



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Diagnosing intramammary infections: Evaluation of definitions based on a single milk sample

Citation for published version:

Dohoo, IR, Smith, J, Andersen, S, Kelton, DF, Godden, S, Zadoks, R & Mastitis Res Workers, C 2011, 'Diagnosing intramammary infections: Evaluation of definitions based on a single milk sample', *Journal of Dairy Science*, vol. 94, no. 1, pp. 250-261. <https://doi.org/10.3168/jds.2010-3559>

Digital Object Identifier (DOI):

[10.3168/jds.2010-3559](https://doi.org/10.3168/jds.2010-3559)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Publisher's PDF, also known as Version of record

Published In:

Journal of Dairy Science

Publisher Rights Statement:

© American Dairy Science Association®, 2011

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.





Diagnosing intramammary infections: Evaluation of definitions based on a single milk sample

I. R. Dohoo,*¹ J. Smith,* S. Andersen,* D. F. Kelton,† S. Godden,‡
and Mastitis Research Workers' Conference²

*Centre for Veterinary Epidemiological Research, Atlantic Veterinary College, University of Prince Edward Island, Charlottetown, Prince Edward Island, Canada C1A 4P3

†Ontario Veterinary College, University of Guelph, Guelph, Ontario, Canada N1G 2W1

‡College of Veterinary Medicine, University of Minnesota, St. Paul 27712

ABSTRACT

Criteria for diagnosing intramammary infections (IMI) have been debated for many years. Factors that may be considered in making a diagnosis include the organism of interest being found on culture, the number of colonies isolated, whether or not the organism was recovered in pure or mixed culture, and whether or not there was concurrent evidence of inflammation (often measured by somatic cell count). However, research using these criteria has been hampered by the lack of a “gold standard” test (i.e., a perfect test against which the criteria can be evaluated) and the need for very large data sets of culture results to have sufficient numbers of quarters with infections with a variety of organisms. This manuscript used 2 large data sets of culture results to evaluate several definitions (sets of criteria) for classifying a quarter as having, or not having an IMI by comparing the results from a single culture to a gold standard diagnosis based on a set of 3 milk samples. The first consisted of 38,376 milk samples from which 25,886 triplicate sets of milk samples taken 1 wk apart were extracted. The second consisted of 784 quarters that were classified as infected or not based on a set of 3 milk samples collected at 2-d intervals. From these quarters, a total of 3,136 additional samples were evaluated. A total of 12 definitions (named A to L) based on combinations of the number of colonies isolated, whether or not the organism was recovered in pure or mixed culture, and the somatic cell count were evaluated for each organism (or group of organisms) for which there were sufficient data. The sensitivity (ability of a definition to detect IMI) and the specificity (Sp;

ability of a definition to correctly classify noninfected quarters) were both computed. For all species, except *Staphylococcus aureus*, the sensitivity of all definitions was <90% (and in many cases <50%). Consequently, if identifying as many existing infections as possible is important, then the criteria for considering a quarter positive should be a single colony (from a 0.01-mL milk sample) isolated (definition A). With the exception of “any organism” and coagulase-negative staphylococci, all Sp estimates were over 94% in the daily data and over 97% in the weekly data, suggesting that for most species, definition A may be acceptable. For coagulase-negative staphylococci, definitions B (2 colonies from a 0.01-mL milk sample) raised the Sp to 92 and 95% in the daily and weekly data, respectively. For “any organism,” using definition B raised the Sp to 88 and 93% in the 2 data sets, respectively. The final choice of definition will depend on the objectives of study or control program for which the sample was collected.

Key words: intramammary infection, definition, sensitivity, specificity

INTRODUCTION

Worldwide, mastitis is one of the most economically important diseases in dairy production, with subclinical mastitis accounting for almost two-thirds of the economic loss (Seegers et al., 2003; Halasa et al., 2007). Subclinical mastitis implies inflammation within the udder, but not necessarily infection. The inflammatory reaction may be identified by an elevated SCC or other measure of inflammation (e.g., California mastitis test). However, subclinical mastitis is most often due to a bacterial IMI (Djabri et al., 2002), so the terms IMI and subclinical mastitis are often used interchangeably (Barkema et al., 1997; Deluyker et al., 2005). The term IMI does refer specifically to the presence of an infectious organism in the udder (Berry and Meaney, 2006). Given the central role of IMI in mastitis, good information on the operating characteristics, sensitivity

Received June 23, 2010.

Accepted October 1, 2010.

¹Corresponding author: dohoo@upei.ca

²Participants at 2008 Mastitis Research Workers' Conference are recognized for their contributions to the process and their endorsement of the views expressed in this manuscript. A complete list of participants is included in the Appendix.

(Se) and specificity (Sp), of procedures used to classify quarters as having an IMI or not is required for correct interpretation of test results.

Classifying Quarters with Regard to IMI

Historically, several definitions (or classification rules) of mastitis and IMI have been used. In 1987, the International Dairy Federation published definitions and guidelines for diagnosis of bovine mastitis, which covered both definitions for IMI and for mastitis in various situations (e.g., experimental research, observational studies, control programs; Griffin et al., 1987). Information considered in the guidelines included both culture results and measures of inflammation and discussion ensued about the number of samples that should be taken. However, no specific recommendations for classifying sample results were made. In the same year, the National Mastitis Council (NMC) published guidelines for classifying results from quarter milk samples: 1 = not significant, 2 = questionable significance, 3 = probable significance, or 4 = highly significant based on a combination of factors consisting of the species isolated, the number of colonies isolated, and whether the organism was isolated in a pure or mixed culture (National Mastitis Council, 1987). However, we frequently need to classify quarters as infected or not (when interpreting culture results in either clinical or research settings), so knowing the operating characteristics (Se and Sp) of the classification procedure is important for correct interpretation of the culture results. Such data were not available at the time of publication of the NMC guidelines.

Information that has gone into those definitions and guidelines has included some or all of the following: the presence of an organism of interest, the number of colonies of the organism grown (typically from a 0.01-mL sample of milk), whether the organism was isolated in pure or mixed culture, and some indicator of inflammatory response (perhaps SCC over a specified threshold). Single, duplicate, and triplicate quarter milk samples over various periods have been used to determine IMI status (Dingwell et al., 2003; Hillerton et al., 2007). In addition to variation in terms of which criteria are used to classify a quarter as having an IMI, there has also been variation in the thresholds used for some of those criteria. For example, Zadoks et al. (2001) used a minimum colony count of 1,000 cfu/mL when using single samples to determine infection status with *Streptococcus uberis*. When the SCC criterion was added to the definition, thresholds used have ranged from 100,000 to 300,000 cells/mL (Schukken et al., 2003; Bansal et al., 2005; Deluyker et al., 2005). With respect to the

number of organisms cultured in the samples, some researchers considered a sample contaminated if 3 or more species were cultured (Parker et al., 2008) and others did not place any restrictions on the number of bacterial species cultured (Berry and Meaney, 2006).

