



J. Dairy Sci. 100:1592–1603

<https://doi.org/10.3168/jds.2016-11365>

© 2017, THE AUTHORS. Published by FASS and Elsevier Inc. on behalf of the American Dairy Science Association®.  
This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/3.0/>).

## Validation of the M-stage scoring system for digital dermatitis on dairy cows in the milking parlor

L. Solano,<sup>1</sup> H. W. Barkema, C. Jacobs, and K. Orsel

Department of Production Animal Health, University of Calgary, Calgary, AB, T2N 4N1, Canada

### ABSTRACT

A high prevalence of digital dermatitis (DD) and the benefits of early topical treatment highlight the need for simple tools for routine DD detection. The objective of this study was to determine the accuracy of scoring DD lesions using the 5 M-stage scoring system in the milking parlor compared with the trimming chute as the gold standard. Three observers inspected 3,585 cows and 6,991 hind feet from 9 farms in the milking parlor using a mirror (glued to a plastic kitchen spatula) and a headlamp, followed by inspection in a trimming chute within 5 d. Interobserver agreement for scoring DD in various settings was  $\geq 82\%$  ( $\kappa > 0.74$ ; weighted  $\kappa > 0.76$ ). At trimming chute inspections, 68% of cows had at least 1 DD lesion, 19% had 1 hind leg affected, and 49% had both hind legs affected. Within-herd DD prevalence ranged from 16 to 81% of cows affected. True within-herd prevalence was 2, 6, 0, 36, and 14% for M1, M2, M3, M4, and M4.1 lesions, respectively. At the foot level, DD prevalence was the same (58%) in the milking parlor and trimming chute inspection, but distribution of M-stages differed. Milking parlor inspection as a means of identifying the presence of DD lesions had a sensitivity of 92% and specificity of 88%, with positive and negative predictive values of 91 and 89%, respectively. Agreement between milking parlor and trimming chute inspections was 73% ( $\kappa = 0.59$ , weighted  $\kappa = 0.65$ ) for the 5 M-stage scoring system and 90% ( $\kappa = 0.80$ ) if only the presence of a lesion was noted. Test characteristics varied greatly among M-stages, with the highest sensitivity for detecting M4 (82%) and M2 (62%) lesions, and the lowest for detecting M4.1 (20%), M1 (7%), and M3 (0%) lesions. In the milking parlor, 20% of M2 lesions were misclassified as M4.1, 8% of M4 lesions were misclassified as M0, and 68% of M4.1 lesions were misclassified as M4. The majority (87%) of DD lesions were located between the heel bulbs; 10 and 2% of DD lesions affected the interdigital space and the front of

the foot, respectively. The sensitivity to detect the presence of a lesion when it occurred between the heel bulbs was 93%, but  $< 67\%$  if it occurred elsewhere on the foot. We concluded that inspection of the rear feet in the milking parlor was an inexpensive and simple method of detecting and scoring DD lesions. If the objective is to determine herd-level DD prevalence and routine monitoring, this method was adequately reliable. However, if the objective is to follow up DD in cows with history of interdigital hyperplasia or to detect M1 or M4.1 lesions, this method was not sufficiently reliable. Although DD scoring in the milking parlor as a routine practice should facilitate early detection, prompt treatment interventions, and herd monitoring, it was not sufficiently reliable to replace definitive identification of M-stages in the trimming chute.

**Key words:** dairy cattle, diagnostic test, lesion scoring, hoof trimming, digital dermatitis

### INTRODUCTION

Digital dermatitis (DD) is a widespread bacterial foot lesion in cattle that typically develops on the bulb of the heel, causing ulcerative lesions that may be uncomfortable or very painful (Cheli and Mortellaro, 1974; Döpfer et al., 1997). It is the most common foot lesion in confined dairy systems, with a prevalence ranging from 15 to 49% (Holzhauer et al., 2006; USDA, 2009; Solano et al., 2016) and causing substantial economic losses due to high incidence and prevalence and the related costs of decreased reproductive performance, increased risk of culling, treatment and labor (Bruijnis et al., 2010; Cha et al., 2010; Gomez et al., 2015b).

As a consequence of pain, cows affected with DD often change their gait or posture to avoid contact with the floor, exhibiting decreased mobility, lifting or shaking the affected leg, or walking with a toe-down posture (Rodriguez-Lainz et al., 1998; Shearer and Van Amstel, 2013). Digital dermatitis is considered a multifactorial and polybacterial disease, with treponemes consistently isolated from DD lesions (Döpfer et al., 2012b; Gomez et al., 2012; Krull et al., 2016). The pathogen can remain endemically present in a herd, and infected cattle can develop active and chronic stages of infection. Chroni-

Received April 26, 2016.

Accepted September 9, 2016.

<sup>1</sup>Corresponding author: [laurasolanoquesada@gmail.com](mailto:laurasolanoquesada@gmail.com)

cally infected cattle are a reservoir of infection and a potential source of outbreaks (Döpfer, 2009; Döpfer et al., 2012a). Additionally, affected cows can experience changes in the heel area, favoring persistence and occurrence of heel horn erosion, an infectious lesion (Gomez et al., 2015a). Thus, DD is a serious animal welfare concern due to painful episodes that can be long lasting, with recurrent outbreaks (Bruijnis et al., 2012; Döpfer et al., 2012a; Gomez et al., 2015b).

Routine monitoring of DD enables early detection and treatment, key factors for the effective management of the disease (Döpfer et al., 2012a). Numerous classification systems for DD have been developed (Laven, 1999; Manske et al., 2002; Vink, 2006; Krull et al., 2014), but over the last 15 years, the scientific community has widely used the M-stage scoring system developed by Döpfer et al. (1997) and amended by Berry et al. (2012). Based on visual observation, this scoring system characterizes various clinical stages of DD over the course of the disease, allowing for observation of transitions between active, chronic, and healed stages. This information provides researchers, farmers and hoof trimmers a tool for monitoring the effectiveness of DD control programs at both individual-animal and herd levels (Döpfer et al., 2012a).

The M-stage scoring system enables macroscopic scoring of DD lesions and is internationally recognized as the most accurate and detailed DD identification system (Greenough et al., 2008), although misclassification bias is expected as diagnosis is based on visual inspection (Relun et al., 2011). Lifting the cow's foot for inspection in the trimming chute continues to be the gold standard for DD detection, although it is expensive, labor-intensive, and stressful for cattle (Thomsen et al., 2008; Relun et al., 2011; Stokes et al., 2012). Furthermore, trimming chute inspections are not practical for assessing disease prevalence, either on a regular basis, or for early DD detection and treatment.

