SUPPLEMENTARY INFORMATION A – DAIRY HERD TESTING EQUIPMENT APPROVAL PROCEDURES AND PRODUCTION DATA STANDARDS (Informative)

The supplementary information presented in this section includes:

- An overview (SA1) of the background and rationale to the approach taken to the herd testing standards described in Appendix A including examples for aggregating data;
- Guidelines that may assist with obtaining data for assessing devices (SA2);
- Accuracy requirements including examples of graphs required when presenting information (SA3)
- Guidelines that may assist identifying cow specific bias when aggregating data (SA4); and
- Guidelines that summarise the expectations for regulatory compliance as outlines by the Ministry for Primary Industries (SA5).

SA1 - Overview

Appendix A of the New Zealand standard: NZS 8100:2015 *Dairy herd testing* provides the performance standards for devices used in herd testing. Herd testing underpins two key purposes simultaneously. The first purpose relates to sires. The data are used to enable research and animal evaluation of the genetic productive potential of the national herd for the benefit of all New Zealand dairy farmers. The second purpose relates to cows. The farmers use the herd test data to monitor individual cows in their herd for cow performance as well as for management decisions. The former purpose currently 'piggy backs' on the farmer testing. The accuracy requirements need to be cognisant of both purposes.

The Herd Test Standards review committee agreed to retain the accuracy requirements from the New Zealand standard: NZS 8100:2007 for milk yield in the New Zealand standard: NZS 8100:2015 (Appendix A) while further research is undertaken to assess the AE Evaluation requirements. The stated intent of the Herd Test Standards review committee is to move toward the ICAR standard where the accuracy for milk meters is stated as a bias and standard deviation as separate components for accuracy. The decision for milk yield is pending and is expected to be reviewed by June 2017.

The accuracy requirements for milk fat concentrations from the ICAR guidelines (2012) have been adopted in the New Zealand standard: NZS 8100:2015 for milk meters and PIME and for aggregated data. However ICAR guidelines (2012) are silent on the accuracy requirements for protein, lactose, and somatic cell counts for milk meters used for single measurements. It is hypothesised that this has occurred because historically, the milk components were all sub-sampled together. Under such circumstances, the accuracy was focused on the most variable component, milk fat. With the changes in technology milk components can now be determined independently of each other. Also in New Zealand milk composition components are strongly influenced by the seasonal nature of a pasture based farm system combined with feed inputs, which vary significantly, when minimising feed deficits. It is therefore important that, in addition to milk fat, accuracies are set for milk protein, and somatic cell counts, at levels consistent with current technologies, to ensure that existing assumed accuracies are retained or improved. The committee has agreed that somatic cell count accuracy limits will be set by June 2017 following further research to assess the AE Evaluation requirements.

ICAR guidelines (2012) are described for PIME devices. However the review committee found the information ambiguous. ICAR guidelines (2012) assume data aggregation when the variation is random and that there is no cow specific bias. It appears that ICAR guidelines (2012) sets accuracy limits for a single measurement by an in-line sensor before aggregation (section 11.3.7 Table 11.2b),

but not following aggregation. The review committee have set accuracy requirements for milk yield and milk fat for aggregated data. They have also set provisional accuracy requirements for milk protein which will be reviewed and set by June 2017, following further research to assess the AE Evaluation requirements.

The calculation methods for aggregating observational data collected on-farm are described in the New Zealand standard: NZS 8100:2015 *Dairy herd testing* Appendix A. Briefly, observation data collected from consecutive milking on-farm are used to calculate the 24-hour average for each cow for each day and then the aggregated data for each cow in the aggregation period. When aggregating data, the milk composition results are weighted according to the yield for the milk collection from which the sample was taken. The aggregation data collection periods for the device and the reference shall be the same. An example for estimating aggregated data for milk components milk fat and milk protein for one cow for batch milk is shown in Table SA1 and for distributed milking in Table SA2.