Procedures for Evaluating Diagnostic Tests

The ideal situation for evaluating a diagnostic test is to have a gold standard test to which the results from the diagnostic test of interest can be compared. For IMI, a gold standard would be a test (or combination of tests) that would correctly classify both infected and noninfected quarters 100% of the time. In reality, a perfect gold standard rarely exists for any condition. In the absence of a perfect gold standard, the options available for evaluating a diagnostic test include using an acceptable gold standard, while recognizing its limitations, comparing the test of interest to a reference test of known sensitivity and specificity, or using a latent class modeling approach, which does not assume that the accuracy of either the reference test or the test of interest is known (Dohoo et al., 2009).

No reference tests of IMI exist for which the sensitivity is known, so option 2 is not possible. Latent class modeling is based on several important assumptions, namely that the 2 (or more) tests being compared are biologically independent, that data are available from multiple populations with different prevalences, and that the Se and Sp of the tests are constant across those populations (Enøe et al., 2000). For tests for IMI, meeting the last 2 assumptions may be possible but the need to have 2 or more biologically independent tests is problematic. Culture procedures are generally accepted as the only method of reliably detecting IMI even though they may be combined with measures of inflammation.

Consequently, the only viable approach to evaluating various definitions for an IMI based on a single milk sample is to compare those tests to an acceptable gold standard, while recognizing the limitations and possible effect of any deficiencies in that gold standard. Triplicate milk samples are often considered to be the gold standard for testing quarters for IMI. A consensus agreement on how results from triplicate samples should be interpreted has recently been published (Andersen et al., 2010). Based on 3 consecutive samples, a quarter was considered to have an IMI with a specific pathogen on the middle sampling date if the organism was isolated with more than 1,000 cfu/mL (10 colonies from a 0.01-mL sample) from that middle sample, or if any 2 of the 3 samples were positive for the organism of interest.

Objectives

This study was conducted as the second step in a multi-part process with the overall goal of determining the operating characteristics of various definitions of IMI. The first step was to develop a consensus gold standard based on 3 consecutive milk samples (Andersen et al., 2010). This manuscript addresses the issue of diagnosing an IMI based on a single milk sample, whereas subsequent steps will include evaluating the merits of duplicate and triplicate sampling compared with basing a decision on a single sample, and evaluating the accuracy of a cow-level diagnosis based on a composite sample compared with a set of quarter samples.

The primary objective of this study was to evaluate a set of rules for classifying the infection status of an udder quarter based on a single milk sample, by comparing those classifications to a gold standard classification based on 3 consecutive milk samples. The evaluated rules used information about the organism isolated, whether or not the organism was isolated in pure or mixed culture, the number of organisms cultured, and the SCC of the quarter milk sample. The evaluation was carried out in 2 separate data sets, one in which the gold standard classification was based on 3 consecutive weekly samples and one in which it was based on 3 samples taken 2 d apart.

MATERIALS AND METHODS

Data Sets

Two data sets were used in this evaluation. The first, referred to as “weekly data,” was based on 3 consecutive quarter milk samples taken at weekly intervals. The second, referred to as “daily data,” was based on duplicate milk samples collected daily over a 5-d period.

Weekly Data. The weekly data were collected as part of the National Cohort of Dairy Farms data collection platform carried out by the Canadian Bovine Mastitis Research Network (CBMRN) during the summer of 2007. Details of the data collection and culturing procedures have been described previously (Reyher et al., in review). Briefly, samples were collected weekly for 7 wk from 15 cows in each study herd. Five of these cows were the most recently calved cows (i.e., early lactation), whereas the other 10 were randomly chosen from the lactating cows that were expected to remain in the milking herd for at least another 60 d. A set of 3 weekly samples made up a triplicate set so each quarter could potentially contribute 5 triplicate sets (e.g., wk 1, 2, and 3; wk 2, 3, and 4). Results from all of the samples in the triplicate set were used to classify the

cow as having or not having an IMI (i.e., considered the gold standard), whereas the results from the middle test day were evaluated to determine the ability of this single sample to correctly classify the quarter.

All milk samples were frozen for storage, thawed once, and cultured using standardized protocols based on the NMC guidelines for bacteriological culture and species identification (National Mastitis Council, 1999). Isolates were classified as belonging to 1 of 20 species (or groups of species—but, hereafter, referred to as species), although only the following 7 species were considered in the analyses for this paper because of insufficient data in other categories: any organism isolated (**ANY**); CNS; *Staphylococcus aureus*; *Corynebacterium* spp. (presumed to be primarily *C. bovis*); enterococci (the CBMRN recorded organisms other than *Strep. uberis*, *Strep. dysgalactiae*, and *Strep. agalactiae* as *Streptococcus* spp., but we assumed these isolates were primarily enterococci and they are referred to as such in this manuscript); *Streptococcus* spp. (consisting of *Strep. uberis* and *Strep. dysgalactiae*); and *Escherichia coli*. In fact, very few samples were positive for *E. coli*, but these were included in the analyses in order that one gram-negative, environmental pathogen was included.

Daily Data. The daily data were collected during the summer of 2003 from 9 herds at 4 different locations (Minnesota, n = 5; Prince Edward Island, n = 2; Ontario, n = 1; New York, n = 1). Duplicate quarter milk samples A and B were collected at morning milking for 5 consecutive days from a total of 197 cows, evenly distributed among fresh and mid- to late-lactation cows. Fresh cows were defined as being close to calving (1 to 5 DIM if possible). Mid- to late-lactation cows were defined as being at least 150 DIM at d 1 of sampling. The cows were randomly selected among those in the herd with no recent or concurrent clinical mastitis. Each cow's parity and DIM at the start of the sampling were recorded. Samples from d 1, 3, and 5 made up the triplicate set that was used to classify the quarter has having or not having an IMI with a specific pathogen for the 5-d period. Samples from d 2 and 4 (n = 4) were evaluated individually for their ability to correctly classify the quarter.