Consequently, alternative diagnostic tools for DD detection have been developed; for example, inspection of hind feet during milking in the parlor with the assistance of a swiveling mirror (Relun et al., 2011) or borescope (Laven, 1999; Vink, 2006; Stokes et al., 2012), or with no specialized tools (Rodriguez-Lainz et al., 1998; Thomsen et al., 2008). In addition, DD has been diagnosed during pen walks while cows were restrained in headlocks (University of Wisconsin, 2013) or by using blood tests to detect active DD cases (Gomez et al., 2014). The 6 studies that evaluated the accuracy of DD detection in the milking parlor used various classification systems for DD, and the majority did not distinguish among DD M-stages. It is noteworthy that the 1 study that used the M-stage scoring system (Relun et al., 2011) was conducted before the release

of the updated scoring system (Berry et al., 2012). In addition, the other 5 studies used a tool that was either expensive or impractical for field conditions (Laven, 1999; Vink, 2006; Stokes et al., 2012) or insufficiently accurate (Rodriguez-Lainz et al., 1998; Thomsen et al., 2008). Simple, inexpensive, and effective tools are needed for routine DD inspection. The objective of this study was to determine the accuracy of detection and scoring DD lesions using the updated M-stage scoring system in the milking parlor, as compared with the trimming chute.

## MATERIALS AND METHODS

### *Farm and Cow Selection*

A total of 17 freestall dairy farms in Alberta, Canada, clients of 2 hoof trimmers, were contacted by telephone to be enrolled as part of a longitudinal study examining the effectiveness of a standardized footbath protocol for prevention of DD (Solano et al., 2017). To ensure that participating farms were representative of the majority of farms in Alberta, farms were selected that had free stall housing systems with no access to pasture, and a herd size of  $\geq 90$  Holstein-Friesian lactating cows. Farms were also selected based on convenience criteria: cows had to be milked in a milking parlor, and farms had to have  $\geq 10\%$  DD prevalence, based on hoof trimming records from the past year. Ten farms met all criteria and agreed to participate in the study. One farm withdrew 1 mo after field data collection had started, because of a change in farm ownership. Data were collected between November 2013 and June 2014 by the corresponding author and 2 trained observers from the University of Calgary (Calgary, AB, Canada). All methods were approved by the Animal Care Committee and Research Ethics Board of the University of Calgary (AC13-0082).

### *DD Lesion Assessment*

The DD lesions were scored using the 5-point scale according to Döpfer et al. (1997) and Berry et al. (2012). Briefly, lesions were classified as M0 if skin was normal with no lesions compatible with DD; M1 if a small ( $< 2$  cm in diameter) focal active lesion was observed, with a red-gray surface and scattered small ( $\sim 1$  mm in diameter) red foci; M2 if an ulcerative active lesion  $\geq 2$  cm in diameter was observed, with a red-gray surface; M3 (healing stage) if lesion presented a dry brown and scab-like tissue, typically seen within few days after topical treatment; M4 (chronic stage) if the lesion surface was raised by brown or black tissue, thickened epithelium, and proliferative or hyperkera-

otic growth; and M4.1 if the chronic stage of M4 had an M1 lesion within its perimeter.

Before the start of the study, the first author conducted a training program with 2 other observers to identify and score DD lesions. The training program was similar to that developed by Gibbons et al. (2012). It consisted of a classroom session with a PowerPoint (Microsoft Corp., Redmond, WA) presentation using digital color photographs to demonstrate each DD stage, along with a detailed description. Each observer received a laminated reference card with a summary table of the scoring system, a characteristic photograph, and brief definition of each M-stage. Observers had ready access to the reference card at all times during both training and data collection. To evaluate interobserver agreement, the 3 observers scored 40 photographs of DD lesions. One day later, during a live session, observers scored the hind feet of 110 cows during milking in the parlor. The following day, observers scored a sample of 40 cows (from the 110 cows scored the day before) in the trimming chute. A mid-way check (12 wk after initial data collection) was conducted using the same number of cows, but different photographs than the initial training, to maintain high agreement throughout the study.

### Milking Parlor Inspection

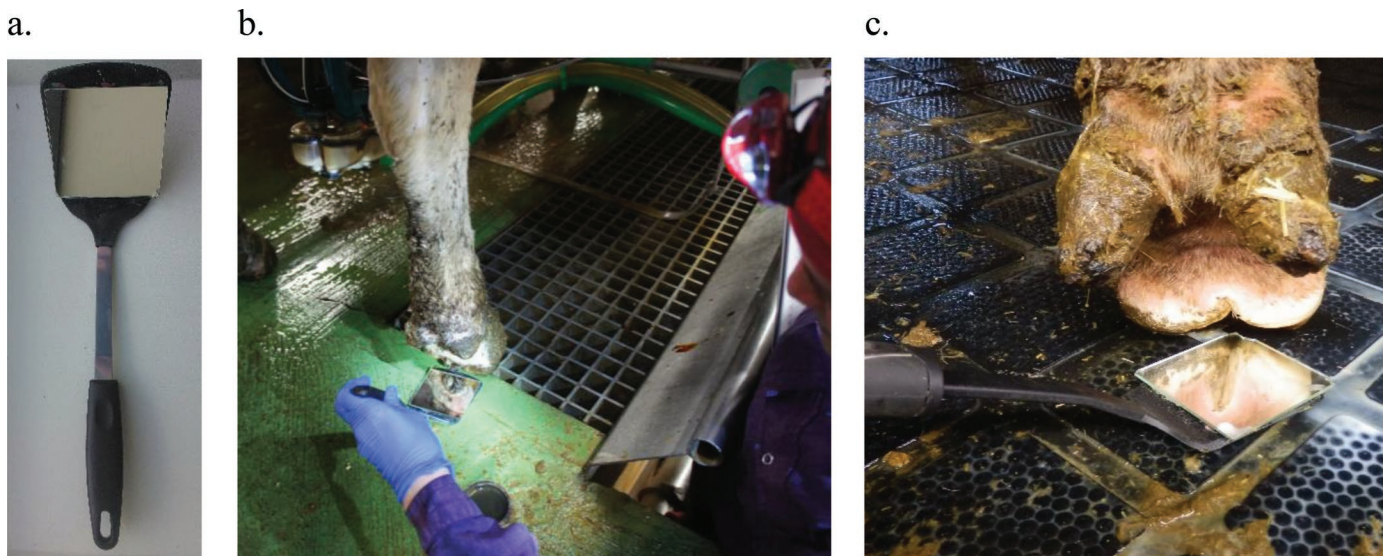
Each farm was visited 3 times during the study period. Each visit consisted of inspection of the entire lactating herd in the milking parlor, followed by inspec-

tion in a trimming chute scheduled within 48 h. The 48-h interval was selected to decrease the possibility of observing a different M-stage in the trimming chute than in the milking parlor, which can occur due to the natural rapid transition between DD stages (Nielsen et al., 2012).

Cows were scored for DD lesions during milking using a scoring method adapted from Relun et al. (2011). Cows' hind feet were washed with water from a hose before examination. Scoring DD lesions was done using a 7.6 × 7.6 cm (3 × 3 in) mirror glued with contact cement to the flat end of a 25-cm plastic kitchen spatula, and extra lighting was provided by an LED headlamp (Figure 1). Inspection of the hind feet only was based on reports that approximately 94% of DD lesions occur on the hind feet and 85% of DD lesions affect the area between heel bulbs (Cramer et al., 2008; Relun et al., 2011; Solano et al., 2016). A total of 3,756 cows were scored, but information was not recorded on 164 feet. Reasons for missing data included poor foot conformation that impeded accurate scoring (i.e., low claw heels), feet covered with dried caked manure that persisted after feet were washed, and cow behavior (e.g., frequent lifting of feet or kicking).

