Methods of Testing for Component Content*. NYS Department of Agriculture and Markets. Division of Milk Control, 1 Winners Circle. Albany, NY.
- Wisconsin Administrative Code. 2005. *Dairy Plants*. Chapter 80 in *Agric., Trade & Consumer Protection*. Wisconsin State Register. <http://www.legis.state.wi.us/rsb/code/atcp/atcp080.pdf>