Technical note: A digital technique and platform for assessing dairy cow teat-end condition

P. S. Basran,^{1*} M. Wieland,² and I. R. Porter¹

¹Department of Clinical Sciences, College of Veterinary Medicine, Cornell University, Ithaca, NY 14853

²Department of Population Medicine and Diagnostic Sciences, College of Veterinary Medicine, Cornell University, Ithaca, NY 14853

ABSTRACT

Because infections with pathogenic bacteria entering the mammary gland through the teat canal are the most common cause of mastitis in dairy cows, sustaining the integrity of the teat canal and its adjacent tissues is critical to resist infection. The ability to monitor teat tissue condition is therefore a key prerequisite for udder health management in dairy cows. However, to date, routine assessment of teat-end condition is limited to cow-side visual inspection, making the evaluation a time-consuming and expensive process. Here, we illustrate and demonstrate a method for assessing teat-end condition of dairy cows through digital images and software. A digital workflow has been designed where images of dairy cow teats are obtained and processed to display individual teats, and the cow and teat images are labeled and displayed through a graphical user interface. The interface then allows an evaluator to assess quarter- and cow-level teat-end condition and store the results for review and future analysis. The digital workflow permits several advantages such as the ability to perform remote teat-end condition assessments, and assess inter- and intrarater variability of teat-end condition scoring. We demonstrate the image-based teat-end condition assessment of 194 dairy cows that also had cow-side teat-end condition assessments by 2 expert evaluators. Weighted Cohen's kappa statistic (κ) was computed to measure the evaluators' concordance of categorical scores of quarter- and cow-level assessments when using cow-side and image-based assessments. Substantial agreement ($0.61 \leq \kappa \leq 0.80$) was observed between an evaluator's cow-side and image-based assessments at the quarter and cow level. Moderate agreement ($0.41 \leq \kappa \leq 0.60$) was observed between evaluators when using image-based assessments at the quarter and cow level. Near perfect agreement ($\kappa = 0.89$, 95% confidence interval 0.78–1.00) was observed

between evaluators when using cow-side assessments at the quarter level, and substantial agreement ($\kappa = 0.66$, 95% confidence interval 0.53–0.79) was observed when using cow-side assessments at the cow level. This suggests that image-based teat-end condition classification is possible, and coupled with improvements in image acquisition and image processing, this method can be used to assess teat-end condition in a systematic and convenient manner.

Key words: teat condition, hyperkeratosis, mastitis, digital imaging, reproducibility

Technical Note

Implementation of mastitis control programs has been the focus of a great number of studies in recent decades. Yet mastitis continues to be the most common and costly disease of dairy cows worldwide, accounting for the majority of antimicrobial treatment in the US dairy industry (Pol and Ruegg, 2007; Ruegg, 2012). The majority of mastitis cases are caused by IMI, where pathogenic bacteria enter the mammary gland through the teat canal. The teat canal is therefore considered the cow's first line of defense against mastitis pathogens (O'Shea, 1987). Sustaining teat canal integrity and pliability of its adjacent tissues is critical to resist IMI, and monitoring teat tissue condition in an efficient and timely manner is fundamental in udder health management. However, routine assessment of, for example, milking machine-induced long-term changes to the teat tissue (i.e., teat-end callosity thickness and roughness; Neijenhuis et al., 2000) is limited to cow-side inspection of all 4 teats. Current industry guidelines recommend evaluating at least 80 cows, or a minimum of 20% of cows in herds larger than 400 cows (Mein et al., 2001), which makes the evaluation a laborious and cost-intensive process. This hampers the ability to routinely monitor teat tissue condition and develop alert systems on dairy operations that detect changes in teat condition after, for instance, a change in milking liner type or vacuum settings. Another limitation of observation-based scoring systems is

Received April 6, 2020.

Accepted June 10, 2020.

*Corresponding author: psb92@cornell.edu

Figure 1. The digital workflow for assessing dairy cow health through image-based teat-end condition assessment through the collection of udder images (A), image cropping to segment the udders from the cow image (B), teat-end segmentation (C), a graphical user interface for classifying teat-end condition by an evaluator (D), and digital storage of results for epidemiological analysis (E). Repeating the process (dashed line) provides opportunities for trend analysis.

a reportedly low inter- and intrarater agreement, which can lead to uncertainties about the importance of teat-end condition. As a result, there is an industry need to improve methods of assessing teat-end condition in dairy cows.

