J. Dairy Sci. 96:4047–4058
http://dx.doi.org/10.3168/jds.2012-6190

Development of protocols to evaluate in-line mastitis-detection systems

C. Kamphuis,*[1] B. Dela Rue,* G. Mein,† and J. Jago*
*DairyNZ Ltd., Private Bag 3221, Hamilton 3240, New Zealand
†Werribee South, Victoria 303, Australia

ABSTRACT

This paper proposes and discusses a methodology to evaluate the performance of automated mastitis-detection systems with respect to their practical value on farm. The protocols are based on 3 on-farm requirements: (1) to detect cows with clinical mastitis promptly and accurately to enable timely and appropriate treatment, (2) to identify cows with high somatic cell count to manage bulk milk SCC levels, and (3) to report the mastitis infection status of cows at the end of lactation to support decisions on individual cow dry-cow therapy. Separate protocols for each requirement are proposed and discussed, including gold standards, evaluation tests, performance indicators, and performance targets. Aspects that require further research or clarification are identified. Actual field data are used as examples. Further debate is invited, the aim being to achieve international agreement on how to evaluate and report performance of different mastitis-detection technologies. Better performance information will allow farmers to compare different mastitis-detection systems sensibly and fairly before investing. Also, the use of evaluation protocols should help technology providers to refine current, or develop new, automated mastitis-detection systems. Such developments are likely to accelerate adoption of these systems, potentially leading to improved animal health, milk quality, and labor productivity.
Key words: testing protocol, udder health management, mastitis sensing system

INTRODUCTION

In most dairy systems, it is assumed that farmers, informed by the official organization in their country or region, have the responsibility to deliver milk that is of sufficient quality (Hogeveen et al., 2010). Although regulations dealing with milk quality standards differ between dairy-producing countries, general agreement exists that abnormal milk, including milk from diseased [e.g., due to clinical mastitis (**CM**) or highly elevated SCC levels] or injured udders should be excluded from milk supplied for human consumption. Identifying cows with CM involves visual inspection of the udder and manually checking the foremilk from each quarter at each milking, whereas regular herd tests and various other systems (e.g., the California mastitis test) can be applied to identify cows with elevated SCC. However, increasing herd sizes, reliance on less-skilled labor, and an increasing emphasis on lowering bulk milk SCC (**BMSCC**) levels (Adkinson et al., 2001; Brightling et al., 2005; Lacy-Hulbert et al., 2010) are all factors contributing to an increased demand for more consistent and less labor-intensive methods to help farmers manage mastitis and BMSCC levels.

Mastitis-detection systems have been developed to automate, or at least contribute to, the processes of managing mastitis and BMSCC. For these automated mastitis-detection systems to be useful on farm, they need to (1) identify cows with CM promptly and accurately to enable timely treatment, maximize cure rate, and reduce the risk of spread of infection; (2) identify cows with high SCC to manage BMSCC levels by withholding cows temporarily or long term; and (3) identify the infection status of cows at the end of the lactation to support decisions on individual cow dry-cow therapy. Most research to date has focused on assessing the sensors' ability to detect CM (Hogeveen et al., 2010), with little or no evidence of field performance of mastitis-detection systems relative to the other 2 requirements. It is important to define evaluation protocols so that independent and uniform performance information can be provided to farmers. This information will enable farmers to assess if a mastitis-detection system will meet their requirements and they will be able to make more informed investment decisions.

Protocols have been established considering each requirement separately and, where possible, comprise (1) an appropriate and practical gold standard, (2) an appropriate test against the gold standard and (3) minimum performance targets. Data from 2 previous

Received September 23, 2012.
Accepted February 11, 2013.
[1]Corresponding author: claudia.kamphuis@dairynz.co.nz

field studies are used to illustrate the proposed gold standards and evaluation tests, and to identify areas that require further research.