Table SA1 - An example of estimating aggregated data for milk components, milk fat and milk protein for observational data collected on-farm from consecutive milking for one cow (ID 1001) for a two day Aggregation Period

Cow ID	Milking	Date and Time of milking	Comments	Milk volume recorded (kg)	Laboratory results		Calculated 24h estimates yields	
					Milk fat (%)	Milk protein (%)	Milk fat	Milk protein
1001	1	21/02/13 15:00	pm milking	10.8	5.30	3.80	0.572 (10.8 x 5.30%)	0.410 (10.8 x 3.80%)
1001	2	22/02/13 06:13	am milking	29.1	4.10	3.70	1.193 (29.1 x 4.10%)	1.076 (29.1 x 3.70%)
1001			Daily Average[1]	39.9	4.42 (1.765/39.9)*100	3.72 (1.486/39.9)*100	1.765 (0.572+1.193)	1.486 (0.410 + 1.076)
1001	3	22/02/13 15:00	pm milking	11.8	5.40	3.68	0.637 (11.8 x 5.40%)	0.434 (11.8 x 3.68%)
1001	4	23/02/13 06:00	am milking	26.8	4.10	3.60	1.099 (26.8 x 4.10%)	0.965 (26.8 x 3.60%)
1001			Daily Average[1]	38.6	4.50 (1.736/38.6)*100	3.62 (1.399/38.6)*100	1.736 (0.637+1.099)	1.399 (0.434+0.965)
The Aggregate Average for milk yield and milk fat and milk protein concentrations for the two day period is calculated using the Daily Average as shown below								
1001			Aggregated average	39.3 (39.9+38.6)/2	4.46 (4.42+4.50)/2	3.67 (3.72+3.62)/2		
The data is submitted to the core database with the Test Period date as 21/02/13 with half of the volume entered as PM and half as AM.								
Please Note: The Aggregate average for milk yield and milk fat and milk protein concentrations for the two day period **is <u>not</u> calculated** by averaging all the milk component concentrations as shown below.								
1001					4.73 (5.30+4.10+5.40+4.10)/4	3.70 (3.80+3.70+3.68+3.60)/4		

Table SA2 - An example of estimating standardised 24h yield for milk volume, fat and milk protein for one cow (ID 1001) with three complete milkings within a 48-hour Test Period.

Cow ID	**Milking**	**Date and Time of milking**	**Milking interval recorded by management system in hours (*in days*)**	**Milk volume recorded by PIME (kg)**	**Laboratory results**		**Calculated 24h estimates yields**	
					Milk fat (%)	**Milk protein (%)**	**Milk fat**	**Milk protein**
1001	-1	21/02/13 09:24						
Start of a 48h sampling period at 12:00 noon								
1001	1	21/02/13 22:56	13.53 *(0.5639)*	18.8	5.3	3.8	0.996 (18.8 x 5.3%)	0.71 (18.8 x 3.8%)
1001	2	22/02/13 16:13	17.28 *(0.7201)*	29.1	5.1	3.7	1.48 (29.1 x 5.1%)	1.08 (29.1 x 3.7%)
1001	3	23/02/13 11:08	18.92 *(0.7882)*	34.2	5.4	3.8	1.85 (34.2 x 5.4%)	1.30 (34.2 x 3.8%)
End of a 48h sampling period at 12:00 noon								
Total			49.73 *(2.0722)*	82.1			4.3272	3.0907
Standardized 24h yield and composition (calculated)				39.62 (82.1/2.0722)	5.27 (2.088/39.62)	3.76 (1.491/39.62)	2.088 (4.327/0.0722)	1.491 (3.091/2.0722)

SA2 - Guidelines that may assist with obtaining data for evaluating devices

When a herd test is being completed by a Certified Herd Tester an underlying assumption being made is that all the individual milk meters are working equally as well as each other (that is, that each individual device is meeting the performance standard). Consequently it does not matter which cow goes to which device. The same principle is operating when the devices are assessed against the standard outlined in Appendix A.

The following guidelines are offered to assist the applicant with obtaining the data required to assess the device.

The following guidelines are offered as suggestions only. They may assist the applicant with obtaining data that covers the ranges of milk yield and milk component concentrations under which the devices would be expected to operate. The guidelines are suggested to assist with:

(a) Targeting the data collection to reduce the number of reference samples that need to be collected.
(b) Obtaining data over the range of milk yields, milk component measurements (somatic cell counts) and milk concentrations (milk fat and milk protein) required.

