Defining standardized protocols for determining the efficacy of a postmilking teat disinfectant following experimental exposure of teats to mastitis pathogens

Y. H. Schukken,*1 B. J. Rauch,* and J. Morelli†

* Department of Population Medicine and Diagnostic Sciences, College of Veterinary Medicine, Cornell University, Ithaca, NY 14853
†Ecolab Research and Development, 8561 154th Avenue Northeast, Redmond, WA 98052

ABSTRACT

The objective of this paper was to define standardized protocols for determining the efficacy of a postmilking teat disinfectant following experimental exposure of teats to both *Staphylococcus aureus* and *Streptococcus agalactiae*. The standardized protocols describe the selection of cows and herds and define the critical points in performing experimental exposure, performing bacterial culture, evaluating the culture results, and finally performing statistical analyses and reporting of the results. The protocols define both negative control and positive control trials. For negative control trials, the protocol states that an efficacy of reducing new intramammary infections (IMI) of at least 40% is required for a teat disinfectant to be considered effective. For positive control trials, noninferiority to a control disinfectant with a published efficacy of reducing new IMI (efficacy data) are inconsistent among countries. An objective of the National Mastitis Council (NMC; Madison, WI) is to monitor the development of protocols that determine effectiveness of mastitis control products, such as teat disinfectant solutions. The NMC has long been the main resource for such standardized protocols for evaluation of teat disinfectants. This article describes a protocol that was developed from previously defined protocols (Hogan et al., 1990; Nickerson et al., 2004) and updated to allow for trials that can be done using Good Clinical Practice guidelines and recently published reporting guidelines for trials in food animals (O'Connor et al., 2010). More importantly, positive control trials are the standard for treatment evaluation in many countries, and negative control trials are considered unethical (Piaggio et al., 2006; Powers 2008; EMEA, 2009). Therefore, positive control trials should be accommodated in standardized trial protocols, including the protocols to evaluate postmilking teat disinfectants.

Determining the efficacy of postmilking teat disinfectants following experimental exposure of teats to mastitis pathogens is a common procedure for quality-oriented teat disinfectant manufacturers. When such experimental exposure trials are performed, it is important to standardize procedures, thereby assuring a more uniform and accurate comparison of trials that evaluate these products.

The objective of this paper was therefore to describe a standardized protocol for selecting cows for the trial, growing the challenge bacteria before experimental exposure, performing the experimental exposure, evaluating the culture results, performing statistical analyses, and reporting the results. The guidelines defined in this protocol should be carefully followed to allow proper comparisons between trials.

Key words: teat disinfectant, mastitis, protocol, National Mastitis Council

INTRODUCTION

The role of postmilking teat disinfection as a management tool in preventing new IMI dairy cows is well documented (Dodd et al., 1969; for review: Dufour et al., 2011). Success of this practice in reducing new IMI has resulted in the emergence of numerous germicidal products marketed as teat disinfectants. However, governmental agency requirements for data on the ability of a teat disinfectant to reduce incidence of new IMI (efficacy data) are inconsistent among countries. An objective of the National Mastitis Council (NMC; Madison, WI) is to monitor the development of protocols that determine effectiveness of mastitis control products, such as teat disinfectant solutions. The NMC has long been the main resource for such standardized protocols for evaluation of teat disinfectants. This article describes a protocol that was developed from previously defined protocols (Hogan et al., 1990; Nickerson et al., 2004) and updated to allow for trials that can be done using Good Clinical Practice guidelines and recently published reporting guidelines for trials in food animals (O'Connor et al., 2010). More importantly, positive control trials are the standard for treatment evaluation in many countries, and negative control trials are considered unethical (Piaggio et al., 2006; Powers 2008; EMEA, 2009). Therefore, positive control trials should be accommodated in standardized trial protocols, including the protocols to evaluate postmilking teat disinfectants.