All samples were frozen for storage, thawed once, and cultured in university laboratories using standardized procedures consistent with NMC guidelines. Samples from the Canadian sites were cultured at the University of Guelph and samples from the US sites were cultured at the University of Minnesota. The presence of specific pathogens was confirmed and identified using the API identification system (BioMérieux, Durham, NC) in each laboratory. Isolates were classified into 1 of 21 different species or groups of species, although only the

Table 1. Definitions considered for classifying quarters as having or not having an intramammary infection with an organism of interest

Definition	Minimum number of colonies cultured (per 0.01 mL)	Pure or mixed growth ¹	Minimum SCC (cells/mL)
A	≥1	Mixed	No minimum
B	≥2	Mixed	No minimum
C	≥10	Mixed	No minimum
D	≥1	Pure	No minimum
E	≥2	Pure	No minimum
F	≥10	Pure	No minimum
G	≥1	Mixed	200,000
H	≥2	Mixed	200,000
I	≥10	Mixed	200,000
J	≥1	Pure	200,000
K	≥2	Pure	200,000
L	≥10	Pure	200,000

¹Mixed: culture could be mixed growth or pure culture; Pure: organism had to be grown in pure culture.

following 4 categories were considered in the analyses for this paper because of insufficient data in other categories: ANY, CNS, *Streptococcus* spp., and enterococci. *Streptococcus* spp. consisted primarily of *Strep. uberis*, with a moderate number of *Strep. dysgalactiae* and very few *Strep. agalactiae* and *Strep. bovis*.

Gold Standard Definition

The gold standard definitions used in the 2 data sets were based on the previously published consensus standard definition (Andersen et al., 2010). A sample was considered gold standard positive (GS+) for a given pathogen if the same pathogen was isolated from 2 of 3 samples that made up the triplicate set, or if it was isolated with more than 1,000 cfu/mL (10 colonies/0.01 mL) from the middle sample of the 3 samples. As noted above, for the weekly data, the sample from the middle week also served as the test day of interest. For the daily data, the gold standard classification was based on the samples from d 1, 3, and 5, whereas the test day of interest samples were those (n = 4) collected on d 2 and 4 of the sequence. Samples that were not GS+ were considered gold standard negative (GS-).

Single Sample Definitions

A set of 12 definitions based on a single sample were evaluated and these are described in Table 1. The definitions were based on whether or not the species of interest was isolated, the number of colonies observed, whether a pure or mixed culture was obtained, and the SCC of the quarter (above or below 200,000 cells/mL). Somatic cells counts were not consistently recorded for the daily data samples, so the 6 definitions that required, in addition to other criteria, that the SCC be

≥200,000 cells/mL were not evaluated using these data. Consequently, definition A was the most relaxed (i.e., easiest to classify a quarter as having an IMI), whereas definition L was the most restrictive. The selection of candidate definitions, the choice of gold standard, and the general approach to the analyses was reviewed and endorsed by the participants at the 2008 Mastitis Research Workers' Conference (see Appendix for list of participants).

Analyses

Descriptive Statistics and Unconditional Assocations. For each species in each data set, descriptive statistics, including the number of isolates that were GS+ and GS-, were computed. For each definition and each species of interest, the Se and Sp were computed (Dohoo et al., 2009). The confidence interval around these estimates was determined using formulae appropriate for binomial proportions. Plots of Se and Sp and their 95% confidence intervals from unconditional analyses were created for *Staph. aureus* and CNS from the weekly data and *Streptococcus* spp. and CNS from the daily data.

Random Effects Models. Given that observations (samples) were not independent, but rather were clustered within quarters (multiple observations per quarter), within cows (4 quarters per cow), and within herds, the confidence intervals estimated above were likely underestimates of the true confidence intervals for the Se and Sp. Consequently, random effects logistic regression analyses were carried out with random effects for herd, cow, and quarter. Separate analyses were carried out for GS+ and GS- sets of samples, with the test result being the outcome of interest. If models would not converge with all 3 random effects included, one

or more of the random effects was removed until the model converged. Estimates derived from these models were converted to marginal (or population average) estimates based on the total variance of the random effects (Dohoo et al., 2009). Plots comparing estimates of Se and Sp and their 95% confidence intervals from unconditional analyses and random effects models were created for *Staph. aureus* and CNS from the weekly data. All analyses were carried out using Stata, Version 11 (StataCorp, 2009).

RESULTS

Descriptive Statistics

Weekly Data. Culture results from 38,376 milk samples collected during the summer of 2007 were available for these analyses. From all species, except *Staph. aureus*, valid results were available for 18,420 triplicate sets of milk samples from 4,889 quarters in 1,330 cows in 90 herds. Because *Staph. aureus* was recorded even if the sample would otherwise be considered contaminated, 25,886 triplicate sets of samples (5,344 quarters, 1,342 cows, 90 herds) were available for analyses of definitions for classifying *Staph. aureus*. The number of triplicate sets considered GS+ and GS- for all species evaluated in this data set are presented in Table 2.

Daily Data. Data were obtained from 784 quarters in 196 cows in 9 herds. From each quarter, 4 observations were recorded (2 samples on d 2 and 2 from d 4), giving a possible total of 3,136 observations. Of these, 2,415 triplicate sets had usable data for all of the 4 species evaluated (Table 3). No other species had sufficient data to warrant estimation of Se and Sp.

Unconditional Estimates of Se and Sp

Weekly Data. Unconditional estimates of the Se and Sp of each of the definitions are presented in Table 2. With the exception of ANY and CNS, all estimates of Sp were very high (>97%), even for definition A. When isolation of a single CNS organism was considered positive, the specificity decreased to 86 and 91% for mixed and pure culture, respectively.

Estimates of Se for *Streptococcus* spp. ranged from 61 to 87%, whereas for *Staph. aureus*, the range was 44 to 90%. For all other species, estimates of Se were very low, with only definition A resulting in estimates of Se >75%. For the other 5 species, requiring 2 or more colonies to be cultured decreased the Se into the 60% range, whereas any other restrictions on the definition (e.g., requiring 10 or more colonies, requiring that the isolate be in pure culture, or requiring a minimum SCC) decreased the Se to very low levels. Graphs of

Table 2. Unconditional estimates of sensitivity (Se) and specificity (Sp) for several pathogens (or pathogen groups) with the gold standard based on 3 consecutive samples taken 1 wk apart and the estimates based on the middle sample¹

