Noninferiority trial on the efficacy of premilking teat disinfectant against naturally occurring new intramammary infections using a novel 2-step diagnostic process

A. Ceballos-Marquez,*† T. Hemling,‡ B. J. Rauch,* M. Lopez-Benavides,‡ and Y. H. Schukken*§1

*Department of Population Medicine and Diagnostic Sciences, Cornell University, Ithaca, NY 14853
†Grupo de Investigación en Biología de la Producción, Universidad de Caldas, Manizales, Colombia
‡DeLaval Manufacturing, Kansas City, MO 64153
§GD Animal Health, Deventer, 7418 EZ, the Netherlands

ABSTRACT

The objective of the study was to evaluate premilking teat disinfectant efficacy with the use of a novel diagnostic protocol against a positive control. The evaluation of efficacy was based on establishing noninferiority of a new premilking teat disinfectant compared with an existing premilking teat disinfectant. Approximately 200 cows were randomly allocated to 1 of 2 groups. Study personnel were blinded as to cow allocation and identification of the premilking disinfectants. Quarters were disinfected using a dip cup before milking with either the control or experimental product. The disinfectant was left on the teat for approximately 30 s before being wiped off in preparation for milking. Quarter-level bacteriological infection status was established at the beginning of the study and quarter milk samples were collected biweekly thereafter. Bacteriological analyses were performed only when somatic cell counts in milk samples crossed a parity-specific threshold. Poisson regression models were used to analyze data. The difference in the rate of new intramammary infections was small and the upper bound of the 95% confidence interval of the rate difference was smaller than a predefined noninferiority limit. Based on the observed difference in new intramammary infections rate between the experimental and control product, it was concluded that the experimental disinfectant was not inferior compared with the positive control. The protocol used in this study proposes a valid and economically attractive methodology to evaluate the efficacy of teat disinfectants relative to a positive control.

Key words: teat disinfectant, iodine, intramammary infection, noninferiority

INTRODUCTION

A vast number of publications referring to the effectiveness of pre- and postmilking teat disinfection products are available, documenting the importance of teat disinfection as a management tool in preventing new IMI with major pathogens in dairy cows (Boddie et al. 2000; Peters et al., 2000; Oliver et al., 2001; NMC, 2013). In many of these initial studies, iodine teat dips have proven to be effective at reducing the incidence of new IMI with mastitis-causing pathogens (i.e., Pankey et al., 1987). As a consequence, the success of teat disinfection has resulted in the emergence of numerous germicidal products marketed as teat disinfectants, creating a need for a formal efficacy estimation protocol. To standardize the evaluation of efficacy of these germicides, the National Mastitis Council (NMC) has developed several protocols over the years for both experimental and natural challenge conditions (Schukken et al., 2013). The NMC publishes annually a list of teat disinfection products that have been tested using these standardized protocols (NMC, 2013).

Postmilking teat disinfection has proven to be an efficacious method to prevent IMI caused by *Streptococcus agalactiae* and *Staphylococcus aureus* compared with no postmilking teat sanitation (Neave et al., 1969; Foret et al., 2005, 2006). However, pathogens that reside primarily in the environment of the cow have not been controlled by postmilking teat sanitation (Pankey et al., 1987). One recommended approach to control environmental mastitis is to decrease the bacterial load of teat ends to pathogens from the environment (Galton et al., 1984). A decrease in premilking bacterial loads on the teat skin can be obtained when teats are dipped with a disinfectant followed approximately 30 s later by wiping and drying teats with individual paper or cloth towels.

The NMC has published efficacy estimation protocols for teat disinfection products (Schukken et al., 2013), including a protocol for determining the efficacy of
premilking teat disinfectants (Hogan et al., 1991). This protocol for evaluation of premilking teat disinfectants is based on a negative control trial using naturally occurring infections, identified through repeated monthly bacteriology in a 12-mo study. The proposed study design in this protocol lacks specific details, such as a precisely defined study protocol, appropriate power calculations, and statistical analysis guidelines. The long prescribed duration of the study combined with the extensive culture-based diagnostics and low incidence of new IMI under field conditions would lead to very expensive trials. As a consequence, these standardized protocols are currently not routinely performed by manufacturers of premilking teat disinfectants. The suggested use of negative controls is also a limiting factor for the performance of these trials, as many owners of commercial dairy farms are not interested in ending a practice that is generally considered valuable for their dairy. However, a need to formally and objectively evaluate premilking teat disinfectants exists. Therefore, a formal efficacy evaluation protocol for premilking teat disinfection of an experimental teat dip relative to a positive control product in reducing experimental challenge or naturally occurring IMI would be valuable.