One possible approach for assessing teat tissue condition in dairy cows is through the use of digital imaging. Compact and inexpensive digital camera technologies could be used to capture images and videos of the teats of dairy cows. If images are of sufficient image quality and can permit a reliable means of assessment, the images could be reviewed remotely, and a higher number of cows in a herd could be assessed. This approach could be valuable in larger dairy farms with large cow numbers and where various levels of automation are already in place.

Our research focus is improving the efficiency and accuracy of dairy health management through automation. We hypothesize that image-based evaluation of teat-end condition is possible and is an alternative to traditional cow-side evaluation. The purpose of this technical report is to illustrate a workflow and software platform for digital image-based evaluation of teat-end condition. We describe the image collection, the image processing steps required for evaluation, and a simple graphical user interface (**GUI**) that facilitates an image-based assessment of teat condition. We then demonstrate its utility by assessing the agreement between 2 experienced evaluators and image-based and cow-side assessments, respectively.

Figure 1 illustrates the digital workflow for the image-based teat-end condition scoring system. First, images are acquired in a safe and consistent manner. Images of the teats are obtained as close as reasonably possible to mitigate collision with milking equipment, parlor stalls, ancillary support systems and equipment, and the dairy cows themselves. Images are acquired in a systematic manner such that the dairy cow and corresponding teats can be corroborated with an image and

that the field of view of the image captures as many teats as possible. Second, images are downloaded from the camera to a computer, and the raw image data are processed and curated such that a cow can be identified for review. Third, individual teats are identified in the image, and the original image data are subsampled by cropping images of each teat such that an assessment of teat-end condition is possible (Figure 1C). Fourth, the digital images of the 4 teats are classified using a 4-point ordinal scoring system through a GUI, and quarter- and cow-level teat-end condition scores are assessed and saved digitally. Fifth, statistical analyses are performed with the data and archived such that if the entire process is repeated at another time point, trends in the population can be quantified (see dashed line in Figures 1E to 1A).

The feasibility assessment of this workflow was carried out on a commercial dairy farm that belongs to the client base of Quality Milk Production Services (Cornell University, Ithaca, NY) milking approximately 1,600 cows in a rotary parlor. All procedures were reviewed and approved by the Cornell University Institutional Animal Care and Use Committee (protocol-no. 2013-0064). For the purpose of this demonstration, we included cows with no udder abnormalities such as nonlactating quarters or teat injuries. Assessment of quarter-level teat-end condition of 194 cows was performed after machine milking by 2 experienced evaluators (evaluators 1 and 2 hereafter identified as **E1** and **E2**). To increase visibility of the teat end, teats were gently cleaned with a cloth towel to reduce the amount of milk droppings on the teats' orifice before scoring. Thereafter, teat-end conditions of the right front (**RF**), left front (**LF**), right hind (**RH**), and left hind (**LH**) teats were assessed according to the 4-point scoring system as previously described (Mein et al., 2001). Briefly, a score of 1 = no callosity ring present; 2 = callosity ring but no roughness present; 3 = callosity ring and roughness present, keratin fronds extending 1 to 3 mm

from the teat orifice; and 4 = callosity ring present with excessive keratin fronds extending ≥ 4 mm from the teat orifice.

Using a digital camera (iPhone6+, Apple Inc., Cupertino, CA), each cow who had teat scoring performed had an image of the udder taken by a third investigator. Digital images were acquired with the highest resolution ($3,024 \times 3,024$ pixels, RGB with 72×72 resolution) and all 194 images were exported from the smart phone in .jpg format at the original resolution. Images ranged in size from 4 to 9 MB depending on the magnification factor used.

Cows were cataloged according to the parlor stall number (ranging from 1 to 60) and the round number (each round starting and stopping on the first and 60th stall); the time for the first stall scored in each parlor round was also recorded. To facilitate assignment of cows, E1 identified cows that met the inclusion criteria and communicated the respective stall number with E2 and the investigator capturing the digital images. Images of the stall number and the cow teats were taken at the same time as the cow-side assessments. The round number and time were recorded such that each image could be matched with the cow's visual assessment. Blinding of E1 and E2 was ensured by documentation on separate sheets and performing the scoring approximately 5 m apart from each other.

All digital image processing and GUI design were performed in a Matlab environment (version 2019a, MathWorks, Natick, MA). Because the original images contained more image information than necessary, all 194 images were manually cropped to display only the udder and teats. The cropped image was used to establish a frame of reference for the evaluator (Figure 1B). Each image was renamed according to the round and stall number. Software was then written to automatically load a cow image and permit the user to crop the RF, LF, RH, and LH teats (Figure 1C). For images where teats were immediately adjacent to each other, the teat image was cropped such that the teat of interest was in the center of the image. When a teat end was not visible, a blank image was created. The cropped images were then exported to file with the round number, stall number, and teat position appended to the original filename. A total of 776 individually labeled teat images were created.