Organism	Any organism		CNS		Staphylococcus aureus		Corynebacterium spp.		Enterococci		Streptococcus dysgalactiae and uberis		Escherichia coli	
	Se	Sp	Se	Sp	Se	Sp	Se	Sp	Se	Sp	Se	Sp	Se	Sp
Gold standard negative	8,700	8,700	13,800	13,800	25,332	25,332	17,921	17,921	18,177	18,177	18,316	18,316	18,403	18,403
Gold standard positive	9,847	9,847	4,620	4,620	554	554	499	499	243	243	104	104	17	17
Total number of sets	18,547	18,547	18,420	18,420	25,886	25,886	18,420	18,420	18,420	18,420	18,420	18,420	18,420	18,420
Definition ²														
A	85.8	75.1	80.9	86.7	90.4	99.8	78.2	97.5	76.5	97.4	86.5	100.0	76.5	99.9
B	60.8	91.9	56.2	95.0	86.8	99.9	68.9	98.8	60.5	99.3	86.5	100.0	64.7	100.0
C	35.2	100.0	31.8	100.0	72.0	100.0	39.5	100.0	46.5	100.0	80.8	100.0	47.1	100.0
D	63.3	79.2	57.9	90.6	68.8	99.9	41.9	98.8	43.6	98.8	73.1	100.0	41.2	100.0
E	42.9	94.1	39.8	96.8	66.6	99.9	36.9	99.5	36.6	99.8	73.1	100.0	29.4	100.0
F	26.8	100.0	24.2	100.0	57.9	100.0	20.8	100.0	30.9	100.0	67.3	100.0	23.5	100.0
G	20.0	98.6	14.1	99.0	67.2	99.9	9.2	99.8	37.0	99.8	77.9	100.0	58.8	100.0
H	17.3	99.4	11.5	99.6	65.3	100.0	9.0	99.9	34.6	99.9	77.9	100.0	47.1	100.0
I	13.1	100.0	8.1	100.0	55.1	100.0	5.0	100.0	27.6	100.0	72.1	100.0	29.4	100.0
J	15.5	98.8	10.1	99.5	51.8	100.0	3.6	99.9	23.5	99.9	66.4	100.0	35.3	100.0
K	13.3	99.6	8.3	99.8	50.7	100.0	3.4	99.9	22.6	100.0	66.4	100.0	23.5	100.0
L	10.4	100.0	6.3	100.0	44.0	100.0	1.6	100.0	18.9	100.0	60.6	100.0	17.7	100.0

¹Data from the Canadian Bovine Mastitis Research Network National Cohort of Dairy Farms study (Reyher et al., 2011).

²See Table 1 for definitions of categories.

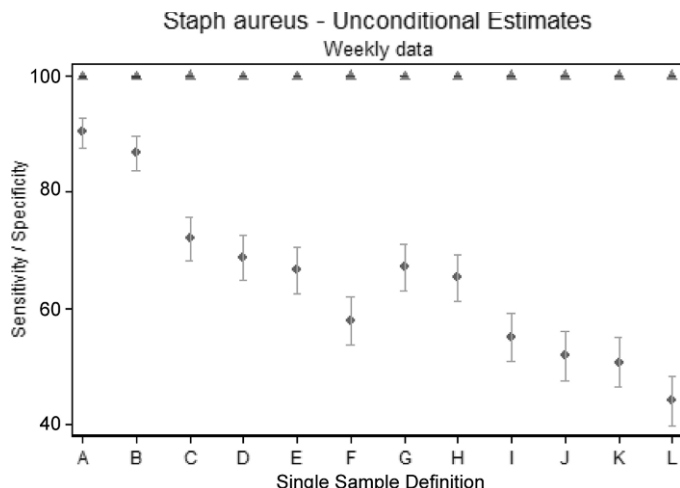


Figure 1. Estimates (and 95% CI) of sensitivity (Se) and specificity (Sp) for various definitions of IMI for *Staphylococcus aureus* compared with a gold standard definition based on 3 consecutive weekly samples (weekly data). See Table 1 for definitions of categories A to L.

the estimates of Se and Sp of the various definitions for *Staph. aureus* and CNS are shown in Figures 1 and 2, respectively.

Daily Data. Unconditional estimates of the Se and Sp of each of the definitions are presented in Table 3. Graphs of the estimates of Se and Sp of the various definitions for *Streptococcus* spp. and CNS are shown in Figures 3 and 4, respectively. As with the weekly data, estimates of Sp were generally high ($\geq 95\%$) for all definitions considered, except for CNS or ANY when only 1 or 2 colonies were required (definitions A, B, D, and E). The estimates of Sp for CNS when a single colony was required in either mixed or pure culture were quite similar to, but slightly lower than, those observed from the weekly data (daily mixed and pure were 84.3 and

86.7%, respectively, and weekly mixed and pure were 88.5 and 90.6%, respectively). In general, estimates of Sp from the daily data were only slightly lower than those based on the weekly data.

Estimates of Se from the daily data were all low and substantially lower than those from the weekly data. For example, for *Streptococcus* spp., estimates of Se derived from the weekly data ranged from 67 to 87% (definitions A to F) but only from 3 to 29% in the daily data.

Random Effects Models

Weekly Data. Random effects models were fit to account for the clustering of observations within quarter, cow, and herd and to allow for an evaluation of the effects of parity and DIM on estimates of Se and Sp. For CNS and *Staph. aureus*, separate models were fit using the GS+ and GS- observations. Although we attempted to include random effects for all levels of clustering (quarter, cow, and herd) in each model, not all models converged. For models of Sp, the total variance of all 3 levels averaged 1.19 and herd, cow, and quarter random effects were successfully estimated in 10, 10, and 5 of the 24 possible models, respectively. For models of Se, the total variance of all 3 levels averaged 2.57 and herd, cow, and quarter random effects were successfully estimated in 21, 15, and 22 of the 24 possible models, respectively.

Days in milk was rarely found to be a significant predictor, even at the $P = 0.1$ level (data not shown), so no further consideration was given to its effect on Se or Sp. Parity effects were not generally linear so parity was converted to a 3-level categorical variable (first, second-third, and fourth+). This variable was a significant predictor of Se and Sp in some models but the

Table 3. Unconditional estimates of sensitivity (Se) and specificity (Sp) for several pathogens (or pathogen groups) with the gold standard based on 3 consecutive samples taken 2 d apart (d 1, 3, and 5) and the estimates based on samples from d 2 and 4¹

Organism	Any organism		CNS		<i>Streptococcus</i> spp.		Enterococci	
	Se	Sp	Se	Sp	Se	Sp	Se	Sp
Gold standard negative	1,564		1,735		2,284		2,301	
Gold standard positive	851		680		131		114	
Total number of triplicate sets	2,415		2,415		2,415		2,415	
Definition ²	Se	Sp	Se	Sp	Se	Sp	Se	Sp
A	65.1	77.3	61.2	84.3	29.1	94.8	40.4	96.3
B	51.7	88.0	49.6	92.9	16.0	97.3	31.6	97.7
C	26.7	97.0	26.0	98.1	6.9	99.3	14.9	99.5
D	51.8	81.6	47.2	88.5	19.1	97.0	18.4	98.3
E	38.5	92.3	38.1	95.5	10.7	98.6	14.0	99.0
F	20.1	98.3	21.2	99.0	3.1	99.8	7.9	99.8

¹Data from a joint study conducted by the University of Prince Edward Island, Charlottetown, PE, Canada; University of Guelph, Ontario, Canada; University of Minnesota, St. Paul, and Cornell University, Ithaca, NY.

²See Table 1 for definitions of categories.