Particularly in the case of premilking teat disinfectant efficacy trials using naturally occurring IMI, the amount of diagnostics that is required to identify new IMI is large. As all quarters are disinfected postmilking in such trials, the expected rate of new IMI is low, resulting in a predictably large number of bacteriological tests that are culture negative. For example, previous published studies reported 0.38 new IMI/100 quarter-weeks in quarters that were disinfected premilking versus 0.64 new IMI/100 quarter-weeks in control quarters (Oliver et al., 2001). This study had a duration of 78 wk (18 mo) in approximately 175 cows and resulted in a total of approximately 280 new IMI detected in an estimated 13,000 quarter milk bacteriology tests. The costs of these diagnostics are high and it may be expected that such trials would generally be cost-prohibitive. As SCC testing is less expensive than bacteriology, and SCC above a given threshold increases the likelihood of identifying bacterial organisms substantially (Schepers et al., 1997), a 2-step diagnostic process for identifying new IMI is proposed. For example, Schepers et al. (1997) estimated that quarter milk SCC with a threshold of either 100,000 or 200,000 SCC/mL had a sensitivity of 83 and 75%, respectively, and a specificity of 81 and 90%, respectively, for diagnosing IMI. Indicating that, with SCC as an initial screen, approximately 80% of true IMI would be detected by bacteriological culture and only approximately 15% of uninfected quarters would incorrectly be subjected to bacteriological culture. Or, at the cost of not finding 20% of truly infected quarters, 85% of uninfected quarters are not subjected to bacteriology. Using the overall incidence of approximately 0.5 per 100 quarter-weeks in 175 cows from Oliver et al. (2001), the prescreening with SCC would result in 114 quarters tested using bacteriology and 11 out of 14 IMI detected per sampling, compared with approximately 700 quarters tested when using only bacteriological culture and 14 IMI detected. However, the use of such a 2-step diagnostic protocol has not been tried under natural infection conditions in commercial dairy herds.

The objective of this study was to use a novel 2-step diagnostic protocol to evaluate the efficacy of premilking teat disinfectant to reduce naturally occurring new IMI. We compared the efficacy of an experimental teat dip with that of a positive control product with the objective of showing noninferiority of the experimental premilking iodine-based teat disinfectant compared with an existing iodine-based premilking teat disinfectant in commercial dairy herds.

**MATERIALS AND METHODS**

A 10-wk trial was conducted in a commercial dairy herd in western New York. Two groups of lactating cows with 100 and 99 animals each were selected. Cows were apparently in good health at the time of recruitment, based on a visual inspection of cows and udders, and had not been treated with antibiotics or anti-inflammatory products in the 30 d before recruitment. The cows were housed in freestall barns and milked twice daily, following recommended premilking procedures (Schukken et al., 2013). Groups within the herd did not differ in feeding and management practices.

Selected cows were allocated to 1 of 2 groups (control = 100; experimental = 99). Groups were balanced by parity, DIM, and status of infection at the beginning of the trial. Study personnel were blinded to treatment allocation and origin of the disinfectants. The same postmilking teat disinfectant (Theratec Plus, GEA Farm Technologies Inc., Naperville, IL) was used for both study groups. The cows in the groups received either Theratec Plus [control group (CG)] or F-2326 [DeLaval Inc. Kansas City, MO; experimental group (EG)] as a premilking teat disinfectant. Pre- and postmilking products were all a ready-to-use 0.5% iodine teat disinfectant with emollients (5% glycerin) and were applied using a nonreturn dip cup.

Before the start of the trial, a site visit was conducted to train study personnel on study procedures. At the same time, the milking system was evaluated and checked whether the system met the criteria defined in the ISO standard procedure 6690:2007, “Milking machine installations – Construction and performance” (ISO, 2007). General information on farm management...
practices were collected using a brief questionnaire. The adequacy of storage of teat disinfection material was checked and found to be in compliance with manufacturer recommendations. Color-coded dip cups and matching leg bands were provided to the farm to further ensure proper treatment application during the trial.

**Sample Collection and Laboratory Analysis**

The efficacy of F-2326 compared with Theratec Plus was evaluated by defining the rate of new IMI in each study group. At the start of the study and whenever a cow entered the study, all quarters from each cow were aseptically sampled for aerobic culture and a second sample was taken for SCC. The entry sample was used to define the infection status or lack of infection (culture negative) in all quarters of the cows entering the study. Throughout the study period, 2 milk samples from each quarter were collected biweekly by trained technicians. One sample was always used for SCC analysis and a second sample was taken aseptically and stored under refrigeration for a maximum of 24 h until it was decided whether the sample was going to be used for aerobic culture. The results of the SCC analyses were used to decide whether the quarter sample would be submitted for bacteriological culture. The threshold for culturing was an SCC value > 100,000 cells/mL for first lactation cows and > 200,000 cells/mL for second and greater lactation cows. These thresholds for culturing were based on previously suggested optimal thresholds to distinguish infected from noninfected quarters (Laevens et al., 1997; Schepers et al., 1997). Over the course of the study, any quarter with clinical mastitis was sampled aseptically by farm personnel for bacterial culture before treatment was applied. Somatic cell count analyses were conducted at Dairy One (Ithaca, NY) using automated cell counting equipment (Fossomatic, Foss, Hillerød, Denmark). All aerobic cultures were conducted at Quality Milk Production Services of Cornell University following standard protocols established by the NMC (Oliver et al., 2004). Briefly, approximately 0.01 mL of milk was streaked on trypticase soy agar containing 5% sheep blood and 0.1% esculin (PML Microbiologicals, Mississauga, Canada) and plates were incubated aerobically at 37°C for 48 h. After observation of colony morphology and hemolytic patterns on blood agar, isolates were examined further by means of 3% KOH, Gram staining of organisms, catalase and oxidase testing, and additional biochemical and metabolic evaluations as needed. Gram-negative organisms were identified by colony morphology on MacConkey’s agar (NMC, 1999).