The GUI imports the images of a dairy cow and provides a means of assessing teat condition (Figure 2). An evaluator starts the application (Figure 2A, identified in the GUI as operator), and opens an image of a dairy cow (Figures 2B and 2D). The individually cropped teat-end images are then also automatically retrieved (Figure 2E displays the RF teat). A frame of reference is achieved by displaying the udder images along with

the 4 teat images. The evaluator then uses drop-down selections to score each teat. Pigmentation status is also recorded (Figures 2F and 2G).

Each teat end is classified with the 4-point scoring system as described above. We also included an option for "incomplete data" because some teat images are not visible due to obfuscation from anatomy or environment, poor image quality (e.g., poor lighting, insufficient zoom, unfocused image, inaccurate cropping), or milk dropping obfuscation of the teat end. A cow-level score of "all pass," "partial pass," "partial fail," and "all fail" is coded in the software based on the quarter-level teat-end condition scores. If any teat score is 3 or higher, an "all fail" is recorded for the cow-level score, and if all teat scores are 2 or lower, an "all pass" is recorded (Figure 2C). To accommodate the scenario where 1 or more teat images are classified as "incomplete," 2 additional cow-level scores of "partial pass" and "partial fail" are provided. After scoring each teat image, a tab-delimited text file is written and updated, recording the identification of the cow, the quarter-level teat-end condition assessments including pigmentation status, as well as the cow-level score.

Of the 194 dairy cows, data from 190 cows were available. Teat scoring assessments from 4 cows were not analyzed due to mismatching errors, resulting in a total of 760 teats evaluated through cow-side evaluations. Of the 190 dairy cows, the assessments from 59 cows were unblinded for quality assurance of interrater scoring. Here, scores obtained by E1 and E2 were discussed to ensure consistency in the scoring methodology for the subsequent unblinded teat-end assessments. Therefore, a total of 131 of the 194 (69%) of all dairy cows assessments and 524 cow-side teat assessments were blinded. To compare cow-side assessments with image-based evaluations at the cow level, we created a new variable in accordance with the image-based cow-level evaluation. An "all pass" at the cow level was considered if all 4 teats had a score of 1 or 2 during the cow-side assessment, whereas an "all fail" was considered if 1 or more teats had a score of 3 or 4 during the cow-side assessment.

Of the 194 dairy cows, image-based assessments were possible for 193 cows. Data from 1 cow was not analyzed due to incomplete image data. The same 2 evaluators (E1 and E2) scored 193 cows using the GUI. For the image-based assessments, no observations were used as "training" data, and 722 teat-end assessments were performed. Of the 772 teats, E1 and E2 provided a classification of "incomplete data" in 35 and 40 teat images, respectively, or were able to score 93.3%, and 92.4% of all teats, respectively. In 22 of the 193 cows, both evaluators were unable to score 1 or more teats. Thus, a total of 193 of the 194 (99%) of all dairy cows

teat-end condition with a higher score in both cow-side and image-based assessments.

Weighted Cohen’s kappa statistics were computed to measure the concordance of categorical scores of quarter- and cow-level assessments (score 1 through 4) for both evaluators and for the cow-side and image-based scores, and are displayed in Table 1 (Cardillo, 2007). Kappa results of 0.10 to 0.20 are interpreted as slight agreement, 0.21–0.40 as fair, 0.41–0.60 as moderate, 0.61–0.80 as substantial, and 0.81–1.00 as near perfect agreement. Moderate agreement was observed with the image-based quarter-level assessments and near perfect agreement with cow-side evaluations. Moderate agreement was observed with the image-based cow-level assessments, and substantial agreement for cow-side assessments, but, arguably, these 2 results are within error. Although not the focus of this technical report, this analysis illustrates challenges of interrater variability in quarter- and cow-level assessments. Our findings are

Figure 3 displays a histogram of the relative percentage of teat-end scores ($n = 524$) for the 2 evaluators using cow-side and image-based quarter-level assessments, respectively. Note the relatively consistent evaluations for scores 3 and 4 for E1 and E2, and less consistent evaluations for scores 1 and 2 when using either cow-side or image-based assessments. Differences in the assessments may be explained by inter- and intrarater error. Compared with E2, E1 was more likely to classify

Figure 2. The graphical user interface for assessing teat-end condition for a dairy cow. The evaluator (A) loads an image file (B) that displays the teats and udder (D), along with individually cropped teats (E) using the standard frame of reference. Each teat is then scored (F) along with the pigmentation status (G), and a cow-level evaluation is produced (C). The upper right section permits the evaluator to load, save the results, and quit. RF = right front; LF = left front; RH = right hind; and LH = left hind. Q-score is the quarter-level classification of the teat-end condition using a 4-point scoring system, along with option for “Incomplete Data,” as described in the text.