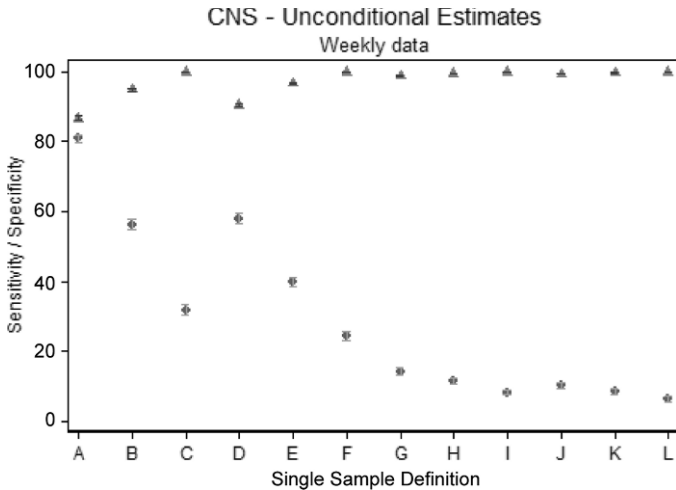


Figure 2. Estimates (and 95% CI) of sensitivity (Se) and specificity (Sp) for various definitions of IMI for CNS compared with a gold standard definition based on 3 consecutive weekly samples (weekly data). See Table 1 for definitions of categories A to L.

effects were generally small. For example, the estimated Se of definition A for CNS varied from 80.7 to 84.1% across the 3 parity groups. Given the generally small effects, the effects of parity were not considered further and single estimates of Se and Sp for each organism were obtained.

Subject-specific estimates of Se and Sp (and their 95% confidence intervals) derived from the random effects models were converted to population averaged (marginal) estimates. In general, these marginal estimates were very close to the unconditional associations described above. However, the confidence intervals were

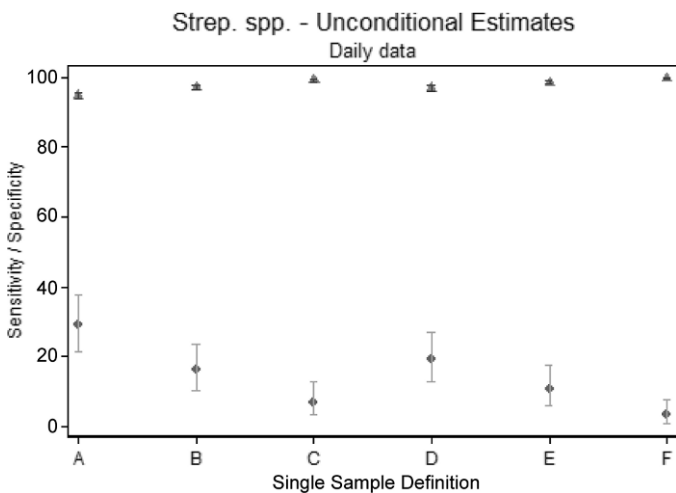


Figure 3. Estimates (and 95% CI) of sensitivity (Se) and specificity (Sp) for various definitions of IMI for *Streptococcus* spp. compared with a gold standard definition based on 3 consecutive samples 2 d apart (daily data). See Table 1 for definitions of categories A to F.

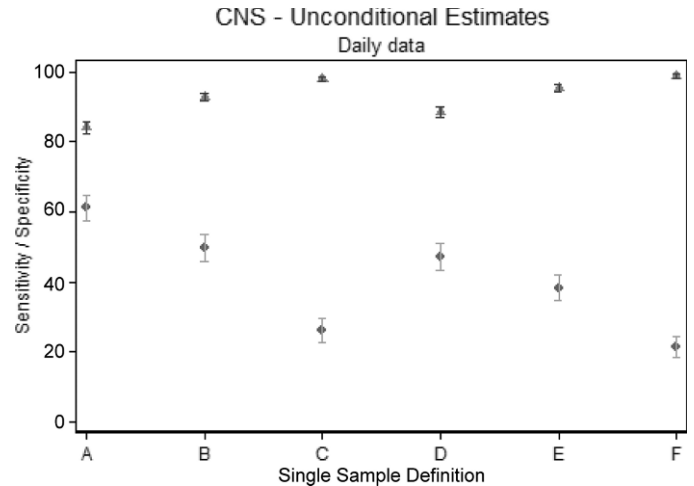


Figure 4. Estimates (and 95% CI) of sensitivity (Se) and specificity (Sp) for various definitions of IMI for CNS compared with a gold standard definition based on 3 consecutive samples 2 d apart (daily data). See Table 1 for definitions of categories A to F.

substantially larger (particularly for estimates of Se). Figures 5 and 6 show the unconditional and marginal estimates of Se for *Staph. aureus* and CNS.

Daily Data. The sum of the variances for the random effects for definition A using the CNS data were 3.52 and 1.07 for the Se and Sp models, respectively. The subject-specific estimates derived from these random effects models were then marginalized and converted to population average estimates of Se and Sp. The unconditional and population average estimates of Se for

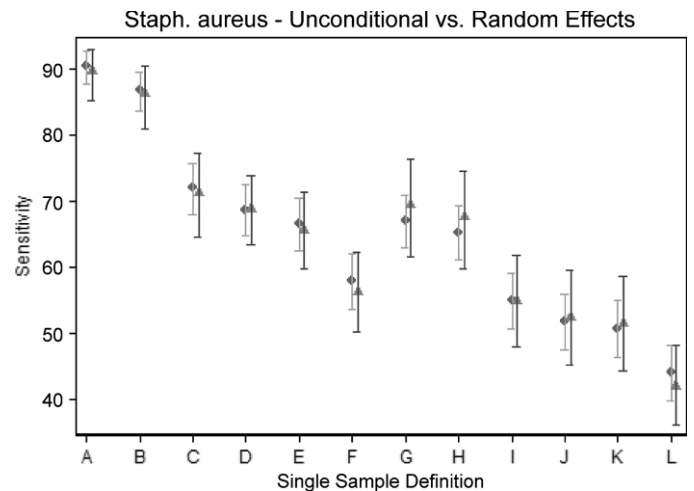


Figure 5. Comparison of unconditional (circles) and marginalized random effects estimates (triangles) of sensitivity (Se) for definitions for *Staphylococcus aureus* compared with a gold standard definition based on 3 consecutive weekly samples (weekly data). Vertical bars are 95% confidence intervals. See Table 1 for definitions of categories A to L.