**Definition of New IMI**

The initial bacteriological status of each quarter was established at the beginning of the trial using a single sample (Dohoo et al., 2011; Reyher and Dohoo, 2011). Once a quarter was identified as being infected with a particular organism, any repeat infection of the same quarter with the initially identified pathogen was not considered a new IMI. Therefore, an IMI was considered a new IMI based on the presence of a microorganism in a single quarter sample that was not identified in that quarter previously. Samples from clinical cases were considered eligible for identifying IMI. Once a new IMI was identified, the quarter became eligible for infection again for all organisms except the organisms that the quarter had been infected with before. Hence, a bacterial organism not previously found in the same quarter would be eligible to cause a second or third new IMI. Samples containing more than 2 bacterial species were considered as contaminated and were not informative of IMI status.

**Statistical Analysis**

The incidence of new IMI during the whole study period was calculated as the number of new IMI divided by the amount of quarter-time at risk for a new IMI across all time points. Efficacy of a teat disinfection product was evaluated in terms of reduction in the rate of new IMI observed among quarters disinfected with the experimental product compared with the rate observed in control quarters (Nickerson et al., 2004). The efficacy is calculated as follows:

\[
\text{efficacy of experimental disinfectant} = 1 - \frac{\text{rate of new IMI with experimental disinfectant}}{\text{rate of new IMI with control product}}. \tag{1}
\]

Note that if the rate of new IMI in the experimental disinfectant group is larger than the rate of new IMI in the positive control, the efficacy becomes negative and is defined at 0. The observed difference in the rate of new IMI between experimental and control disinfectants with a 95% CI was calculated. The difference in rate and 95% CI were used for evaluation of noninferiority. The study hypothesis was that no differences exist in the incidence of IMI between a control and experimental premilking iodine-based teat dips in reducing naturally occurring infections.

The trial was designed as a positive control field trial, where the objective was to show noninferiority of the new product versus the existing product used.
as control. Therefore, the main interest was to show a difference equal to zero (or a relative efficacy equal to zero). Even if the experimental product was slightly worse than the control product, the products may still be considered to be equivalent (Schukken et al., 2013). The experimental disinfectant was considered inferior only when the difference was larger than a determined critical value (Piaggio et al., 2006; Schukken et al., 2013).

Efficacy of a premilking teat disinfectant for reducing new IMI relative to no premilking disinfection is estimated at approximately 50% (Pankey et al., 1987; Oliver et al., 2001). For an experimental product to be considered noninferior we need to define a noninferiority margin ($d$). Using the data from the example above, a positive control would be expected to have an incidence of new IMI of approximately 1 new IMI per 100 quarter-weeks. The upper 95% CI of the incidence of new IMI in the experimental group would need to be smaller than $1 + p_3$, where $p_3$ is the risk of new IMI in group $i$. Using the data from the example, we define $p_3 = 0.004$ for a control teat disinfectant. Note that we use risk and rate here interchangeably. The incidence of new IMI is measured per quarter-week at risk. As a short time period is the unit of analysis, the distinction between risk and rate diminishes. The relationship between risk (cumulative incidence) and rate is given by

$$\text{risk} = 1 - \exp(-\text{rate} \times \Delta t),$$

where $\Delta t$ is measured in quarter-weeks.

For small risk, a good approximation for risk is rate $\times \Delta t$, and, hence, when risk is expressed as new IMI per week, rate and risk can be used interchangeably (Rothman et al., 2008). Therefore, we define $1 - (p_i/p_3) > 0.2$, and $1 - (p_i/p_3) = 0.5$. Then $p_i/p_3 = (p_i/p_3) \times (p_3/p_3)$, then, $1 - (0.5 \times p_i/p_3) > 0.2$ or $p_i/p_3 < 1.6$. This means that the risk of infection in the test product group divided by the risk of infection in the positive control group should be less than 1.6. To relate this relative effect (efficacy is based on a ratio of 2 infection risks) to the critical value of the difference between infection risk in treatment and positive control ($d$), $d = p_1 - p_3$ applies, for $p_1 \leq 1.6 \times p_3$ this becomes $d = 1.6 \times p_3 - p_3 = 0.6 \times p_3$. With this definition of $d$, the study size can be calculated. For $n = n_1 = n_3$, the sample size per group ($N$) is given by (Christensen, 2007)

$$n = (Z_{2a} + Z_{b})^2 \times [p_1(1 - p_1) + p_3(1 - p_3)]/d^2,$$  

where $Z$ is the standard normal distribution evaluated at either $2a$ or $b$, where $2a$ is twice the type I error and $b$ is the type II error. Using an expected IMI risk in the positive control of approximately 1 new IMI per 100 quarter-week (e.g., Oliver et al., 2001), a one-sided significance level of 5%, and an estimated power of 80%, the trial size for a noninferiority study would have approximately 3,500 quarter-weeks per treatment group. Correction for the loss of diagnostic efficacy because of the prescreening with SCC (sensitivity estimated at 80%), approximately 4,400 quarter-weeks per treatment group. For premilking teat disinfection trials, the positive control product should have a published efficacy against a negative control of approximately 50% or higher for the mastitis organisms of interest. Efficacy of premilking teat disinfectants meeting this ≥50% efficacy requirement have been reported (NMC, 2013). In general, we propose here that a premilking teat disinfection product is considered to be adequate if the left hand side of the 95% CI of the efficacy is at least 20%.