### Trimming Chute Inspection

Cows' hind feet were thoroughly cleaned with paper towels before being scored for DD. Paper towels were preferred to water, because temperatures during the winter could have resulted in ice formation, creating



**Figure 1.** Inspection of digital dermatitis lesions during milking using a simple tool: (a) mirror glued to a kitchen spatula; (b) use of headlamp to improve lighting while scoring lesions; (c) showing an M4 stage lesion that would be difficult to detect without a mirror. Color version available online.



unsafe conditions in the workspace. In addition to M-scoring, the anatomical location of DD lesions was recorded according to Relun et al. (2011): BH = between the heel bulbs; MH = medial heel; LH = lateral heel; D = around the dewclaws; C = on the coronet; F = on the front; UF = under the foot in the interdigital space; HYP = on an interdigital hyperplasia. All lactating cows were scheduled for inspection in the trimming chute within 2 d following the milking parlor inspection. However, the interval between milking parlor and trimming chute inspection was occasionally longer than scheduled (range 0 to 5 d), due to winter conditions [e.g., snowstorms that impeded driving and cold weather (below  $-25^{\circ}\text{C}/-13^{\circ}\text{F}$ ) that caused trimming chute malfunctions. In addition, farms with a large herd size (>150 cows) required 2 d of trimming inspection. On 9 of 27 farm visits (31% of cows), cows were inspected in the trimming chute 3 to 5 d after milking parlor inspection.

Depending on the farm setting and herd size, 1 or 2 trimming chutes were present, with 1 observer per chute. The observer at the milking parlor inspection was not necessarily the same as the observer at the trimming inspection. A total of 3,765 cows were inspected in the trimming chute, but information was inadvertently not recorded on 18 feet. It was not possible to match data between milking parlor and trimming chute inspections for 180 cows, due to missing cow identification.

### Statistical Analyses

Data were entered into Access (Microsoft Corp.), and all statistical analyses were performed using Stata 13.1 (StataCorp, College Station, TX). For all analyses, a  $P$ -value <0.05 was considered significant. The foot was considered the statistical unit. All analyses were based on 5 M-stages. Scores were also combined into a simplified scoring system: no lesions (M0), active lesions (M1, M2, and M4.1 merged) and chronic lesions (M3 and M4 merged). This simplified system aimed to detect active stages (M1, M2, and M4.1; according to Berry et al., 2012), which are relevant for early topical treatment and to decrease the spread of infection (Döpfer et al., 2012a).

If >1 DD lesion was observed on the same foot at trimming chute inspections, analysis was done using the most clinically relevant M-stage, as follows:  $M2 > M4.1 > M1 > M4 > M3$  (Relun et al., 2011; Gomez et al., 2014). For example, a foot with an M2 and M4.1 lesion was classified as M2, a foot with an M4.1 and M4 lesion was classified as M4.1, and a foot with an M1 and M4 lesion was classified as M1. An M4.1 lesion was prioritized over an M1, because it is a chronic and

recurrent stage. For the simplified scoring system, a foot with an active and chronic lesion was classified as active, whereas a foot with a chronic lesion and no active lesion was classified as chronic. This classification attempted to address the accuracy of detecting the M-stages in the milking parlor that could benefit most from early treatment and control measures to prevent M2 lesions.

Interobserver agreement was calculated by assessing DD scores attributed by the 3 observers during the classroom session with pictures and live in the milking parlor and trimming chute. Agreement among observers was assessed by the percentage of agreement [number of exact agreements/total number of observations  $\times 100$ , ( $\text{PA}_o$ )], whereas actual agreement beyond chance was assessed by kappa ( $\kappa$ ; Cohen, 1960). To quantify the magnitude of discrepancy among raters, linearly weighted kappa ( $\kappa_w$ ) for ordinal scores was calculated using a weight matrix (Cohen, 1968; Fleiss et al., 2003) according to the order  $M2 > M4.1 > M1 > M4 > M3 > M0$  (5 M-stages), and active > chronic > no lesion (simplified scoring system). Agreement as measured using  $\kappa$  was interpreted according to Landis and Koch (1977) as  $\leq 0$  = poor, 0.01 to 0.20 = slight, 0.21 to 0.40 = fair, 0.41 to 0.60 = moderate, 0.61 to 0.80 = substantial, and 0.81 to 1.00 = almost perfect.

Digital dermatitis scores attributed during trimming chute inspections were considered the gold standard or reference category in all pertinent analyses. Overall agreement was estimated by comparing DD scores during milking and at the trimming chute. Lesions that occurred in the interdigital space (under the foot or in a hyperplasia) were excluded from these analyses, as it was not possible to evaluate this location during milking parlor inspections. Sensitivity (Se), specificity (Sp), positive predictive value (PPV), and negative predictive value (NPV) were calculated for the absence (M0) and presence (M1, M2, M3, M4, or M4.1) of disease and for specific M-stages. Chi-squared tests were used to compare test characteristics between trimming chute inspections that occurred during 0- to 2-d and 3- to 5-d intervals after milking parlor inspections. No significant differences were detected between the 2 periods, so all data were included in the analyses.

Apparent prevalence and accuracy estimates were used to calculate true DD prevalence per M-stage. True prevalence was calculated as apparent prevalence  $+ (\text{Sp} - 1)/\text{Se} + (\text{Sp} - 1)$  (Dohoo et al., 2009).

## RESULTS

Average herd size was 151 lactating cows (ranging from 92 to 200), with a mean daily milk production

of 33 kg/cow and a median parity of 2. Three study farms had herringbone milking parlors, and 6 farms had parallel parlors.

**Interobserver Agreement**

Interobserver agreement for scoring DD was substantial to almost perfect (agreement >82%;  $\kappa > 0.74$ ;  $\kappa_w > 0.76$ ) throughout the study period using the 5 M-stages and the simplified scoring system (Table 1). Agreement at the mid-way check improved in the classroom session using the 5 M-stages ( $\kappa = 0.77$  to  $0.83$  at the first and mid-way check picture sessions, respectively), and overall when using the simplified scoring system. However, applying  $\kappa_w$  had little or no effect for the simplified scoring system.

**Trimming Chute Inspection**

A total of 6,991 feet were scored in the milking parlor and in the trimming chute. At trimming chute inspections, 67.6% of cows had at least 1 DD lesion; 18.8% had 1 hind leg affected, and 48.8% had both hind legs affected. Within-herd DD prevalence ranged from 16 to 81% of cows affected. A total of 10.4 and 0.3% of legs affected had 2 and 3 DD lesions on the same foot, respectively (Table 2). The majority (87.0%) of DD lesions were located between the heel bulbs, 9.8% affected the interdigital space (including on an interdigital hyperplasia), and 2.4% affected the front of the foot (Table 2). A total of 78.4 and 19.1% of DD lesions occurred in the interdigital space and at the front of the foot, respectively, if >1 lesion was detected.