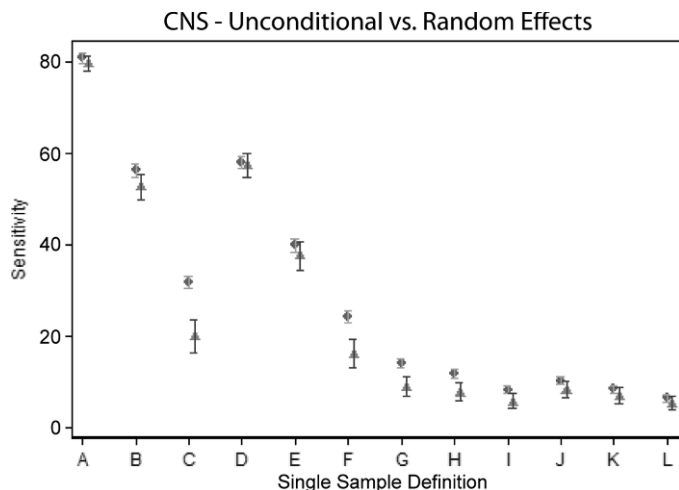


Figure 6. Comparison of unconditional (circles) and marginalized random effects estimates (triangles) of sensitivity (Se) for definitions for CNS compared with a gold standard definition based on 3 consecutive weekly samples (weekly data). Vertical bars are 95% confidence intervals. See Table 1 for definitions of categories A to L.

CNS (and 95% confidence intervals) were 61.2 (57.4 and 64.9%) and 61.8% (56.2 and 67.1%), respectively, whereas the corresponding estimates for Sp were 84.3 (82.5 and 85.9%) and 84.5% (81.9 and 86.8%), respectively.

DISCUSSION

Study Populations

The data required to estimate the operating characteristics of a test are extensive. The existence of a large data set from the CBMRN National Cohort of Dairy Farms along with a data set from a joint project conducted by researchers at the Universities of Prince Edward Island, Guelph, Minnesota, and Cornell provided an opportunity to estimate the operating characteristics of a range of possible definitions of an IMI based on a single milk sample. Given the extensive sampling required, herds in both studies were purposively selected. However, they were broadly based geographically and considered representative of the dairy populations in those regions. Although the protocol for classifying isolates was slightly different in the 2 projects, they were substantively the same and followed NMC guidelines.

Gold Standard

In the absence of a perfect gold standard, using the best diagnostic protocol available as the gold standard is important in any test evaluation. Culture results from triplicate quarter milk samples have often been considered the gold standard for classifying quarters with regard to IMI status. However, we recently found

that, even when complete data from 3 independent samples are available, there can be considerable disagreement among mastitis experts as to the probability of an IMI within the quarter (Andersen et al., 2010). Consequently, a consensus standard, agreed to by participants in the Mastitis Research Workers' Conference, was used as the gold standard for this evaluation.

Consideration must be given to what biases may have resulted from the use of an imperfect gold standard. Truly uninfected quarters were unlikely to meet the criteria laid out to be considered gold standard positive (GS+), so there were probably very few false positives among the GS+ samples. Any false positives that did exist would have had a high probability of being classified as negative from the single sample and this would bias the estimated Se downward. However, given that there were probably very few false positives, this bias would have been very small.

Some quarters with an IMI on the test day of interest may have failed to meet the criteria required to be considered GS+. Two consequences arise from this misclassification. First, quarters classified as gold standard negative (GS-) would have included some truly infected quarters. If these came up positive on one or more of the single sample definitions, they would have been incorrectly classified as false positives. Consequently, the estimates of Sp in this paper may be somewhat negatively biased. However, given that almost all of the Sp estimates were very high, the magnitude of this bias cannot have been large. The second consequence of missing some infected quarters from the GS+ group was that quarters classified as GS+ represented a biased subset of all infected quarters and samples in this subset were presumably easier to classify as positive based on a single sample. Consequently, the estimates of Se derived in this paper may well be biased upward.

In the weekly data, the test sample of interest (i.e., the sample to which the single sample definitions were applied) was the middle sample of the triplicate set used to set the GS. Using this sample both as part of the reference (GS) test and the sample of interest has the potential to bias the estimates of both the Se and Sp upward. This partially accounts for the large discrepancy in estimates of Se between the weekly data and daily data (the former having much higher estimates of Se). No obvious bias was found in the estimates of Sp because for the 3 species in common (ANY, CNS, and enterococci), the estimates of Sp from the 2 data sets were very similar.

Estimates of Sensitivity and Specificity

The results show that all culture procedures have limited Se and requiring anything other than the iso-

lation of 1 organism from 0.01 mL of milk (100 cfu) exacerbates the problem of limited Se. This does not imply that definition A is universally the best definition. The appropriate definition to use will depend on the objectives of the study and the relative costs of false negatives (IMI not detected) and false positives (uninfected quarters classified as having an IMI).

The results from ANY and CNS were very similar because the vast majority of IMI in the ANY category were CNS. The sensitivity of detection of an IMI with ANY was only 65% (based on daily data - slightly higher for weekly data) with definition A. This Se fell to 27% if 10 organisms per 0.01 mL (1,000 cfu/mL) were required for a positive diagnosis (definition C). Values of Sp were also quite low (77 and 88% for definitions A and B, respectively) and were only reasonably high when 10 cfu/0.01 mL were required (definitions C and F).

Culture of a single milk sample worked best when the target organism was *Staph. aureus*. Definition A resulted in a Se of 90.4% and Sp of 99.8%, suggesting that this definition will always be appropriate for detecting IMI due to *Staph. aureus*.

For *Corynebacterium* spp., enterococci, and *E. coli*, results from the weekly data were broadly similar (Se of approximately 75% and Sp >97% for definition A, with Se decreasing to 40 to 50% for definition C, whereas the Sp increased to 100%). However, a large discrepancy was found between the Se estimates for enterococci from the weekly data and the daily data (Se = 40 and 15% for definitions A and C, respectively). As discussed above, the Se estimates from the weekly data were likely biased upwards, suggesting that those from the daily data may be more appropriate.

For *Streptococcus* spp., the estimates of Sp were consistently high in both data sets. Only with definition A in the daily data did the Sp decrease below 97%. However, the estimates of Se for *Streptococcus* spp. were much higher (61–87%) in the weekly data compared with those in the daily data (3–29%). Although part of the explanation for the discrepancy may be the expected upward bias in the Se estimates from the weekly data (described above), part of the explanation lies in the fact that the nature of the *Streptococcus* spp. infections were very different in the 2 data sets. In the weekly data, the 104 triplicate sets that were GS+ came from only 43 quarters with 74 of the 104 GS+ sets coming from quarters that produced 3 or more triplicate sets. In addition, 89% of all *Streptococcus* spp. isolates in the weekly data had more than 1,000 cfu/mL on the culture. These results suggest that these infections were generally persistent, chronic infections shedding large numbers of organisms. In contrast, of the 39 quarters that were GS+ in the daily data, only 6 were positive

on all 3 d (1, 3, and 5) and only 16 had 10 or more colonies on d 3. Of all *Streptococcus* spp. isolates in the daily data, only 19% had more than 1,000 cfu/mL. These results suggest that the *Streptococcus* spp. isolates in the daily data were much more transient infections with the quarters shedding fewer organisms. Why the patterns of infection with these 2 organisms were so different between the 2 data sets is not known.