Table 1. The degree of agreement between quarter-level and cow-level teat-end cow-side and image-based assessments for 524 teat ends (quarter level) or 131 dairy cows¹

Item	κ -Value	95% CI	
		Lower	Upper
Evaluator 1 vs. evaluator 2			
Quarter level-image-based	0.56	0.42	0.70
Quarter level-cow-side	0.89	0.78	1.00
Cow level-image-based	0.58	0.44	0.72
Cow level-cow-side	0.66	0.53	0.79
Image-based vs. cow-side			
Quarter level-evaluator 1	0.66	0.52	0.81
Quarter level-evaluator 2	0.72	0.66	0.78
Cow level-evaluator 1	0.74	0.62	0.86
Cow level-evaluator 2	0.77	0.67	0.88

¹Note that the weighted Cohen's kappa statistic (κ) values ranging from 0.41 to 0.60, 0.61 to 0.80, and 0.81 to 1.00 are considered to be in moderate, substantial, and near perfect agreement, respectively, between 2 observers or methods.

consistent with those from Douphrate et al. (2019) who investigated interrater agreement for the assessment of teat hygiene scoring and found fair agreement between different raters (overall kappa statistic: 0.43).

Table 1 also displays the kappa statistics for both evaluators when performing a cow-side or image-based teat-end condition assessment. Substantial agreement was observed between evaluators for quarter-level and cow-level teat-end scores when using either the cow-side or image-based method. For image-based cow-level

assessments, the results from “pass” and “partial pass” are combined and compared with cow-side “pass” and “fail” assessments, and “fail” and “partial fail” are also combined for the computation of confusion matrices and subsequent kappa statistics. This finding suggests that image-based teat-end condition assessments could replace cow-side evaluations.

We have described a data collection strategy, the image processing steps required for evaluation, and a simple GUI that can be used to facilitate an image-based assessment of teat-end condition. We then demonstrated its utility by comparing image-based and cow-side assessments of teat-end condition by 2 evaluators, and compared image-based and cow-side teat-end assessments. This work has several shortcomings and opportunities. First, the image acquisition step has great room for improvement. Image acquisition is most convenient during the milking process. If images are acquired before milking, care must be taken to ensure that the teats are sufficiently cleaned for an image-based assessment. If done after, care must be taken to ensure that milk droppings do not obfuscate the teat. The image system must provide high-resolution images such that fine details can be revealed, and the imaging system must be robust enough to withstand typical environmental conditions. Multiple cameras focused on all teats will increase the number of viewpoints and hence the amount of information for the image-based assessment, and could provide an opportunity to view all teats irrespective of their orientation and locations relative to each other. Improvements in lighting will also be beneficial. Second, the image processing and teat identification process has room for improvement. The image cropping and image preparation steps in this work were manual, time consuming, and prone to human error. However, this process can be automated with image detection and segmentation algorithms commonly used in machine learning: this is the focus of our current research. Finally, numerous improvements could be made in the GUI design and functionality. In the current design, the evaluator is required to manually select images. This process is also prone to human error as an evaluator may evaluate the same cow or skip cows altogether. No significant effort was directed toward optimizing the software, deploying it for different computer environments, such as a stand-alone executable program, or improving functionality and operability.

Despite these challenges, using an image-based teat-end condition scoring system has several advantages and opportunities. First, the important question of inter- and intrarater variability can be explored in a systematic manner. Inter- and intrarater agreement experiments are challenging to perform in the field. In

Figure 3. The relative percentage of 524 teat-end scores as a function of cow-side and image-based evaluations from the 2 evaluators. 1 = no callosity ring present; 2 = callosity ring but no roughness present; 3 = callosity ring and roughness present, keratin fronds extending 1 to 3 mm from the teat orifice; and 4 = callosity ring present with excessive keratin fronds extending ≥ 4 mm from the teat orifice. Note that score N shows the number of images/teats that could not be scored.