For premilking teat disinfection trials, the positive control product should have a published efficacy against a negative control of approximately 50% or higher for the mastitis organisms of interest. Efficacy of premilking teat disinfectants meeting this ≥50% efficacy requirement have been reported (NMC, 2013). In general, we propose here that a premilking teat disinfection product is considered to be adequate if the left hand side of the 95% CI of the efficacy is at least 20%. The derivation of $d$, the maximum acceptable difference, can then be shown mathematically by defining $p_i = \text{risk of IMI in group } i$, $p_3 = \text{risk in test disinfectant}$, $p_2 = \text{risk in negative control}$, and $p_3 = \text{risk in positive control disinfectant}$. Note that we use risk and rate here interchangeably. The incidence of new IMI is measured per quarter-week at risk. As a short time period is the unit of analysis, the distinction between risk and rate diminishes. The relationship between risk (cumulative incidence) and rate is given by

$$\text{risk} = 1 - \exp(-\text{rate} \times \Delta t),$$

where $\Delta t$ is measured in quarter-weeks.

For small risk, a good approximation for risk is rate $\times \Delta t$, and, hence, when risk is expressed as new IMI per week, rate and risk can be used interchangeably (Rothman et al., 2008). Therefore, we define $1 - (p_i/p_3) > 0.2$, and $1 - (p_i/p_3) = 0.5$. Then $p_i/p_3 = (p_i/p_3) \times (p_3/p_3)$, then, $1 - (0.5 \times p_i/p_3) > 0.2$ or $p_i/p_3 < 1.6$. This means that the risk of infection in the test product group divided by the risk of infection in the positive control group should be less than 1.6. To relate this relative effect (efficacy is based on a ratio of 2 infection risks) to the critical value of the difference between infection risk in treatment and positive control ($d$), $d = p_1 - p_3$ applies, for $p_1 \leq 1.6 \times p_3$ this becomes $d = 1.6 \times p_3 - p_3 = 0.6 \times p_3$. With this definition of $d$, the study size can be calculated. For $n = n_1 = n_3$, the sample size per group ($N$) is given by (Christensen, 2007)

$$n = (Z_{2a} + Z_{b})^2 \times [p_1(1 - p_1) + p_3(1 - p_3)]/d^2,$$  

where $Z$ is the standard normal distribution evaluated at either $2a$ or $b$, where $2a$ is twice the type I error and $b$ is the type II error. Using an expected IMI risk in the positive control of approximately 1 new IMI per 100 quarter-week (e.g., Oliver et al., 2001), a one-sided significance level of 5%, and an estimated power of 80%, the trial size for a noninferiority study would have approximately 3,500 quarter-weeks per treatment group. Correction for the loss of diagnostic efficacy because of the prescreening with SCC (sensitivity estimated at 80%), approximately 4,400 quarter-weeks per treatment group. For premilking teat disinfection trials, the positive control product should have a published efficacy against a negative control of approximately 50% or higher for the mastitis organisms of interest. Efficacy of premilking teat disinfectants meeting this ≥50% efficacy requirement have been reported (NMC, 2013). In general, we propose here that a premilking teat disinfection product is considered to be adequate if the left hand side of the 95% CI of the efficacy is at least 20%. The derivation of $d$, the maximum acceptable difference, can then be shown mathematically by defining $p_i = \text{risk of IMI in group } i$, $p_3 = \text{risk in test disinfectant}$, $p_2 = \text{risk in negative control}$, and $p_3 = \text{risk in positive control disinfectant}$. Note that we use risk and rate here interchangeably. The incidence of new IMI is measured per quarter-week at risk. As a short time period is the unit of analysis, the distinction between risk and rate diminishes. The relationship between risk (cumulative incidence) and rate is given by

$$\text{risk} = 1 - \exp(-\text{rate} \times \Delta t),$$

where $\Delta t$ is measured in quarter-weeks.
the middle between 2 samplings. Prevalence of any IMI during the study period was calculated as the number of IMI at a given time point divided by the number of quarters sampled at that specific time point.

The quarter was the unit of interest accounting for within-cow and within-quarter clustering by fitting a generalized linear mixed model, including fixed effects of time and the interaction with treatment as categorical variables and quarter within cow as a repeated effect. Repeated measurements on the same individual are expected to lead to dependence between the responses of observations in a group (or cluster), such as control or experimental groups (Dohoo et al., 2009). This means that within-cow correlations of infection events should also be accounted for in the statistical analyses (Ceballos-Marquez et al., 2011). Analyses were carried out using PROC GLIMMIX in SAS version 9.2 (SAS Institute Inc., Cary, NC). This procedure fits statistical models to data with correlations or nonconstant variability and where the response is not necessarily

Figure 1. Sample size and related power of the study in noninferiority studies per treatment group for an estimated incidence of new IMI of 1 per 100 quarter-weeks at risk. The noninferiority limit is defined at 0.006, 0.005, and 0.004, related to a minimum efficacy of a test premilking disinfectant of 20, 25, and 30%, respectively. Color version available in the online PDF.
normally distributed (Littell et al., 2006). The generalized linear mixed model used was

\[ f(\cdot) = \text{intercept} + \text{treatment} + \text{time} + (\text{treatment} \times \text{time}) + \text{cow(random)} + \text{Re}, \tag{3} \]

where \( f(\cdot) \) is link function of the probability or rate of a new IMI. The probability of a new IMI is modeled in a logistic regression, whereas the rate of a new IMI is modeled in a Poisson regression. The link function is either a logit link when a binomial distribution is used or a log link for a Poisson distribution, treatment is teat disinfection product (control or experimental), and time corresponds to the time of milk sample collection relative to the cow entry into the trial. Time would be 0, 2, 4, 6, 8, 10, or the actual wk of occurrence of a clinical case or a special sample, such as calving, dry-off, or culling. For cows entering the trial after the start or cows leaving the trial before the completion of the study, less time would be included in the study and the maximum time of 10 wk would not be reached. Cow is a random effect to correct for correlation within cow. Finally, Re is a complex error structure allowing for longitudinal within-quarter correlation.