The guidelines are as follows:

(a) It is suggested that a conventional herd test is completed to enable a subset of animals with low and high milk yields, and varying concentrations of somatic cells to be identified for the assessment. Alternatively if PIME are available, these devices could be used for selecting cows with high or low milk yields, however somatic cells counts will not be known. Selecting low and high milk yields should assist with providing the range of milk volumes required as well as the range of milk concentrations.
(b) Selecting herds which are milking twice a day, with morning and afternoon milkings will also assist with attaining a wider range of milk yields and milk compositions.
(c) Including a Jersey breed as one of the herds in late lactation will further extend the range for both milk yield and milk component with lower milk yields.

The following suggestions are made to assist with ensuring that operator errors and laboratory repeatability does not impact on the assessment of the device.

Any differences in results will be a combination of the sub-sampling of the reference sample (that is, the sample expected to represent the true result) or device sample and the repeatability of measurement by the instrument in the laboratory These cannot be separated entirely but the replicated subsamples and repeat analysis completed by the laboratory will determined if the differences were due to sub-sampling techniques or measurement by the laboratory instrument. It is therefore suggested that:

(a) Multiple sub-samples (that is, three or four) are prepared from the reference sample and that all of these are submitted to the laboratory for testing. This approach ensures that the laboratory sample result is a good estimate of the entire reference sample. It will identify if there is a lack of repeatability by the technician preparing the samples. The samples can be retained after analysis and re-analysed through the same Milkoscan instrument if the data are variable. Reanalysis assists with elucidating laboratory analysis repeatability and repeatability by the technician of preparing the reference subsamples.
(b) Where the sample volume permits, duplicate samples from the sampling device are submitted, together with the reference samples, to an approved laboratory for analysis (that is, milk fat and protein concentrations and somatic cell counts). Where there is sufficient volume for only a single sample to be submitted from the device, reanalysis at the testing laboratory is suggested. Reanalysing the sample assists with assessing the repeatability of the laboratory result.

SA3 - Accuracy requirements

Revision of the New Zealand standard

The New Zealand standard and the ICAR guidelines (2012) express accuracy limits using different terminology. ICAR guidelines (2012) define accuracy in terms of separate values for standard deviation and bias. The New Zealand standard combined the standard deviation and mean bias and set accuracy limits whereby 95% of the data and 99% of the data shall fall between defined limits. This means that the New Zealand standard and the ICAR guidelines (2012) cannot be easily compared.

With regard to PIME and aggregating data, although the ICAR guidelines (2012) set accuracy limits for in-line analysers (PIME) the review committee found the information ambiguous.

The consensus of the review committee was that:

(a) having a New Zealand standard as well as international guidelines (that is ICAR) made it difficult for equipment manufactures to comply with effectively two standards. It was therefore agreed that the New Zealand standard should transition to the ICAR guidelines, or their equivalent.
(b) the standards adopted needed to be fit for purpose for New Zealand
(c) Accuracy requirements for milk meters and PIME with single and multiple measurements and for aggregated data should be set out separately in the standard
(d) accuracy limits for milk protein are required because of the;
    (i) high weighing in the AE model for milk protein
    (ii) variation in milk protein concentration present in the New Zealand dairy industry which results from the seasonality of the pastoral based farm system and the uniqueness these systems bring in terms of feed inputs, which may vary dramatically during the season.
(e) As existing milk meters could meet milk protein accuracy requirements that equated to the ICAR accuracy limits for a single measurement for milk fat (i.e. a standard deviation of 0.1% and a mean bias of 0.05), that these accuracy limits should be accepted for the New Zealand standard for milk protein.
(f) The accuracy requirements should follow the ICAR guideline approach with setting standard deviation and mean bias limits separately by 2020.

The review committee therefore agreed accuracy limits for milk meters and PIME used for single measurements (Appendix A; Tables A3 or A4 and A5) or for multiple measurements (Appendix A; Tables A3 or A6 and A5) or for aggregated data (Appendix A; Tables A3 and A7).

Examples of graphs required when presenting information

Graphs are required when presenting data as they provide a quick visual assessment of the data including the correlation between the device and reference data, whether the data are affected by systematic bias and/or heteroscedasticity. The graph shall show the device measurement versus the mean reference measurement. Figure SA1 shows an example of graph for milk yield data where the 1-1 line is included on the graph.