**Accuracy of Milking Parlor Inspection**

At milking parlor inspections, 66.4% of cows had at least 1 DD lesion. At the foot level, DD prevalence was the same (58%) in the milking parlor and trimming chute inspection, but distribution of M-stages differed. Apparent prevalence at milking parlor inspections for each M-stage was 0.3, 5.4, 0.8, 44.9, and 6.6% for M1, M2, M3, M4, and M4.1, respectively. Among the 3 most frequent M-stages, the majority of misclassifications involved M2 with M4.1, M4 with M0, and M4.1 with M4 stages, when considering trimming chute scores as the reference (Table 3). In the milking parlor, 20% of M2 lesions were misclassified as M4.1, 8% of M4 lesions were misclassified as M0, and 68% of M4.1 lesions were misclassified as M4 (Table 3). When using the simplified scoring system, 51% of active lesions were misclassified as chronic. Agreement between milking parlor and trimming chute inspections was moderate ( $\kappa = 0.59$ ) and substantial ( $\kappa_w = 0.65$ ) for the 5 M-stage scoring

**Table 1.** Interobserver agreement for scoring digital dermatitis on dairy cows with pictures and live in the milking parlor and trimming chute using the 5 M-stage system and a simplified scoring system

Setting	Agreement coefficients of 5M-stages <sup>1</sup>			Agreement coefficients of simplified system <sup>2</sup>		
	PA <sub>o</sub> <sup>3</sup>	$\kappa^4$ (95% CI)	$\kappa_w^5$ (95% CI)	PA <sub>o</sub>	$\kappa$ (95% CI)	$\kappa_w$ (95% CI)
Classroom with pictures	82.9 (70.2-94.3)	0.77 (0.67-0.86)	0.84 (0.68-0.92)	90.6 (73.2-98.4)	0.83 (0.73-0.92)	0.83 (0.72-0.98)
On-farm in milking parlor	82.1 (72.6-90.2)	0.74 (0.69-0.78)	0.76 (0.61-0.85)	86.3 (76.1-92.5)	0.78 (0.71-0.85)	0.79 (0.68-0.89)
On-farm in the trimming chute	87.5 (78.7-93.5)	0.82 (0.72-0.90)	0.89 (0.79-0.92)	90.9 (82.9-96.0)	0.86 (0.75-0.95)	0.91 (0.88-0.93)
Classroom with pictures MW <sup>6</sup>	88.3 (76.3-98.4)	0.83 (0.74-0.90)	0.88 (0.73-0.95)	93.3 (79.6-99.9)	0.87 (0.77-0.95)	0.89 (0.76-1.00)

<sup>1</sup>M0 to M4.1 (Berry et al., 2012).

<sup>2</sup>Simplified scoring system: no lesion (M0), active lesion (M1/M2/M4.1), and chronic lesion (M3/M4).

<sup>3</sup>Mean overall percentage agreement.

<sup>4</sup>Kappa = agreement beyond chance ( $\kappa$ ).

<sup>5</sup>Weighted kappa ( $\kappa_w$ ) according to order: M2 > M4.1 > M1 > M4 > M3 > M0 (5 M-stage), and active > chronic > no lesion.

<sup>6</sup>Mid-way check (12 wk after initial data collection).

**Table 2.** Distribution (no.; % in parentheses) of anatomical locations of digital dermatitis (DD) lesions identified at trimming chute inspection on 4,161 hind feet (N) of dairy cattle

DD stage <sup>1</sup>	N	Location <sup>2</sup>						
		BH	D	MH	LH	F	HYP	UF
Sensitivity <sup>3</sup>	—	92.7	33.3	0.0	66.7	17.7	0.0	0.0
First DD lesion	4,161	4,053 (97.4)	3 (0.1)	2 (0.1)	21 (0.5)	17 (0.4)	39 (0.9)	26 (0.6)
M1	126	92 (73.0)	1 (0.8)	2 (1.6)	3 (2.4)	2 (1.6)	10 (7.9)	16 (12.7)
M2	437	401 (91.8)	1 (0.2)	0 (0)	6 (1.4)	7 (1.6)	15 (3.4)	7 (1.6)
M3	14	13 (92.9)	0 (0)	0 (0)	0 (0)	0 (0)	1 (7.1)	0 (0)
M4	2,604	2,570 (98.8)	1 (0)	0 (0)	11 (0.4)	8 (0.3)	11 (0.4)	3 (0.1)
M4.1	980	977 (99.7)	0 (0)	0 (0)	1 (0.1)	0 (0)	2 (0.2)	0 (0)
Second DD lesion	483			3 (0.6)	8 (1.7)	91 (18.8)	242 (50.1)	139 (28.8)
M1	200			3 (1.5)	4 (2.0)	8 (4.0)	67 (33.5)	118 (59.0)
M2	121				2 (1.7)	11 (9.1)	95 (78.5)	13 (10.7)
M4	127				1 (0.8)	55 (43.3)	66 (52.0)	5 (3.9)
M4.1	35				1 (2.8)	17 (48.6)	14 (40.0)	3 (8.6)
Third DD lesion	13					4 (30.8)	5 (38.5)	4 (30.8)
M1	5						1 (20.0)	4 (80.0)
M2	5						3 (60.0)	2 (40.0)
M4	3						1 (33.3)	2 (66.7)

<sup>1</sup>Scored according to Berry et al. (2012).

<sup>2</sup>Anatomical areas defined by Relun et al. (2011): BH = between the heel bulbs; D = around the dewclaws; MH = medial heel; LH = lateral heel; F = on the front; HYP = on an interdigital hyperplasia; UF = under the foot in the interdigital space.

<sup>3</sup>Sensitivity = proportion of DD lesions affecting a specific anatomical location identified by milking parlor inspection.

system, and substantial ( $\kappa = 0.62$  and  $\kappa_w = 0.67$ ) for the simplified scoring system. Agreement improved but was still considered substantial ( $\kappa = 0.80$ ) if only the presence of a lesion was considered (Table 4).

Milking parlor inspection as a means of identifying presence of DD lesions had a Se of 92.2% and Sp of 87.6%, with PPV and NPV of 90.9 and 89.3%, respectively (Table 5). True within-herd prevalence was 2, 6, 0, 36, and 14% for M1, M2, M3, M4, and M4.1 lesions, respectively. Test characteristics varied greatly among DD M-stages, with the highest Se for detecting M4 (81.6%) and M2 (61.5%) lesions, and the lowest for detecting M4.1 (20.2%), M3 (0.0%), and M1 (6.7%) lesions. The Se for detecting presence of a lesion when it occurred between the heel bulbs was 93%, but <67% if the lesion occurred elsewhere on the foot (Table 2). The results of test characteristics and comparison of

M-stage pairs between milking parlor and trimming chute inspections specific by farm are included as an Appendix (Tables A1 and A2). The percentage of misclassifications among M-stage pairs and test characteristics varied among farms and stages. For example, 40 to 76% of M4.1 lesions were misclassified as M4 among all farms. Farms with the lowest M4.1 prevalence also had the lowest misclassification percentage (40%). Prevalence of each M-stage affected test characteristics differently on each farm.