In the case of Poisson regression, a weighting factor was used describing the weeks at risk for each time period. Alternatively, in the case of Poisson regressions, the number of new IMI per time period may be used as the outcome variable and an offset term with the natural log of the number of weeks at risk should be included in the explanatory part of the model.

The coefficients of the regression models were expressed as odds ratios for the logistic regression and incidence rate ratios for the Poisson regression. Both of these parameters give the relative effect of predipping with an experimental product on the incidence of new IMI across all quarters compared with a control product (Dohoo et al., 2009). The Wald’s test was used to determine statistical significance of individual terms in the regression model.

The dependent variable, SCC, obtained from each quarter sample (0, 2, 4, 6, 8, 10 wk after treatment), was analyzed using the linear mixed model

\[ \log_{10}(\text{SCC}) = \text{intercept} + \text{treatment} + \text{time} + \text{treatment} \times \text{time} + \text{cow(random)} + \text{Re}, \tag{4} \]

where \( \log_{10}(\text{SCC}) \) is the base 10 log of the measured quarter SCC in thousands per milliliter, and the fixed and random effects in the model are the same as in model [3]. Final analyses were carried out using PROC MIXED in SAS version 9.2 (SAS Institute Inc.). For all analyses, a \( P \)-value of 0.05 was used for deciding on statistical significance.

**RESULTS**

The study in the herd took place between February and May, and the trial performance was without issues.
Environmental temperature ranged between −12.3 and 25.4°C. Milking equipment was checked twice, at the beginning and midway through the study. Both times the equipment fully met ISO standards. Teat disinfectants were applied according to protocol without any evidence for protocol violations. Teat disinfectant usage was approximately 7 mL/cow per day.

The effect of the 2 premilking teat disinfectants (F-2326 in EG and Theratec Plus in CG) was evaluated using information available from 4,017 quarter samples (Table 2). A total of 630 in CG and 587 in EG were cultured. Samples were cultured at study entry, calving, dry-off, exit, and during the trial when SCC was above the threshold value. A total of 1,374 (68.6%) and 1,426 (70.8%) milk samples for CG and EG, respectively, did not meet these specifications; therefore, they were not submitted for bacteriological culture. Of all samples submitted for culture, 306 (CG) and 229 (EG) quarter milk samples met the qualifying requirements for bacteriological culture with the quarters having an SCC >100,000 cells/mL in first lactation cows or >200,000 cells/mL in lactation 2 and higher cows. A total of 6 cases of clinical mastitis were observed; 5 cases were in CG and 1 case in EG. All clinical cases were cultured; 4 samples were culture negative, 2 samples cultured *Escherichia coli*, and contributed 2 new IMI to the study.

The incidence rate of having a new IMI was calculated based only on the information available for the quarters sampled during the study period. The effect of 2 premilking teat disinfectants on the risk of new IMI was evaluated based on a total of 6,273 quarter-weeks at risk, 3,109 in CG and 3,164 in EG (Tables 3 and 4). The overall rate of new IMI was 0.55 (SE = 0.13) and 0.44 (SE = 0.12) new IMI per 100 quarter-weeks at risk in the CG and EG, respectively. Therefore, the relative rate of new IMI was 1.25 times higher in the CG compared with the EG, resulting in a relative efficacy of the experimental dip compared with the control product of 19%. The Poisson regression model converged and

---

**Table 2. Isolation of mastitis pathogens in quarter milk samples collected from cows treated with different pre- and postmilking teat disinfectants**