DEVELOPMENT OF EVALUATION PROTOCOLS

Requirement 1: Detecting Clinical Mastitis Promptly and Accurately

Background Information. The National Mastitis Council defines CM as the presence of flakes, clots, or other gross alterations in milk appearance, irrespectively of its SCC level (Smith et al., 2001). Whereas this is a clear definition, the development of a practical and objective assessment of the condition of (quarter) milk for large-scale field evaluation of mastitis-detection systems is less clear. Previous research on CM detection used a variety of gold standards to define CM (Hogeveen et al., 2010) and Claycomb et al. (2009) demonstrated that even minor differences in gold standard definitions had a large effect on the number of true CM cases used for evaluation purposes. In addition, differences in data used for analyses (Kamphuis et al., 2010a) and time windows applied (Kamphuis et al., 2010b) are factors contributing to the variation of reported performance of mastitis-detection systems. Recently, there has been effort to develop meaningful and consistent methods of field evaluation of automatic mastitis-detection systems in relation to an appropriate gold standard, data used for analyses, evaluation tests, and performance targets (ISO, 2007; Mein and Rasmussen, 2008; Claycomb et al., 2009; Hogeveen et al., 2010). The information contained in these reports has formed the basis of the protocol proposed to evaluate the ability of a mastitis-detection system to detect CM.

Proposed Gold Standard. For a field evaluation, a CM episode is defined as the observation of clots (>2 mm in average diameter) at 2 out of 3, or all 3, consecutive cow milkings (Mein and Rasmussen, 2008). A single CM episode is not necessarily limited to 3 consecutive milkings but can continue as long as clots are found at 2 out of 3 consecutive milkings. Presence of clots should be assessed using in-line filters (Claycomb et al., 2009). Situation A in Figure 1 illustrates how clot observations should be used to identify CM: 5 of 15 consecutive cow milkings had clots present. Using the proposed gold standard, this example shows one CM episode that started at the first observation of clots (milking 6). The second clot observation at milking 8 serves as a confirmation that the first observation was the start of a CM episode. For the purposes here, the CM episode ends at the last milking in which clots are observed where this is followed by 4 consecutive cow milkings without clot observation (milking 9). Situation A (Figure 1) also has 2 stand-alone clot observations at milking 1 and milking 14. These clot observations, however, fall outside the proposed gold standard definition and are, therefore, not considered to be (part of) a CM episode.

Proposed Evaluation Test. Field data should be collected from a minimum of 3 commercial farms and collection should continue until a minimum of 20 CM episodes per farm have been identified (ISO, 2007; Hogeveen et al., 2010). A mastitis alert should be generated by the detection system within 4 milkings (48-h time window) around the first cow milking with clots in a confirmed CM episode (Hogeveen et al., 2010; Table 1). Time-window analysis (Sherlock et al., 2008) should be used to link alerts generated by a mastitis-detection system with CM episodes. This time-window analysis is illustrated in Figure 1, where 4 situations (A, B, C, and D) show 15 consecutive cow milkings, with 5 having clots present; there is 1 CM episode (milking 6 through milking 9) and 2 stand-alone clot observations (milking 1 and milking 14). The time window of 4 cow milkings is centered around the start of the milking with the first clot observation [i.e., a time window of 2 cow milkings before the first cow milking with clots and 1 cow milking after the first cow milking with clots of a confirmed CM episode (milking 4 through milking 7)]. Situation A in Figure 1 illustrates a situation where the mastitis-detection system generated 3 alerts. The first occurred at the start of the time window of 4 cow milkings and is, therefore, considered a true positive (**TP**) alert. The other 2 alerts occurred outside that time window and are, therefore, false positive (**FP**). Situation B (Figure 1) represents a situation where 3 alerts are generated but all fall outside the time window of 4 cow milkings in which a mastitis alert should be expected and, therefore, they are all FP. This also implies that the CM episode in situation B remains undetected and, therefore, the time window of 4 cow milkings receives a single false-negative (**FN**) alert. Situation C in Figure 1 illustrates a more complex situation where consecutive mastitis alerts occurred within 1 CM episode. The first alert at milking 3 in situation C (Figure 1) is an FP alert. The alert at milking 4 is the first alert within the time window in which a mastitis alert should be expected and is, therefore, TP. If that first alert had occurred at milking 5, 6, or 7, then this would have been classed as TP also. In the situation of consecutive mastitis alerts within the time window of 4 cow milkings (alerts at milking 5 and 7 in situation C), these should be considered as confirmative alerts of the first mastitis alert. Therefore, these confirmative alerts are merged with the first alert and they are counted as 1 TP alert. The mastitis alerts at milkings 8, 9, and 10 in situation C are also confirmative alerts, as they appear within 2