## DISCUSSION

Our study apparently included the largest sample of cows used to investigate the accuracy of M-stage scoring for DD in the milking parlor compared with examination in a trimming chute as the gold standard.

**Table 3.** Comparison of M-stage pairs (no., % in parentheses) scoring digital dermatitis (DD) in the milking parlor and in a trimming chute, considering the trimming chute score as the reference category

DD M-stage in milking parlor	DD M-stage <sup>1</sup> in trimming chute						
	M0	M1	M2	M3	M4	M4.1	Total
M0	2,618 (88)	41 (39)	18 (4)	2 (17)	210 (8)	42 (4)	2,931
M1	2 (0)	7 (7)	3 (1)	0 (0)	4 (0)	4 (0)	20
M2	10 (0)	3 (3)	257 (62)	3 (25)	47 (2)	59 (6)	379
M3	8 (0)	0 (0)	1 (0)	0 (0)	41 (2)	8 (1)	58
M4	337 (11)	44 (42)	54 (13)	7 (58)	2,046 (82)	655 (68)	3,143
M4.1	12 (0)	10 (9)	85 (20)	0 (0)	158 (6)	195 (20)	460
Total	2,987	105	418	12	2,506	963	6,991

<sup>1</sup>Scored according to Berry et al. (2012).

**Table 4.** Agreement<sup>1</sup> of digital dermatitis scores attributed during milking parlor and trimming chute inspections

Digital dermatitis stage	Agreement (%)	$\kappa$ (95% CI)	$\kappa_w$ (95% CI)
5 M-stages <sup>2</sup>	73.3	0.59 (0.57–0.60)	0.65 (0.63–0.66)
Simplified system <sup>3</sup>	76.3	0.62 (0.61–0.63)	0.67 (0.65–0.68)
Absence vs. presence <sup>4</sup>	90.2	0.80 (0.79–0.81)	

<sup>1</sup>Kappa ( $\kappa$ ) = agreement beyond chance; weighted kappa ( $\kappa_w$ ) according to order: M2 > M1 > M4.1 > M4 > M3 > M0 (5 M-stages), and active > chronic > no lesion.

<sup>2</sup>M0 to M4.1 (Berry et al., 2012).

<sup>3</sup>Simplified scoring system: no lesion (M0), active lesions (M1/M2/M4.1), and chronic lesions (M3/M4).

<sup>4</sup>Absence (M0) vs. presence (M1 to M4.1).

Scoring DD in the parlor was highly accurate in determining the presence of a lesion, although accuracy varied depending on the M-stage.

The training program to identify and score DD achieved substantial agreement among observers that was maintained throughout the study period. Interobserver reliability estimates in the milking parlor seemed higher than those reported by Relun et al. (2011) in a similar study ( $PA_o = 82\%$ ,  $\kappa = 0.74$ ,  $\kappa_w = 0.76$  in the present study vs.  $PA_o = 66\%$ ,  $\kappa = 0.51$ ,  $\kappa_w = 0.63$  in Relun et al., 2011). However, the number of observers in the present study was lower (3 vs. 5 observers), potentially reducing a source of variation. In addition, intraobserver agreement, another potential source of error, was not evaluated in the present study. Perhaps the effectiveness of our training program was a result of implementing 3 consecutive training sessions that were progressively more detailed (i.e., first scoring with pictures, then live in the milking parlor, and finally in the trimming chute), and encouraging observers to use the reference card at all times during scoring. In addition, it is important to check repeatability in different settings and time points to ensure accurate and reliable data collection throughout the study (Gibbons et al.,

2012). In our view, a training program that successfully obtains high repeatability is essential for achieving reliable data and to prevent misclassification bias.

The cow-level DD prevalence of 68% detected in this study was rather higher than findings in Denmark and France (27%; Thomsen et al., 2008; 45%; Relun et al., 2011), but this was expected, as the study herds were selected based on  $\geq 10\%$  DD prevalence. It was not surprising that most lesions were located between the heel bulbs; however, 10% of DD lesions occurred in the interdigital space, an area not visible during milking parlor inspections. Additionally, when more than one DD lesion was detected, the vast majority of the second and third lesions detected occurred in the interdigital space. These results and the findings of other studies that reported interdigital hyperplasia as a risk factor for DD (Holzhauer et al., 2006; Solano et al., 2016) indicate that early detection and treatment of feet in a trimming chute should not be neglected, especially for cows with a history of interdigital hyperplasia.

Detection of DD in the milking parlor resulted in high test characteristics when determining the presence/absence of a lesion. The Se of 92% seemed higher than findings from any other validation study (72%,

**Table 5.** Test characteristics<sup>1</sup> for scoring each digital dermatitis stage in the milking parlor, considering scoring in the trimming chute as the gold standard (n = 6,991 feet)

Digital dermatitis stage	Prevalence (95% CI)	Se (95% CI)	Sp (95% CI)	PPV (95% CI)	NPV (95% CI)
Absence vs. presence <sup>2</sup>	57.3 (56.1–58.4)	92.2 (91.3–93.0)	87.6 (86.4–88.8)	90.9 (90.0–91.8)	89.3 (88.1–90.4)
M1 <sup>3</sup>	1.5 (1.2–1.8)	6.7 (2.7–13.3)	99.8 (99.7–99.9)	35.0 (15.4–59.2)	98.6 (98.3–98.9)
M2 <sup>3</sup>	6.0 (5.4–6.6)	61.5 (56.6–66.2)	98.1 (97.8–98.5)	67.8 (62.8–72.5)	97.6 (97.2–97.9)
M3 <sup>3</sup>	0.2 (0.1–0.3)	0.0 (0.0–26.5)	99.2 (98.9–99.4)	0.0 (0.0–6.2)	99.8 (99.7–99.9)
M4 <sup>3</sup>	35.8 (34.7–37.0)	81.6 (80.1–83.1)	75.5 (74.3–76.8)	65.1 (63.4–66.8)	88.0 (87.0–89.1)
M4.1 <sup>3</sup>	13.8 (13.0–14.6)	20.2 (17.8–22.9)	95.6 (95.1–96.1)	42.4 (37.8–47.1)	88.2 (87.4–89.0)
Active lesion (M1, M2, M4.1)	21.3 (20.3–22.2)	41.9 (39.4–44.5)	95.7 (95.1–96.2)	72.5 (69.4–75.5)	85.9 (85.0–86.8)
Chronic lesion (M3, M4)	36.0 (34.9–37.2)	83.2 (81.6–84.6)	75.3 (74.0–76.5)	65.4 (63.7–67.1)	88.8 (87.8–89.8)

<sup>1</sup>Sensitivity (Se) = proportion of feet affected with an M-stage of digital dermatitis identified by milking parlor inspection; specificity (Sp) = proportion of feet not affected with an M-stage of digital dermatitis identified by milking parlor inspection; positive predictive value (PPV) = probability that given an M-stage of digital dermatitis identified in the milking parlor, the foot is affected; negative predictive value (NPV) = probability that not given an M-stage of digital dermatitis identified in the milking parlor, the foot is not affected.

<sup>2</sup>Absence (M0) versus presence (M1 to M4.1).