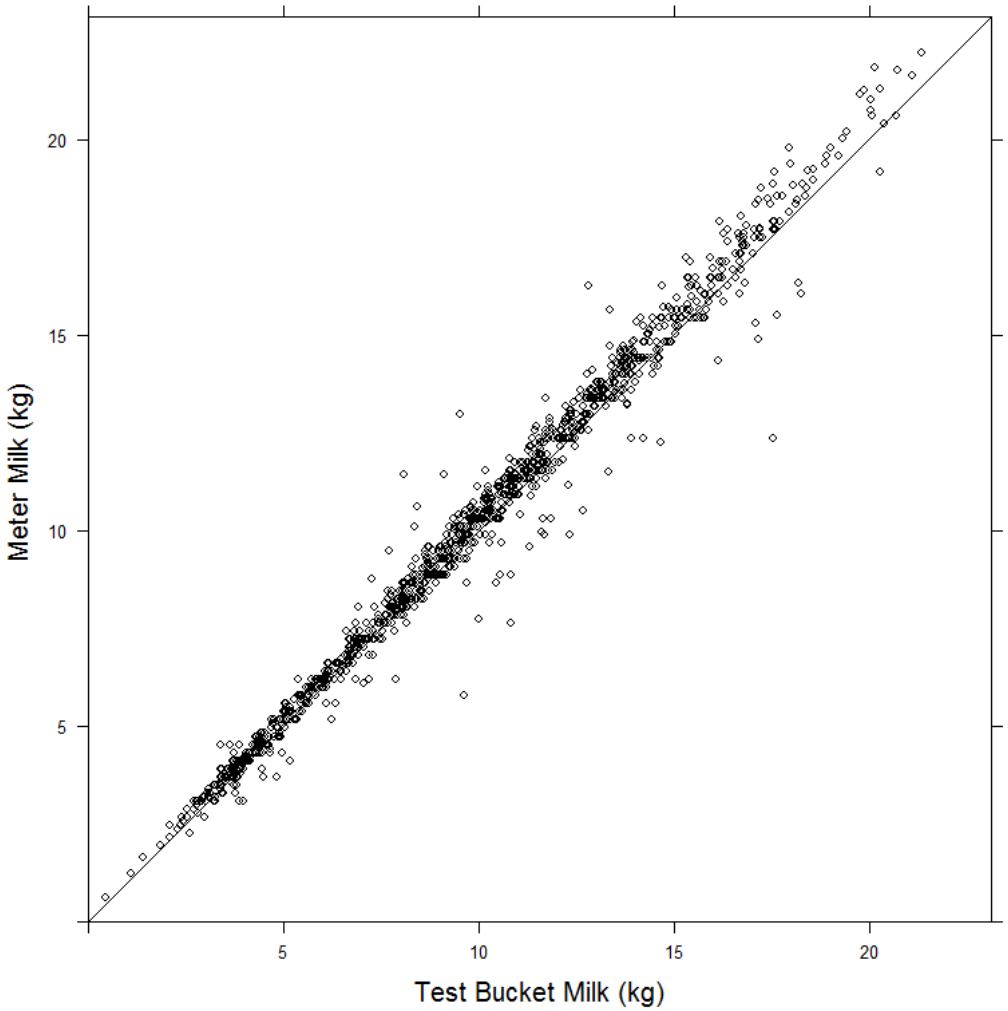


Figure SA1: An example showing the relationship between the results for milk yield derived from the device (i.e. milk meters) and the reference data, and including a 1-1 line.

Figures SA2 and SA3 show examples of a graph for milk yield showing the difference between device and mean reference sample (expressed as a percentage of reference for milk yield) vs. mean reference sample. Horizontal lines are included to indicate bounds within which 95% or 99% of the differences are required to lie for a device where a single measurement (Figure SA2) or multiple measurements (Figure SA3) are used to derive an accurate assessment of the milk yield that is representative for the entire milking event for an individual cow.

Figure SA2: An example of a graph showing the difference between device and mean reference sample (expressed as a percentage of reference for milk yield) vs. mean reference sample where the device is used for a single measurement. The horizontal lines are included to indicate bounds within which the milk yield data shall fall: 95% of the differences are required to fall between the dotted lines at ± 7.5% and 99% of the data are expected to fall between the dashed lines at ±10%.