conjunction with cow-side assessments, image-based assessments can be used to supplement inter- and intra-rater agreement studies, as demonstrated in this technical note. Matching assessments with a specific dairy cow parlor stall and round number was achieved by carefully documenting the time, stall, and round number. This matching raises the possibility of identifying cows and their corresponding health assessment, and thus a more targeted approach to dairy health. Consensus-based assessments can be explored where multiple observers can review images and collectively evaluate quarter- and cow-level scores. Upon completion of a cow-side assessment, users can perform image-based assessments while reviewing cow-side assessments made by themselves and their colleagues. When used in this way, image-based assessments become an educational tool. Second, and more importantly, solely image-based assessments may provide the opportunity for remote teat health assessments. Although cow-side and image-based assessments have yet to be proven to be comparable, remote teat condition scoring has the potential to reduce costs and inconvenience in field-based assessments. Third, the digitization of the entire process provides vast opportunities to adopt “big-data” and artificial intelligence approaches for improving the quality of health of dairy cows. For example, image-based classification data can be curated, and the process can be repeated at different time points for analyzing epidemiological trends for a population. Deep learning algorithms can be easily adopted to train a neural network for automated teat condition classification to further automate the process.

To conclude, we illustrated and demonstrated a simple and effective method for assessing teat-end condition through digital images and software. We have found that image-based teat-end condition classification is possible, and that the method can be used to assess teat-end condition in a systematic manner. Our preliminary data suggest that digital image assessment of teat-ends may be an acceptable alternative to time-intensive and costly cow-side evaluation for teat-end condition in dairy cattle. The workflow described in this manuscript describes how a GUI streamlines the digital classification procedure. In the future, optimization and implementation of automated digital image

evaluation of teat condition offers the potential to regularly and efficiently monitor teat-end condition and therefore prevent mastitis by allowing implementation of corrective actions in a more timely fashion.

ACKNOWLEDGMENTS

The authors thank the farm owners and their employees for their willingness to participate in the study and their support during data collection. We gratefully acknowledge the technical assistance from Belinda Gross (Quality Milk Production Services, Cornell University, Ithaca, NY) with data entry. The authors have not stated any conflicts of interest.

REFERENCES

- Cardillo, G. 2007. Cohen's kappa: Compute the Cohen's kappa ratio on a square matrix. <http://www.mathworks.com/matlabcentral/fileexchange/15365>.
- Douphrate, D. I., N. B. Fethke, M. W. Nonnenmann, A. Rodriguez, and D. Gimeno Ruiz de Porras. 2019. Reliability of observational- and machine-based teat hygiene scoring methodologies. *J. Dairy Sci.* 102:7494–7502. <https://doi.org/10.3168/jds.2019-16351>.
- Mein, G. A., F. Neijenhuis, W. F. Morgan, D. J. Reinemann, J. E. Hillerton, J. R. Baines, I. Ohnstad, M. D. Rasmussen, L. Timms, J. S. Britt, R. Farnsworth, N. Cook, and T. Hemling. 2001. Evaluation of bovine teat condition in commercial dairy herds: 1. Non-infectious factors. Pages 347–351 in *Proc. 2nd International Symposium on Mastitis and Milk Quality*. National Mastitis Council Inc., Vancouver, BC, Canada.
- Neijenhuis, F., H. W. Barkema, H. Hogeveen, and J. P. Noordhuizen. 2000. Classification and longitudinal examination of callused teat ends in dairy cows. *J. Dairy Sci.* 83:2795–2804. [https://doi.org/10.3168/jds.S0022-0302\(00\)75177-0](https://doi.org/10.3168/jds.S0022-0302(00)75177-0).
- O'Shea, J. 1987. Machine milking and mastitis. Section 2: Machine milking factors affecting mastitis. A literature review. *Bull. Int. Dairy Fed.* 215:5–32.
- Pol, M., and P. L. Ruegg. 2007. Relationship between antimicrobial drug usage and antimicrobial susceptibility of gram-positive mastitis pathogens. *J. Dairy Sci.* 90:262–273. [https://doi.org/10.3168/jds.S0022-0302\(07\)72627-9](https://doi.org/10.3168/jds.S0022-0302(07)72627-9).
- Ruegg, P. L. 2012. Mastitis in dairy cows. *Vet. Clin. North Am. Food Anim. Pract.* 28:xi–xii. <https://doi.org/10.1016/j.cvfa.2012.04.003>.

ORCIDS

- P. S. Basran <https://orcid.org/0000-0002-1573-1549>
- M. Wieland <https://orcid.org/0000-0003-0513-1782>
- I. R. Porter <https://orcid.org/0000-0002-5501-4683>