<sup>3</sup>Scored according to Berry et al. (2012).

Rodriguez-Lainz et al., 1998; 67%, Laven, 1999; 65%, Thomsen et al., 2008; 90%, Relun et al., 2011), except Stokes et al. (2012), who reported a Se of 100% in a study involving only 80 cows. Variation of test characteristics among studies could be the consequence of the scoring system or diagnostic tool used, interobserver reliability, length of intervals between test and gold standard evaluation, and DD prevalence. In the present study, the high within-herd mean DD prevalence influenced the test's predictive values. Although Se and Sp are not directly affected by prevalence in the way that predictive values are, they may be prevalence-related (Greiner and Gardner, 2000). Other factors that may have affected the current study's test characteristics included type of milking parlor (i.e., in herringbone parlors, the distance between the observer and one of the hind legs may be relatively long; Thomsen et al., 2008), observation conditions (i.e., poor lighting, quantity of manure contamination, water pressure for cleaning feet), foot conformation (i.e., low claw heels in older cows), different observers between milking parlor and trimming chute inspections, and lesion location (i.e., under or at the front of the foot).

Despite the good accuracy of our tool, detecting only presence/absence of DD does not provide enough information for effectively managing the disease. For instance, early detection of M2 lesions followed by rapid treatment can increase cure rates and reduce the recurrence and spread of infection. Detection of a higher prevalence of M4.1 lesions can be an indicator of upcoming outbreaks, suggesting that prevention measures should be undertaken (e.g., modify footbath program; Döpfer et al., 2012a). Accurate detection of M-stages contributes to the development of preventive, control and therapeutic strategies for DD (Döpfer, 2009). In the present study, the Se for all M-stages except M4 was low (0 to 62%), but the Sp for all stages except M4 was high (96 to 100%). The low Se and high Sp resulted in a moderately low PPV (0 to 67%) but a high NPV (88 to 100%) for each specific M-stage, resulting in a low proportion of false-positive cows. Therefore, milking parlor inspection was a useful herd-level screening tool, as it gave high confidence that a cow that did not exhibit any stage of DD was truly disease-free, although it could also result in unnecessary individual treatments. Cows identified as positive in the milking parlor could be followed closely or selected for further inspection in the trimming chute.

The lower Sp for M4 lesions was due to many false positives. Approximately 10 and 20% of lesions detected as M4 in the milking parlor were truly M0 and M4.1, respectively. Misclassification of M0 as M4 may have occurred because M4 lesions are characterized by thickened epithelium, and some cows could naturally

have a thicker epithelium resembling an M4, or the epithelium could remain thickened after an infection has healed. Additionally, dried manure could be mistaken for scab-like material, which is characteristic of M4 lesions (as well as M3 and M4.1 lesions), when foot inspection is challenged by poor lighting, parlor setting, cow behavior (e.g., constant lifting of feet or kicking), or insufficient washing of feet. Misclassification of M4.1 as M4 could be due to difficulty detecting the small active M1 focus, because it commonly occurs in the foot's most anaerobic areas (e.g., the posterior end of the interdigital cleft). Furthermore, the low Se of the M4.1 stage (20%) due to frequent misclassification as M4 may have serious clinical implications, because these lesions should be treated promptly to prevent aggravation into M2 and potential outbreaks.

The low Se for detecting M2 lesions (62%) and resulting low PPV led to a high proportion of false negatives. Of those, 20% were M4.1 lesions, meaning that the active M1 focus detected in the milking parlor is larger than it appears (M2) when inspected in the trimming chute. Therefore, we inferred that control measures for M4.1 lesions should not be underestimated; once detected in the parlor, they should be monitored closely and potentially be treated as seriously as M2 lesions.

The lowest Se but highest Sp was for the detection of M1 and M3 stages. However, the prevalence of these stages was very low (collectively <2%). In the present study, M3 were commonly misclassified as M4 lesions, and M1 were commonly misclassified as M0 or M4 lesions, consistent with the study of Relun et al. (2011). Misclassification of M3 as M4 lesions was not surprising, due to their similar scab-like characteristics and because M3 is a short-lived stage that can rapidly transition to M4. In agreement with Relun et al. (2011), misclassification frequency and the infectiousness relevance of the stages can justify merging M3 and M4 into 1 category for practical purposes. Misclassification of M1 as M0 could be because M1 lesions were too small to be detected in the milking parlor. It is noteworthy that 10% of all DD lesions occurred in the interdigital space or on an interdigital hyperplasia, an area not visible in the milking parlor and therefore not included in analyses presented in Tables 3 to 5. This was particularly the case for M1 lesions, because the majority (65%) occurred in the interdigital space and went completely undetected during milking parlor inspections.

Potential bias could have been introduced because observers did not use blinding techniques between inspections in the milking parlor and trimming chute. However, it was unlikely that data from >90 cows would be recalled between inspections. In addition, the present study did not record time spent evaluating feet



in the milking parlor (15 s to 2 min; Rodriguez-Lainz et al., 1998; Thomsen et al., 2008; Relun et al., 2011), because this information was outside our objectives. However, farmers consistently reported that observers had no effect on the farms' milking routine or time required for milking. Thus, we believe the practice of scoring feet for DD can be integrated into the milking routine without delaying milking. Unfortunately, our methodology cannot be used easily or safely to detect DD in precalving heifers or in farms with automatic milking systems. Further research is needed on the accurate detection of M-stages and easy monitoring in settings other than a milking parlor.

## CONCLUSIONS

Inspecting the feet of dairy cows in the milking parlor was an inexpensive and simple method of detecting DD lesions. However, this method was not sufficiently reliable to replace thorough foot inspection and definitive identification of M-stages in the trimming chute. If the objective is to follow up DD in cows with a history of interdigital hyperplasia or to detect M1 or M4.1 lesions, this method was not sufficiently reliable. However, if the objective is to determine herd-level DD prevalence and routine monitoring, this method was adequately reliable. Implementation of DD scoring as routine monitoring should aid in early DD detection and prompt treatment interventions. Such interventions can be at the individual level, by topical treatment, or at the herd level, by adjusting the frequency and intensity of foot bathing.

## ACKNOWLEDGMENTS

The authors thank participating farmers and hoof trimmers Elbert Koster and Rob Geier. We are especially grateful for the contributions of Charlotte Pickel (University of Calgary, AB, Canada), Ilse Geurts (Wageningen University, Gelderland, the Netherlands), Mart Wolbers, Fokke Sikkema (Utrecht University, Utrecht, the Netherlands), and Miranda Marchand (University of British Columbia, Vancouver, BC, Canada) in data collection. We also thank John Kastelic (Department of Production Animal Health, University of Calgary, AB, Canada) for editing this manuscript. This study was funded by Alberta Milk (Edmonton, AB, Canada), and the Alberta Livestock and Meat Agency (Edmonton, AB, Canada).