<table>
<thead>
<tr>
<th>Sample status</th>
<th>All samples</th>
<th>Theratec Plus²</th>
<th>F-2326³</th>
<th>Samples above SCC threshold</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No.</td>
<td>%</td>
<td>No.</td>
<td>%</td>
</tr>
<tr>
<td>Samples cultured</td>
<td>630</td>
<td>31.4</td>
<td>587</td>
<td>29.2</td>
</tr>
<tr>
<td>Culture negative</td>
<td>542</td>
<td>27.0</td>
<td>502</td>
<td>24.9</td>
</tr>
<tr>
<td>Culture positive</td>
<td>84</td>
<td>4.2</td>
<td>73</td>
<td>3.6</td>
</tr>
<tr>
<td><em>Streptococcus</em> spp.</td>
<td>2</td>
<td>0.1</td>
<td>1</td>
<td>0.0</td>
</tr>
<tr>
<td><em>Streptococcus dysgalactiae</em></td>
<td>4</td>
<td>0.2</td>
<td>7</td>
<td>0.3</td>
</tr>
<tr>
<td><em>Streptococcus uberis</em></td>
<td>18</td>
<td>0.9</td>
<td>2</td>
<td>0.1</td>
</tr>
<tr>
<td><em>Staphylococcus aureus</em></td>
<td>12</td>
<td>0.6</td>
<td>13</td>
<td>0.6</td>
</tr>
<tr>
<td><em>Staphylococcus</em> spp.</td>
<td>35</td>
<td>1.7</td>
<td>34</td>
<td>1.7</td>
</tr>
<tr>
<td><em>Corynebacterium</em> spp.</td>
<td>0</td>
<td>0.0</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>Gram-positive bacillus</td>
<td>0</td>
<td>0.0</td>
<td>1</td>
<td>0.0</td>
</tr>
<tr>
<td><em>Escherichia coli</em></td>
<td>6</td>
<td>0.3</td>
<td>6</td>
<td>0.3</td>
</tr>
<tr>
<td><em>Klebsiella</em> spp.</td>
<td>5</td>
<td>0.2</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td><em>Pasteurella</em> spp.</td>
<td>0</td>
<td>0.0</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td><em>Serratia</em> spp.</td>
<td>0</td>
<td>0.0</td>
<td>1</td>
<td>0.0</td>
</tr>
<tr>
<td><em>Citrobacter</em> spp.</td>
<td>2</td>
<td>0.1</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td><em>Enterobacter</em> spp.</td>
<td>0</td>
<td>0.0</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td><em>Truoperella pyogenes</em></td>
<td>0</td>
<td>0.0</td>
<td>3</td>
<td>0.1</td>
</tr>
<tr>
<td><em>Proteococcus</em> spp.</td>
<td>0</td>
<td>0.0</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>Yeast</td>
<td>0</td>
<td>0.0</td>
<td>5</td>
<td>0.2</td>
</tr>
<tr>
<td>Contaminated</td>
<td>4</td>
<td>0.2</td>
<td>12</td>
<td>0.6</td>
</tr>
<tr>
<td>Samples not cultured¹</td>
<td>1,374</td>
<td>68.6</td>
<td>1,426</td>
<td>70.8</td>
</tr>
</tbody>
</table>

¹Results are shown for all samples and for samples with an SCC value above the defined threshold.
²Theratec Plus, GEA Farm Technologies Inc., Naperville, IL.
³DeLaval Inc., Kansas City, MO.
⁴These samples did not meet the SCC selection criterion for culture or no milk was available for SCC and culture.
showed a good fit to the data. Within-quarter correlation was low, with a correlation coefficient of 0.007. In the Poisson regression model, the relative rate of new IMI during lactation, although numerically higher in CG, was not significantly different between the 2 treatment groups ($P = 0.57$). Time had a significant effect on the overall rate of new IMI ($P < 0.05$), with a decreasing trend of new IMI over the course of the 10-wk study (Table 4). The interaction of treatment by time was not significant ($P > 0.50$). The least squares means of the rate of new IMI for the 2 treatment groups are shown in Figure 2.

As the trial did not show a significant difference between CG and EG, the goal then was to identify whether the conclusion of no difference may result in a noninferiority claim. The claim of noninferiority was based on the right-hand side of the 95% CI of the rate difference being smaller than the predefined noninferiority limit of $d$. In this study, $d$ was defined at $0.6 \times$ rate of new IMI in the positive control group. Therefore, the critical value was $0.6 \times 0.55 = 0.33$ new IMI/100 quarter-weeks. The rate difference and 95% CI are shown in Table 4. The rate difference was $-0.10$ with the upper bound of the 95% CI at 0.24. The upper bound of the CI was therefore below the critical value, $d$. Thus, it can be concluded that noninferiority of the experimental premilking disinfectant relative to the control was shown (Figure 3).

Based on the observed rate of new IMI in the control group, the defined value of $d$, the critical noninferiority difference, and the actual sample size, the actual power of the study was calculated. The actual power when the true difference between the 2 disinfectants was assumed to be 0, and the rate of new IMI in the control group of 0.55 new IMI/100 quarter-weeks with approximately 3,100 observations per treatment group was estimated at 0.55. The within-cow correlation of new IMI was low and estimated at 0.02.

**Effect on SCC**

A total of 4,017 quarter milk samples were collected between 0 and 10 wk after the start of the trial. Of these, 3,960 were analyzed for SCC. Fifty-seven val.

**Table 3.** Number of new IMI and quarters at risk for an IMI on a biweekly basis

<table>
<thead>
<tr>
<th>Treatment</th>
<th>0</th>
<th>2</th>
<th>4</th>
<th>6</th>
<th>8</th>
<th>10</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Control group</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>New IMI</td>
<td>7</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>0</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>Quarter-week at risk</td>
<td>0</td>
<td>729</td>
<td>705</td>
<td>627</td>
<td>546</td>
<td>502</td>
<td>3,109</td>
</tr>
<tr>
<td><strong>Experimental group</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>New IMI</td>
<td>7</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>Quarter-week at risk</td>
<td>0</td>
<td>713</td>
<td>700</td>
<td>631</td>
<td>602</td>
<td>518</td>
<td>3,164</td>
</tr>
</tbody>
</table>

1The incidence rate of all quarters is calculated over the total new IMI and total quarters at risk.