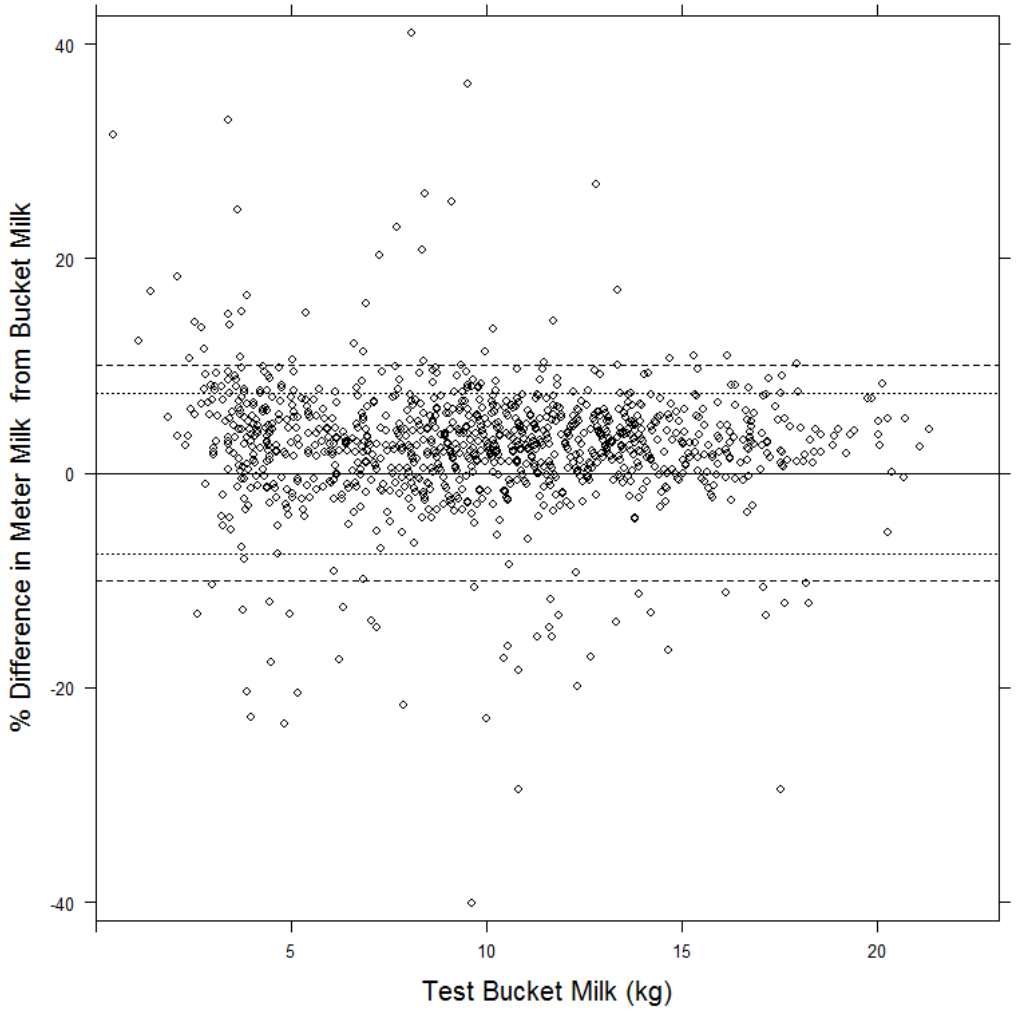
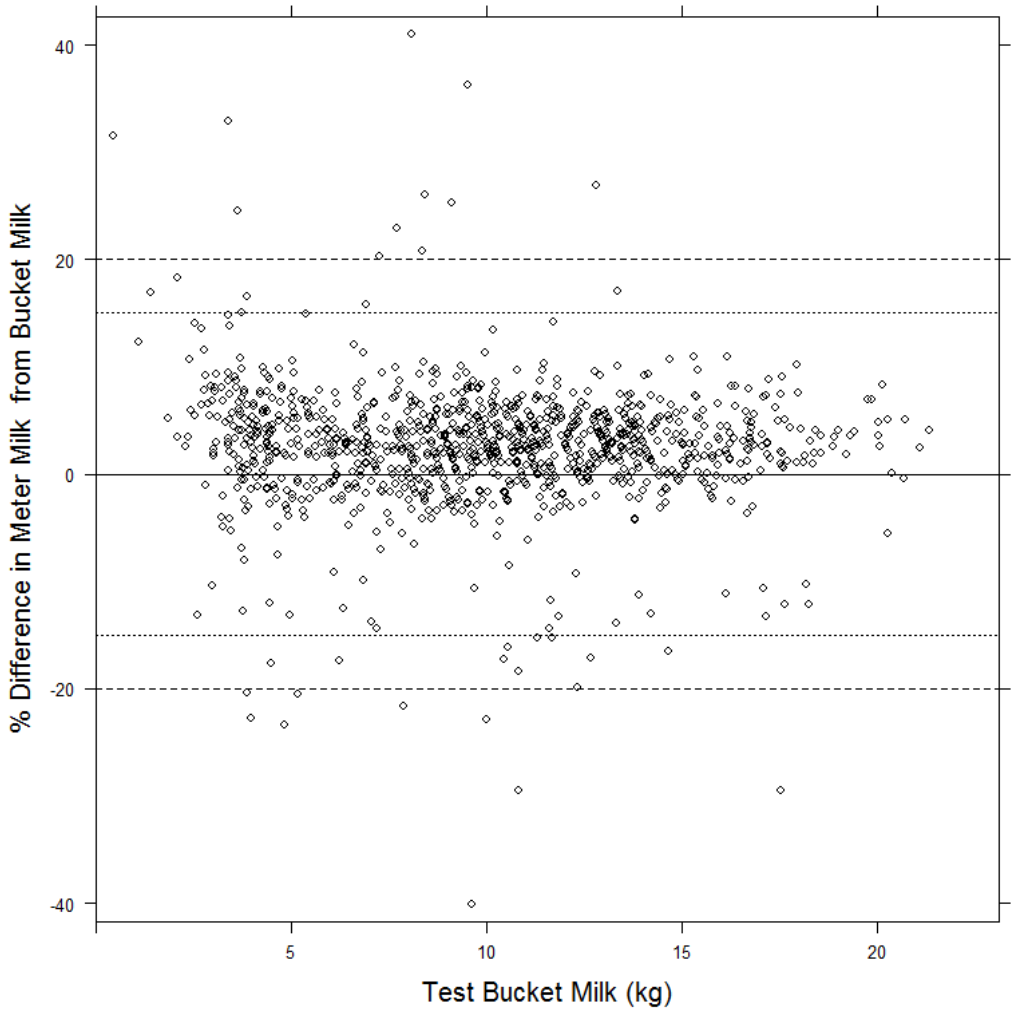