Requiring that an organism be isolated in pure culture generally reduced the Se of the definition with only slight gain in the Sp. Most clinicians and researchers already accept that there can be at least 2 isolates from a quarter, so requiring isolation of a pure culture to classify a quarter as having an IMI seems inappropriate. Adding a requirement of a minimum SCC before a quarter is considered positive drastically reduced estimates of Se with, in most cases, relatively little gain in Sp (except for definitions A and B for ANY and CNS).

Predictive Values

As noted above, the appropriate definition to choose for any particular circumstance will depend on the objectives of the activity. However, clinicians and researchers are often more interested in the positive (PPV) and negative (NPV) predictive values of a test than the Se and Sp. The PPV is the proportion of test positive samples that are truly infected, whereas the NPV is the proportion of test negative samples that are truly not infected (Dohoo et al., 2009). Predictive values depend on the prevalence of disease in the population in which the test is being used in addition to the Se and Sp. Figure 7 shows the PPV and NPV for definitions A and C for the detection of CNS based on the Se and Sp estimates from the daily data. As can be seen, PPV are quite low when the prevalence is <20%, but definition C performs much better than definition A. However, as the prevalence goes up, the amount by which definition A outperforms C in terms of NPV also goes up. In terms of overall correct classification, definition C performs better up to 28% and thereafter, definition A works better (Figure 7).

Random Effects Models

Random effects models were employed in the analyses for 2 reasons. The first was to evaluate the effects of factors such as stage of lactation and parity on the estimates of Se and Sp. Few significant effects were observed and they did not follow any consistent pattern. Despite the large size of these data sets, they may still have had insufficient power to detect the effects of interest. Alternatively, there may truly be no (or very

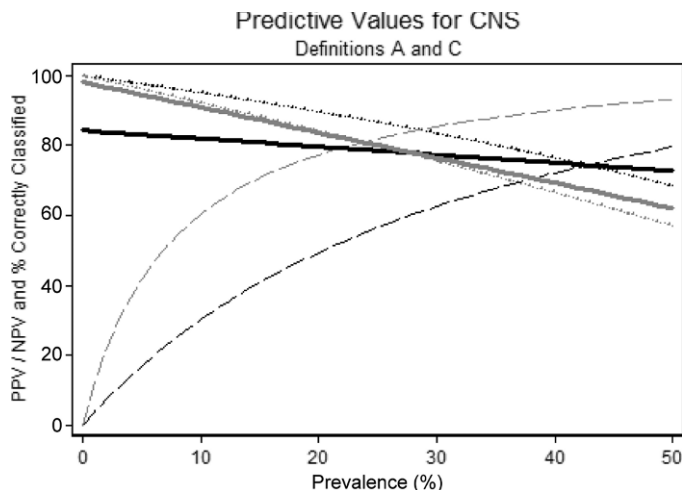


Figure 7. Positive (PPV; dashed lines) and negative (NPV; dotted lines) predictive values for definitions A (minimum of 1 colony per 0.01 mL; black lines) and C (minimum of 10 colonies per 0.01 mL; gray lines) for prevalences ranging from 0 to 50%. Also shown are overall percentage correctly classified (solid straight lines, A = black and C = gray).

small) effects, which would suggest that the Se and Sp of culture is the same regardless of the stage of lactation or parity of the cow from which the milk sample was collected. The second reason was to provide better estimates of the confidence intervals of the Se and Sp estimates. The confidence intervals of the Sp were so narrow that no detectable difference was observed between the 2 analytical approaches. The confidence intervals for the Se were wider when derived from the random effects models (as expected), but the difference was generally not large.

Random effects models produce different point estimates of Se and Sp than do unconditional associations because they weight observations differently (they assign relatively more weight to individual observations derived from groups with small numbers of observations and less from large groups). However, there was no reason to prefer these weighted estimates over the unconditional ones, so for the sake of simplicity, the unconditional estimates were presented. However, the confidence intervals for these unconditional estimates are smaller than they should be.

CONCLUSIONS

The Se and Sp of a variety of definitions of an IMI based on results from a single milk sample were computed by using a set of triplicate milk samples as the gold standard. Two extensive data sets were available for the analysis: one with weekly milk samples making up the triplicate set and the other having the 3 samples

collected over 5 d. For all species except *Staph. aureus*, the Se of all definitions was <90% (and in many cases <50%). Consequently, if identifying as many existing infections as possible is important, then the criteria for considering a quarter positive should be a single colony (from a 0.01-mL milk sample) isolated (definition A). With the exception of ANY and CNS, all Sp estimates were over 94% in the daily data and over 97% in the weekly data, suggesting that for most species definition A may be acceptable. For CNS, definition B (2 colonies from a 0.01-mL milk sample) increased the Sp to 92 and 95% in the daily and weekly data, respectively. For ANY, using definition B increased the Sp to 88 and 93% in the 2 data sets, respectively. The final choice of definition will depend on the objectives of study or control program for which the sample was collected.

ACKNOWLEDGMENTS

The authors acknowledge the intensive work carried out during the summer of 2007 by the national coordinator of the National Cohort of Dairy Farms Kristen Reyher and the regional technicians Theresa Andrews, Natasha Robinson, Meliza Morris, and Jane Saunders from the University of Prince Edward Island (Charlottetown, Canada), Mike MacLean from the University of Guelph (Kemptville, Canada), Francois Dubois from Université de Montreal (Saint-Hyacinthe, Canada), Andrea Wasko and Anke Wellen from University of Calgary (Calgary, Canada). This research was financed by the Natural Sciences and Engineering Research Council (Ottawa, Canada); Alberta Milk (Edmonton, Canada); Dairy Farmers of New Brunswick (Sussex, Canada), Nova Scotia (Lower Truro, Canada), Ontario (Mississauga, Canada), and Prince Edward Island (Charlottetown, Canada); Novalait Inc. (Sainte-Foy, Canada); Dairy Farmers of Canada (Ottawa, Canada); Canadian Dairy Network (Guelph, Canada); Agriculture and Agri-Food Canada (Ottawa, Canada); Public Health Agency of Canada (Ottawa, Canada); Technology PEI Inc. (Charlottetown, Canada); Université de Montreal (Montreal, Canada) and University of Prince Edward Island (Charlottetown, Canada) through the Canadian Bovine Mastitis Research Network (Saint-Hyacinthe, Canada).