<table>
<thead>
<tr>
<th>Sample status (no.)</th>
<th>Theratec Plus(^1) (control)</th>
<th>F-2326(^2) (experimental)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quarter-wk at risk</td>
<td>3,109</td>
<td>3,164</td>
</tr>
<tr>
<td>New IMI</td>
<td>17</td>
<td>14</td>
</tr>
<tr>
<td><em>Streptococcus</em> spp.</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td><em>Streptococcus dysgalactiae</em></td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td><em>Streptococcus uberis</em></td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td><em>Staphylococcus aureus</em></td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td><em>Staphylococcus</em> spp.</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td><em>Corynebacterium</em> spp.</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Gram-positive bacillus</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><em>Escherichia coli</em></td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td><em>Klebsiella</em> spp.</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><em>Pasteurella</em> spp.</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><em>Serratia</em> spp.</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><em>Citrobacter</em> spp.</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td><em>Trueperella pyogenes</em></td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td><em>Protheca</em> spp.</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Yeast</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Rate of new IMI per 100 quarter-week (95% CI)</td>
<td>0.55 (0.31, 0.88)</td>
<td>0.44 (0.24, 0.74)</td>
</tr>
<tr>
<td>Rate difference (95% CI)</td>
<td>-0.10 (-0.45, 0.24)</td>
<td></td>
</tr>
<tr>
<td>Experimental disinfectant efficacy</td>
<td>19%</td>
<td></td>
</tr>
</tbody>
</table>

\(^1\)Theratec Plus, GEA Farm Technologies Inc., Naperville, IL.
\(^2\)DeLaval Inc., Kansas City, MO.
ues were missing because too little milk was present, samples were clinical, or samples were lost for other reasons. Mean SCC in all samples was 274,000 cells/mL, with a geometric mean SCC of 28,000 cells/mL. The SCC values ranged between 1,000 and 20,105,000 cells/mL. No differences in overall mean SCC were observed between CG and EG (P = 0.21). However, a slight but statistically significant increase in SCC was observed during the trial (P < 0.01). Also, the interaction between treatment and time was significant (P = 0.01), and further evaluation indicated that SCC in milk from quarters in CG at the beginning and at the end of the trial was significantly higher compared with the EG (Figure 4). In the linear model, the longitudinal within-quarter correlation of SCC over the study was approximately 0.25, and the within-sample day cross-sectional correlation within cow was also estimated at approximately 0.25, clearly showing a high correlation of SCC measurements over time in the same quarter and within the same mammary gland at any given point in time.

DISCUSSION

The major objective of the study described herein was to evaluate the feasibility of performing noninferiority studies on premilking teat disinfectants on commercial dairy farms, using a novel 2-step diagnostic process for the detection of new IMI. The study showed sufficient new IMI and very small differences between the experimental product and positive control product, such that a conclusion of noninferiority could be drawn from the results.

These results were obtained using a novel 2-step diagnostic process that allowed the study to be completed in a reasonable time frame, with a targeted use of the more precise and expensive diagnostics, whereas a low cost screening test was used extensively. Eventually, approximately 4,000 samples were tested for SCC and 30.7% of these samples were also tested using classical bacteriology. Compared with previously published trials (Oliver et al., 1993a; Peters et al., 2000; Oliver et al., 2001), our proposed noninferiority study with a 2-step diagnostic process resulted in substantial savings while maintaining the ability to draw valid conclusions.

The use of a 2-step diagnostic process, as used in this protocol, is known as testing in series (i.e., high SCC and culture positive) and is therefore expected to result in an increased specificity and decreased sensitivity (Smith and Slenning, 2000; Dohoo et al., 2009). Given the low incidence of new IMI, a strategy to optimize specificity would be generally recommended (Smith and Slenning, 2000). The estimated loss in overall sensitivity is approximately 20% (Schepers et al., 1997), resulting in fewer identified new IMI. The presence of diagnostic error that is identical across the 2 treatments would be expected to introduce little, if any, bias in the estimation of the relative efficacy of the disinfectants. However, diagnostic errors do introduce a bias in comparative studies where the bias generally results in
an efficacy estimate that is closer to 1 (Morant et al., 1988). Although the introduced bias due to diagnostic error was relatively small, it is of particular concern in noninferiority studies where the identification of a relative efficacy of 1 is the preferred outcome of the study. The estimated bias in efficacy estimates due to diagnostic error is small when the true relative efficacy is close to 1 (Morant et al., 1988); as the observed difference between the 2 treatments in the trial was small, we expect that the introduced bias due to diagnostic error is small. Additionally, in the current study, the sample size was corrected for the estimated loss in sensitivity and the overall ability to identify a difference, if it was truly present, turned out to be acceptable. In hindsight, the duration of the trial could have been extended to improve the observed power. Still, the difference in new IMI favoring the experimental product eventually allowed a conclusion of noninferiority.

Keeping track of new IMI and time at risk is essential to estimate the rate of new IMI in each treatment group. Even though we used biweekly sampling, the unit of time in our analysis was weeks. By choosing biweekly sampling rather than weekly sampling, the rate of new IMI will be lower, as many short-term infections will be missed. This principle is shown in Figure 5. Quarters 1, 3, 4, and 5 have an observed new IMI, whereas quarters 2 (2 new IMI), 3, 5, and 6 have a new IMI that occurs between samplings and is short in duration so that it is missed by the biweekly sampling scheme. Note that both in the first and second week between samplings, new IMI occur that can be observed; in quarter 4, a new IMI occurs in wk 3, whereas, in quarter 5, a new IMI occurs in wk 5. Hence, both weeks are potentially at risk for new IMI. In reality, the actual time of new IMI is not observed, and therefore all new IMI are assumed to occur at the midpoint between samplings. After the IMI occurs, the quarter is not at risk for a new IMI for 1 wk, as indicated by a hashed bar in Figure 5. After 1 wk a new IMI can occur with a different bacterial species than the one observed in previous IMI. Quarters at risk are also counted in Figure 5. In the case of biweekly sampling, the observed rate of new IMI will be 4 per 43 quarter-weeks at risk. In the case of weekly sampling, the observed rate would be 8 per 43 quarter-weeks at risk. The denominator is the same as with the weekly sampling, with only 0.5 wk at risk lost after a new IMI. Hence, with more frequent sampling, the required sample size may be smaller, as a higher rate of new IMI is likely observed (Wickramaratne 1995).