Figure SA3: An example of a graph showing the difference between device and mean reference sample (expressed as a percentage of reference for milk yield) vs. mean reference sample where the device is used for multiple measurements. The horizontal lines are included to indicate bounds within which the milk yield data shall fall: 95% of the differences are required to fall between the dotted lines at ± 15% and 99% of the data are expected to fall between the dashed lines at ±20%. In addition, the graph will show if the devices are likely to meet the bias limits of no more than 0.2 litres or 2% of the mean reference yield, whichever is greater.



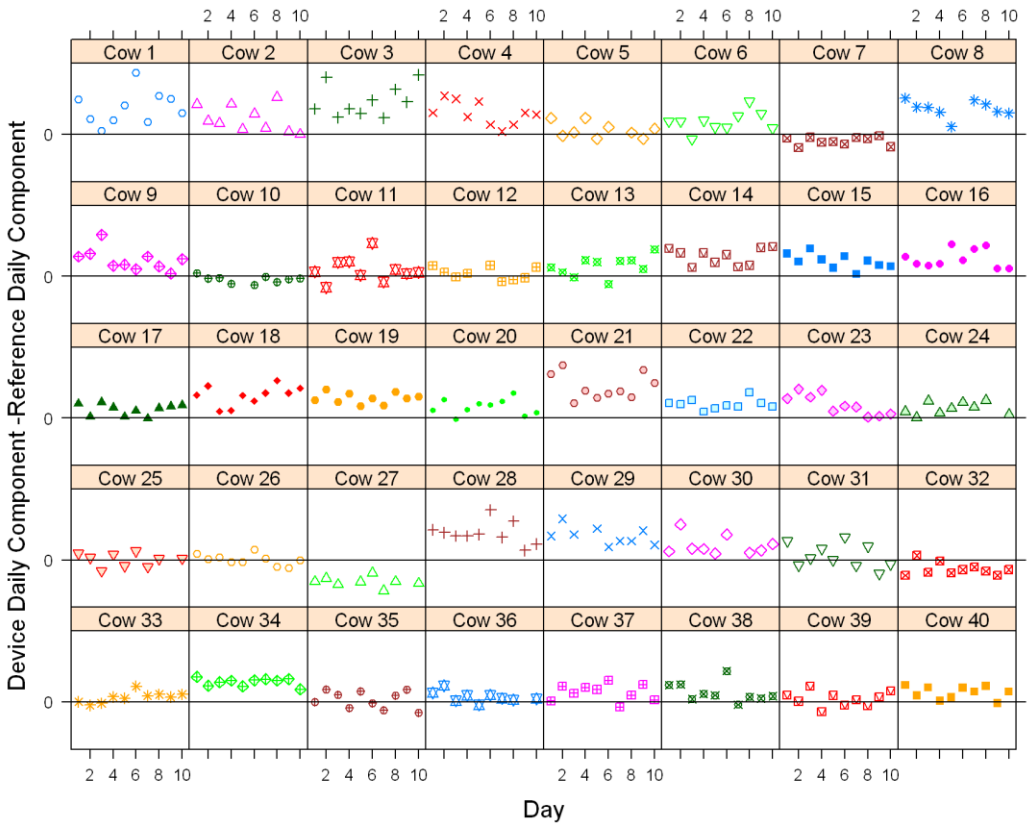
NOTE – The graphs are additional and do not replace the statistical evaluation comparing the predicted data compared with the reference data.