## REFERENCES

Berry, S. L., D. H. Read, T. R. Famula, A. Mongini, and D. Döpfer. 2012. Long-term observations on the dynamics of bovine digital

- dermatitis lesions on a California dairy after topical treatment with lincomycin HCl. *Vet. J.* 193:654–658.
- Bruijnijis, M. R. N., B. Beerda, H. Hogeveen, and E. N. Stassen. 2012. Assessing the welfare impact of foot disorders in dairy cattle by a modeling approach. *Animal* 6:962–970.
- Bruijnijis, M. R. N., H. Hogeveen, and E. N. Stassen. 2010. Assessing economic consequences of foot disorders in dairy cattle using a dynamic stochastic simulation model. *J. Dairy Sci.* 93:2419–2432.
- Cha, E., J. A. Hertl, D. Bar, and Y. T. Grohn. 2010. The cost of different types of lameness in dairy cows calculated by dynamic programming. *Prev. Vet. Med.* 97:1–8.
- Cheli, R., and C. Mortellaro. 1974. La dermatite digitale del bovino. Pages 208–213 in *Proc. 8th Int. Conf. Dis. Cattle*, Piacenza, Milan, Italy. Tipografia Editoriale Piacentina Gallarati, Piacenza, Italy.
- Cohen, J. 1960. A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.* 20:37–46.
- Cohen, J. 1968. Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychol. Bull.* 70:213–220.
- Cramer, G., K. D. Lissemore, C. L. Guard, K. E. Leslie, and D. F. Kellton. 2008. Herd- and cow-level prevalence of foot lesions in Ontario dairy cattle. *J. Dairy Sci.* 91:3888–3895.
- Dohoo, I., W. Martin, and H. Stryhn. 2009. *Veterinary Epidemiologic Research*. 2nd ed. VER Inc., Charlottetown, PE, Canada.
- Döpfer, D. 2009. The dynamics of digital dermatitis in dairy cattle and the manageable state of disease. Pages 1–5 in *Proc. CanWest Veterinary Conference*, Banff, AB, Canada. <http://www.hoofhealth.ca/Dopfer.pdf>.
- Döpfer, D., K. Anklam, D. Mikheil, and P. Ladell. 2012b. Growth curves and morphology of three *Treponema* subtypes isolated from digital dermatitis in cattle. *Vet. J.* 193:685–693.
- Döpfer, D., M. Holzhauer, and M. van Boven. 2012a. The dynamics of digital dermatitis in populations of dairy cattle: Model-based estimates of transition rates and implications for control. *Vet. J.* 193:648–653.
- Döpfer, D., A. A. H. M. ter Huurne, J. L. Cornelisse, A. J. A. M. van Asten, A. Koopmans, F. A. Meijer, Y. H. Schukken, I. Szakáll, W. Klee, and R. B. Bosma. 1997. Histological and bacteriological evaluation of digital dermatitis in cattle, with special reference to spirochaetes and *Campylobacter faecalis*. *Vet. Rec.* 140:620–623.
- Fleiss, J. L., B. Levin, and M. C. Paik. 2003. *Statistical Methods for Rates and Proportions*. J. Wiley & Sons, Hoboken, NJ.
- Gibbons, J., E. Vasseur, J. Rushen, and A. M. de Passillé. 2012. A training programme to ensure high repeatability of injury scoring of dairy cows. *Anim. Welf.* 21:379–388.
- Gomez, A., K. S. Anklam, N. B. Cook, J. Rieman, K. A. Dunbar, K. E. Cooley, M. T. Socha, and D. Döpfer. 2014. Immune response against *Treponema* spp. and ELISA detection of digital dermatitis. *J. Dairy Sci.* 97:4864–4875.
- Gomez, A., N. B. Cook, N. D. Bernardoni, J. Rieman, A. F. Dusick, R. Hartshorn, M. T. Socha, D. H. Read, and D. Döpfer. 2012. An experimental infection model to induce digital dermatitis infection in cattle. *J. Dairy Sci.* 95:1821–1830.
- Gomez, A., N. B. Cook, J. Rieman, K. A. Dunbar, K. E. Cooley, M. T. Socha, and D. Döpfer. 2015a. The effect of digital dermatitis on hoof conformation. *J. Dairy Sci.* 98:927–936.
- Gomez, A., N. B. Cook, M. T. Socha, and D. Döpfer. 2015b. First-lactation performance in cows affected by digital dermatitis during the rearing period. *J. Dairy Sci.* 98:4487–4498.
- Greenough, P. R., C. K. W. Muelling, D. Döpfer, and D. J. Tomlinson. 2008. International atlas of lesions of cattle feet. Nomenclature and atlas update. Page 40 in *Proc. 15th Int. Symp. 7th Conf. Lameness Ruminants*, Kuopio, Finland. Savonia University of Applied Sciences, Kuopio, Finland.
- Greiner, M., and I. A. Gardner. 2000. Epidemiologic issues in the validation of veterinary diagnostic tests. *Prev. Vet. Med.* 45:3–22.
- Holzhauer, M., C. Hardenberg, C. J. M. Bartels, and K. Frankena. 2006. Herd- and cow-level prevalence of digital dermatitis in the Netherlands and associated risk factors. *J. Dairy Sci.* 89:580–588.
- Krull, A. C., V. L. Cooper, J. W. Coatney, J. K. Shearer, P. J. Gorden, and P. J. Plummer. 2016. A highly effective protocol for the rapid

- and consistent induction of digital dermatitis in Holstein calves. *PLoS One* 11:e0154481.
- Krull, A. C., J. K. Shearer, P. J. Gorden, V. L. Cooper, G. J. Phillips, and P. J. Plummer. 2014. Deep sequencing analysis reveals temporal microbiota changes associated with development of bovine digital dermatitis. *Infect. Immun.* 82:3359–3373.
- Landis, J. R., and G. G. Koch. 1977. The measurement of observer agreement for categorical data. *Biometrics* 33:159–174.
- Laven, R. A. 1999. The environment and digital dermatitis. *Cattle Pract.* 7:349–354.
- Manske, T., J. Hultgren, and C. Bergsten. 2002. Topical treatment of digital dermatitis associated with severe heel-horn erosion in a Swedish dairy herd. *Prev. Vet. Med.* 53:215–231.
- Nielsen, B. H., P. T. Thomsen, L. E. Green, and J. Kaler. 2012. A study of the dynamics of digital dermatitis in 742 lactating dairy cows. *Prev. Vet. Med.* 104:44–52.
- Relun, A., R. Guatteo, P. Roussel, and N. Bareille. 2011. A simple method to score digital dermatitis in dairy cows in the milking parlor. *J. Dairy Sci.* 94:5424–5434.
- Rodriguez-Lainz, A., P. Melendez-Retamal, D. W. Hird, and D. H. Read. 1998. Papillomatous digital dermatitis in Chilean dairies and evaluation of a screening method. *Prev. Vet. Med.* 37:197–207.
- Shearer, J., and S. Van Amstel. 2013. *Manual of Foot Care in Cattle*. 2nd ed. Hoard's Dairyman, Fort Atkinson, WI.
- Solano, L., H. W. Barkema, E. A. Pajor, S. Mason, S. J. LeBlanc, and K. Orsel. 2016. Prevalence and distribution of foot lesions in dairy cattle in Alberta, Canada. *J. Dairy Sci.* 99:6828–6841. <https://doi.org/10.3168/jds.2016-10941>.
- Solano, L., H. W. Barkema, C. Pickel, and K. Orsel. 2017. Effectiveness of a standardized footbath protocol for prevention of digital dermatitis. *J. Dairy Sci.* 100:1295–1307. <https://doi.org/10.3168/jds.2016-11464>.
- Stokes, J. E., K. A. Leach, D. C. J. Main, and H. R. Whay. 2012. The reliability of detecting digital dermatitis in the milking parlour. *Vet. J.* 193:679–684.
- Thomsen, P. T., I. C. Klaas, and K. Bach. 2008. Scoring of digital dermatitis during milking as an alternative to scoring in a hoof trimming chute. *J. Dairy Sci.* 91:4679–4682.
- University of Wisconsin. 2013. DDPen walk. Accessed Sep.9, 2016. <http://www.vetmed.wisc.edu/dms/fapm/apps/ddpw.htm>.
- USDA. 2009. Dairy 2007, Part IV: Reference of dairy cattle health and management practices in the United States. USDA-APHIS-VS-CEAH, Fort Collins, CO. #N494.0209. Accessed Sep. 9, 2016. [https://www.aphis.usda.gov/animal\\_health/nahms/dairy/downloads/dairy07/Dairy07\\_dr\\_PartIV.pdf](https://www.aphis.usda.gov/animal_health/nahms/dairy/downloads/dairy07/Dairy07_dr_PartIV.pdf).
- Vink, W. D. 2006. Investigating the epidemiology of bovine digital dermatitis: causality, transmission and infection dynamics. PhD Thesis. University of Liverpool, UK.