REFERENCES

- Andersen, S., I. R. Dohoo, R. Olde Riekerink, and H. Stryhn., and Mastitis Research Workers' Conference. 2010. Diagnosing intramammary infections: Evaluating expert opinions on the definition of intramammary infection using conjoint analysis. *J. Dairy Sci.* 93:2966–2975.
- Bansal, B. K., J. Hamann, N. T. Grabowskit, and K. B. Singh. 2005. Variation in the composition of selected milk fraction samples from healthy and mastitic quarters, and its significance for mastitis diagnosis. *J. Dairy Res.* 72:144–152.

- Barkema, H. W., Y. H. Schukken, T. J. G. M. Lam, D. T. Galligan, M. L. Beiboer, and A. Brand. 1997. Estimation of interdependence among quarters of the bovine udder with subclinical mastitis and implications for analysis. *J. Dairy Sci.* 80:1592–1599.
- Berry, D. P., and W. J. Meaney. 2006. Interdependence and distribution of subclinical mastitis and intramammary infection among udder quarters in dairy cattle. *Prev. Vet. Med.* 75:81–91.
- Deluyker, H. A., S. N. v. Oye, and J. F. Boucher. 2005. Factors affecting cure and somatic cell count after pirlimycin treatment of subclinical mastitis in lactating cows. *J. Dairy Sci.* 88:604–614.
- Dingwell, R. T., K. E. Leslie, Y. H. Schukken, J. M. Sargeant, and L. L. Timms. 2003. Evaluation of the California mastitis test to detect an intramammary infection with a major pathogen in early lactation dairy cows. *Can. Vet. J.* 44:413–415.
- Djabri, B., N. Bareille, F. Beaudou, and H. Seegers. 2002. Quarter milk somatic cell count in infected dairy cows: a meta-analysis. *Vet. Res.* 33:335–357.
- Dohoo, I. R., S. W. Martin, and H. Stryhn. 2009. *Veterinary Epidemiologic Research*. 2nd ed. VER Inc., Charlottetown, PEI.
- Enøe, C., M. P. Georgiadis, and W. O. Johnson. 2000. Estimation of sensitivity and specificity of diagnostic tests and disease prevalence when the true disease state is unknown. *Prev. Vet. Med.* 45:61–81.
- Griffin, T. K., S. V. Morant, and F. H. Dodd. 1987. Bovine mastitis. Definition and guidelines for diagnosis. *IDF Bull.* No. 211. International Dairy Federation, Brussels, Belgium.
- Halasa, T., K. Huijps, O. Osteras, and H. Hogeveen. 2007. Economic effects of bovine mastitis and mastitis management: A review. *Vet. Q.* 29:18–31.
- Hillerton, J. E., J. Cooper, and J. Morelli. 2007. Preventing bovine mastitis by a postmilking teat disinfectant containing acidified sodium chlorite. *J. Dairy Sci.* 90:1201–1208.
- National Mastitis Council. 1987. *Laboratory and Field Handbook on Bovine Mastitis*. National Mastitis Council, Arlington, VA.
- National Mastitis Council. 1999. *Laboratory Handbook on Bovine Mastitis*. Rev. ed. National Mastitis Council, Madison, WI.
- Parker, K. I., C. W. Compton, F. M. Annis, C. Heuer, and S. McDougall. 2008. Quarter-level analysis of subclinical and clinical mastitis in primiparous heifers following the use of a teat sealant or an injectable antibiotic, or both, precalving. *J. Dairy Sci.* 91:169–181.
- Reyher, K. K., S. S. Dufour, H. W. Barkema, L. Des Côteaux, T. J. DeVries, I. R. Dohoo, G. P. Keefe, J.-P. Roy, and D. T. Scholl. The National Cohort of Dairy Farms—A data collection platform for mastitis research in Canada. *J. Dairy Sci.* (under review) 10.3168/jds.2010-3180
- Schukken, Y. H., D. J. Wilson, F. Welcome, L. Garrison-Tikofsky, and R. N. Gonzalez. 2003. Monitoring udder health and milk quality using somatic cell counts. *Vet. Res.* 34:579–596.
- Seegers, H., C. Fourichon, and F. Beaudou. 2003. Production effects related to mastitis and mastitis economics in dairy cattle herds. *Vet. Res.* 34:475–491.
- StataCorp. 2009. *Stata Statistical Software*. Release 11 ed. StataCorp LP, College Station, TX.
- Zadoks, R. N., H. G. Allore, H. W. Barkema, O. C. Sampimon, Y. T. Grohn, and Y. H. Schukken. 2001. Analysis of an outbreak of *Streptococcus uberis* mastitis. *J. Dairy Sci.* 84:590–599.

APPENDIX

Table A1. Participants at the 2008 Mastitis Research Workers' Conference who reviewed and endorsed the approach used to evaluate mastitis definitions used in this manuscript¹

Name	Institution	Name	Institution
Andersen, Signe	University of Prince Edward Island (PEI)	Middleton, John	University of Missouri
Andrew, Sheila	University of Connecticut	Mullarky, Isis	Virginia Tech
Barkema, Herman	University of Calgary	Nedrow, Alicia	Virginia Tech
Barlow, John	University of Vermont	Nickerson, Stephen	University of Georgia
Chaffer, Marcello	University of PEI	Olde Riekerink, Richard	Animal Health Service, the Netherlands
De Vliegher, Sarne	Ghent University	Oliver, Steve	University of Tennessee
Dohoo, Ian	University of PEI	Owens, William	Louisiana State University
Dufour, Simon	University of Montreal	Perez-Casal, Jose	University of Saskatchewan
Fox, Larry	Washington State University	Peterson-Wolfe, Christina	Virginia Tech
Hulland, Carol	University of Wisconsin	Piepers, Sofie	Ghent University
Keefe, Greg	University of PEI	Reyher, Kristen	University of PEI
Lacasse, Pierre	Agriculture and Agrifood Canada	Roy, Jean-Philippe	University of Montreal
Lago, Alfonso	University of Minnesota	Scholl, Daniel	University of Montreal
Leslie, Ken	University of Guelph	Schukken, Ynte	Cornell University
Lichtenwalner, Anne	University of Maine	Ster, Celine	Agriculture and Agrifood Canada
Luby, Chris	University of Saskatchewan	Supre, Karlien	Ghent University
MacDonald, Kimberley	University of PEI	Wenz, John	Washington State University
Masiello, Stephanie	Virginia Tech	Wilson, David	Utah State University
McClure, J. T.	University of PEI	Zadoks, Ruth	University of Edinburgh - Moredun
McDougall, Scott	Animal Health Centre - New Zealand		

¹One participant requested that their name not be included.