SA4 - Guidelines that may assist identifying cow specific bias when aggregating data

The following suggestions are made to assist the applicant when aggregating data does not improve the accuracy for the milk yield or a milk component. In such situation, it may be helpful for the applicant to assess data to see if cow specific bias is present and contributing to a loss in improvement in accuracy.

A simple visual check can be completed by plotting, for each individual cow, the differences between the predicted value and the reference value for the aggregation period, for the specific component. Figure SA4 shows an example of such a graph with the individual cow data plotted.

Figure SA4: An example of a graph showing the individual cow data where cow specific bias is present for some cows such as cow 27, 28 and 34 (i.e. the difference between the predicted values is consistently low or high relative to the reference data)



SB5 - Guidelines that summarise the expectations for regulatory compliance as outlined by the Ministry for Primary Industries

Ministry for Primary Industries
Manatū Ahu Matua

February 2015

New Zealand Standard NZS 8100 (Dairy Herd Testing): regulatory requirements for the standard

Purpose

This document outlines the requirements that the New Zealand Standard NZS 8100 (Dairy Herd Testing) must address that are identified under the Dairy Industry (Herd Testing and New Zealand Dairy Core Database) Regulations 2001 (**2001 Regulations**).

Requirements under the 2001 Regulations

The 2001 Regulations identify two clear requirements that the standard should address:

1. *Regulation 6(1)(b)* – the standard should specify the operating and accuracy standards that certified herd testers must meet for equipment and methods used for sampling, measuring, and analysing milk and supplying data to the manager of the core database; and

2. *Regulation 7(1)* – the standard should identify the data requirements that every certified herd tester must meet for the collection and supply of herd testing data to the manager of the core database, relating to the tester's activities, that are specified in Schedule 2 of the 2001 Regulations (*Information to be supplied to manager of the core database*).

Further regulatory provisions to consider

The regulations also provide the certification body must:

1. *"assess applications from persons who want to become certified herd testers against the requirements of the dairy herd testing standard"* (Clause 2(a) Schedule 1); and

2. *"not grant certification to an applicant unless the certification body is satisfied that the applicant has the necessary competencies, capacity, and capability to undertake herd testing in compliance with the dairy herd testing standard."* (Clause 4 Schedule 1).

The Ministry for Primary Industries (**Ministry**) considers, when read together, these provisions are capable of being read two ways:

1. The standard must clearly identify the competencies, capacity and capability that the certification body will use as criteria to determine whether an applicant should be certified as a certified herd tester; or

2. The standard must identify the standards and requirements that certified test herders must comply with in performing their roles as testers following certification.

The Ministry understands that in practice Telarc SAI Limited uses ISO 9001 to assess the competencies, capacity and capability of applicants. The standard could contain an explanatory note that says that the certification body uses ISO 9001 to determine whether an applicant meets the requirements to be certified as a herd tester. That would avoid any issues arising from the ambiguity in clauses 2(a) and 4 of Schedule 1.