## APPENDIX

**Table A1.** Comparison of M-stage pairs (no., % in parentheses) scoring the most frequent digital dermatitis (DD) stage in the milking parlor and in a trimming chute, considering the trimming chute score as the reference category for all farms<sup>1</sup>

DD M-stage in parlor	Farm 1				Farm 2				Farm 3			
	DD M-stage <sup>2</sup> in trimming chute				DD M-stage in trimming chute				DD M-stage in trimming chute			
	M0	M2	M4	M4.1	M0	M2	M4	M4.1	M0	M2	M4	M4.1
M0	291 (89)	5 (9)	25 (8)	14 (8)	406 (95)	0 (0)	3 (7)	0 (0)	209 (90)	0 (0)	37 (13)	1 (2)
M2	1 (0)	26 (46)	2 (1)	9 (5)	1 (0)	2 (67)	0 (0)	2 (40)	0 (0)	8 (100)	6 (2)	10 (25)
M4	36 (11)	5 (9)	263 (83)	108 (62)	19 (5)	0 (0)	37 (88)	2 (40)	19 (8)	0 (0)	210 (71)	16 (40)
M4.1	1 (0)	20 (36)	24 (8)	41 (24)	0 (0)	1 (33)	2 (5)	1 (20)	1 (0)	0 (0)	40 (14)	13 (33)
	Farm 4				Farm 5				Farm 6			
M0	303 (87)	6 (4)	36 (15)	13 (12)	203 (84)	2 (3)	14 (4)	2 (1)	359 (90)	1 (3)	32 (8)	5 (1)
M2	1 (0)	107 (78)	1 (0)	6 (6)	0 (0)	36 (50)	4 (1)	5 (3)	3 (1)	24 (75)	28 (7)	18 (9)
M4	40 (12)	9 (7)	194 (81)	66 (63)	36 (15)	14 (19)	323 (85)	113 (70)	31 (8)	5 (16)	303 (73)	146 (76)
M4.1	4 (1)	12 (9)	9 (4)	19 (18)	2 (1)	20 (28)	27 (7)	36 (22)	1 (0)	2 (6)	35 (9)	24 (13)
	Farm 7				Farm 8				Farm 9			
M0	170 (77)	2 (5)	10 (3)	1 (1)	222 (79)	0 (0)	16 (9)	4 (6)	455 (90)	2 (4)	37 (12)	5 (5)
M2	1 (1)	24 (62)	1 (0)	2 (1)	1 (0)	5 (23)	2 (1)	3 (4)	2 (0)	25 (52)	3 (1)	4 (4)
M4	48 (21)	6 (15)	317 (94)	84 (74)	60 (21)	8 (36)	144 (83)	52 (71)	48 (10)	7 (15)	255 (83)	68 (69)
M4.1	1 (1)	7 (18)	4 (1)	27 (24)	0 (0)	9 (41)	8 (5)	13 (18)	2 (0)	14 (29)	9 (3)	21 (21)

<sup>1</sup>Percentages may add to <100% because DD M-stages with low frequency ( $\leq 2\%$ ) were excluded.

<sup>2</sup>M0 to M4.1 scored according to Berry et al. (2012).

**Table A2.** Test characteristics<sup>1</sup> for the most frequent digital dermatitis M-stages scored in the milking parlor, compared with scoring in the trimming chute as the gold standard for all farms and by farm

Farm	M0 <sup>2</sup>						M2						M4						M4.1							
	Prev	Se	Sp	PPV	NPV		Prev	Se	Sp	PPV	NPV	Prev	Se	Sp	PPV	NPV	Prev	Se	Sp	PPV	NPV	Prev	Se	Sp	PPV	NPV
1	37	88	91	85	93		6	46	99	68	96	36	83	73	63	89	20	24	94	47	84	20	24	94	47	84
2	89	95	92	99	71		1	67	99	40	100	9	88	95	64	99	1	20	99	25	99	1	20	99	25	99
3	40	90	89	84	93		2	100	97	31	100	51	71	88	86	75	7	33	92	24	95	7	33	92	24	95
4	40	87	87	81	91		16	78	99	92	96	28	81	79	60	91	12	18	96	40	90	12	18	96	40	90
5	28	84	97	91	94		8	50	99	78	96	44	85	65	65	84	19	22	93	41	84	19	22	93	41	84
6	38	90	93	89	94		3	75	95	32	99	39	73	70	62	80	18	13	96	38	83	18	13	96	38	83
7	31	77	97	92	91		5	62	99	86	98	47	94	63	69	92	16	24	98	68	87	16	24	98	68	87
8	50	78	93	91	81		4	23	99	42	97	31	83	67	53	90	13	18	96	42	89	13	18	96	42	89
9	53	90	90	91	89		5	52	99	74	98	32	83	81	67	91	10	21	97	45	92	10	21	97	45	92
All farms	43	88	92	89	91		6	62	98	68	98	36	82	76	65	88	14	20	96	42	88	14	20	96	42	88

<sup>1</sup>Prev = prevalence of an M-stage; sensitivity (Se) = proportion of feet affected with an M-stage of digital dermatitis identified by milking parlor inspection; specificity (Sp) = proportion of feet not affected with an M-stage of digital dermatitis identified by milking parlor inspection; positive predictive value (PPV) = probability that given an M-stage of digital dermatitis identified in the milking parlor, the foot is affected; negative predictive value (NPV) = probability that not given an M-stage of digital dermatitis identified in the milking parlor, the foot is not affected.

<sup>2</sup>M0 to M4.1 scored according to Berry et al. (2012).