A high efficacy of iodine-based premilking teat disinfectants against most of the environmental and contagious pathogens has been reported after experimental challenges or natural exposure studies (NMC, 2013). Previous studies have reported an adequate efficacy of iodine disinfectants (Pankey et al. 1987; Oliver et al., 1993); our results indicated that the iodine-based experimental product compared favorably with the iodine-based control disinfectant in reducing the rate of new IMI in a commercial dairy herd using a noninferiority study design. Although not tested here, it may be expected that both products would reduce the rate of new IMI relative to a negative control.

Although the study focuses on all new IMI across bacterial species, the distribution of bacterial species in the 2 groups is of interest. As shown in Table 2, the difference between CG and EG in bacterial species profile is remarkably similar. Possible exceptions include a higher proportion of Streptococcus uberis and Pasteurella spp. in CG and Truepurella pyogenes and yeast in EG. The overall prevalence in IMI was virtually identical in both treatment groups, which is probably the most important parameter to evaluate given the potential transmission of pathogens between cows within a treatment group.

The current NMC protocol for evaluation of premilking teat disinfectants (Hogan et al., 1991) is in need of a rewrite. The defined protocol does not address the issues recently identified by Schukken et al. (2013) with regard to postmilking teat disinfectant efficacy estimation. Among other issues, the protocol does not address positive control trials, noninferiority, quality assurance in data collection, power calculations, or standardized reporting. The studies, as reported herein, may serve as a template for the development of the next generation of such a standardized protocol for evaluation of premilking teat disinfectants.

In the description of the statistical analysis of the data, both Poisson regression and logistic regression were indicated as options. Poisson regression appears to be the more natural choice for data on the rate of new IMI. However, our experience is that convergence
of Poisson models is not always reached and, in such situations, logistic regression is more robust (Schukken et al., 2010). Logistic regression is used with the outcome variable defined as events or trials, which in teat disinfectant studies would be new IMI/weeks at risk; for most of our data that would be 0/2 and, in some cases, 1/2 or even 2/2 IMI/weeks at risk.

When defining a standard protocol for evaluation of premilking teat disinfectants a few critical issues would need to be defined. In this study, we defined the required efficacy of premilking disinfectants to be used as positive controls relative to no disinfectant used as, at least, 50%. Several disinfectants meeting this requirement have been reported (NMC, 2013). To define the noninferiority limit ($d$), the minimal acceptable efficacy of a premilking teat disinfectant was defined at 20%. Although this minimum efficacy is lower than the minimal efficacy used for postmilking teat disinfection (Schukken et al., 2013), the difference between the efficacy of the positive control to the minimal acceptable efficacy is identical to what was defined for postmilking teat disinfection at 30 percentage points (70–40 vs. 50–20%, respectively). It is important to keep in mind that the minimal acceptable efficacy refers to the left-hand side of the 95% CI. Therefore, for any reasonable sample size, the actual point estimate of efficacy of a test product in a noninferiority trial will need to be relatively close to the positive control product. It will be important for the major partners in the area of teat disinfectants to consider the arguments brought forward in this report and define the acceptable noninferiority limit ($d$) to be used in future trials.

**CONCLUSIONS**

A field trial with naturally occurring IMI evaluating an experimental premilking teat disinfectant against a positive control was conducted, and noninferiority of the experimental product relative to the positive control was shown. A novel 2-step diagnostic process was used to identify new IMI. This diagnostic protocol resulted in a more economical and efficient way to identify new IMI. The described protocol may serve as a template for future trials and for the definition of a standardized efficacy evaluation protocol for premilking teat disinfectants.

**ACKNOWLEDGMENTS**

The authors acknowledge the participation of the staff of two dairy farms in upstate New York in the two described trials. The staff and faculty of Quality Milk Production Services at Cornell University are acknowledged for participation in data collection, bacteriology, and reporting. The staff of DairyOne (Ithaca, NY) is acknowledged for SCC testing. The authors thank the anonymous reviewers for the constructive criticism that contributed substantially to the clarity of the

---

**Figure 5.** Description of IMI and new infections in 8 theoretical quarters during 6 wk of observation with 4 biweekly sampling events. New IMI observed are indicated by a shaded triangle, whereas new IMI not observed are indicated by an open triangle. Bacterial species are indicated on the infection lines (Escherichia coli, Staphylococcus aureus, Streptococcus uberis, and Streptococcus dysgalactiae). The time that a quarter is assumed to be not at-risk for another new IMI is indicated with a hashed bar.
arguments in the manuscript. DeLaval Manufacturing (Kansas City, MO) provided funding for the field study that was used as an example.

REFERENCES


Page 295 of 295