Determining the efficacy of postmilking teat disinfectants following experimental exposure of teats to mastitis pathogens is a common procedure for quality-oriented teat disinfectant manufacturers. When such experimental exposure trials are performed, it is important to standardize procedures, thereby assuring a more uniform and accurate comparison of trials that evaluate these products.

The objective of this paper was therefore to describe a standardized protocol for selecting cows for the trial, growing the challenge bacteria before experimental exposure, performing the experimental exposure, evaluating the culture results, performing statistical analyses, and reporting the results. The guidelines defined in this protocol should be carefully followed to allow proper comparisons between trials.
RECOMMENDED METHODS

Experimental Design

All 4 teats of each cow are exposed to suspensions of *Staphylococcus aureus*, *Streptococcus agalactiae*, or other mastitis pathogens on a repeated experimental basis. Both *Staph. aureus* and *Strep. agalactiae* have been used for many years as test organisms (see, for example, Pankey et al., 1985; Leslie et al., 2006) and will therefore be used throughout. The NMC publishes annually the “summary of peer-reviewed publications on efficacy of premilking and postmilking teat disinfectants” in its proceedings, and the vast majority of publications use both *Staph. aureus* and *Strep. agalactiae* as challenge organisms. The long-term use of these 2 organisms therefore allows excellent comparison between trials and between products. The experimental teat disinfectant is applied to 2 teats of each udder, and the remaining teats serve as either negative or positive controls, depending on whether the trial is being conducted to prove efficacy against no teat disinfectant or noninferiority compared with an existing teat disinfectant product. Efficacy is defined as 1 minus the ratio of the infection rate in treated quarters divided by control quarters. It is defined here that an acceptable efficacy of a teat disinfectant against a negative control should be at least 40%. This 40% efficacy is based on the previous protocol (Hogan et al., 1990) and results reported in numerous published peer-reviewed trials as reported in the summary of peer-reviewed publication on efficacy of premilking and postmilking teat disinfectants compiled annually by the NMC research committee and published in the NMC annual proceedings (hereafter, the NMC summary; NMC, 2013). Regardless of treatment design, the numbers of new IMI in quarters with disinfected and control teats will be determined to make treatment comparisons. When comparing with a positive control of known efficacy, the positive control product should have been shown to be efficacious in an experimental exposure or natural exposure trial, and should be listed in the NMC summary. A product may be used as a positive control if the published efficacy against a negative control is at least 70%.

Selection of Experimental Herds, Cows, and Quarters

Trials should be conducted in research herds or herds supervised by competent research faculty or staff. Milking equipment and management practices should be monitored regularly to minimize machine-mediated infections. “Regularly” is defined here as at least once before the start of the trial and at least once during the trial. All quarters are eligible for inclusion in the trial except those infected with microorganisms of the same species as challenge microorganisms and quarters in which teats are deformed, abnormal, or injured before or during the trial. Teat scoring should be done using a documented teat scoring system (Neijenhuis et al., 2000; Mein et al., 2001; Öhnhst et al., 2003). Teat scoring should take place before starting the experimental exposure and at least every 4 wk during the trial. Diagonal teats or teats on either the right or left side of each udder are immersed in the experimental product, and the remaining teats serve as controls. A trial in which either the front or rear quarters are immersed in product and the remaining rear or front quarters are kept as controls is not valid, because the risk of infection has been shown repeatedly to be lower in front quarters than in rear quarters (Barkema et al., 1997). The innate difference in infection risk between front and rear quarters would bias the comparison of products or product versus a negative control. If an application method other than teat dipping (such as spraying or foaming) is used for application of the disinfectant, the application method should be carefully described in the trial report.

Cultures

Trials should be conducted with unexpired reference strains of bacteria. The use of *Staphylococcus aureus* ATCC 29740 [American Type Culture Collection (ATCC), Rockville, MD] and *Streptococcus agalactiae* (ATCC 27956) is suggested. Other strains or species that are used as challenge organisms need to be documented and should be available in publicly available collections, preferably the ATCC. The ATCC collection is available in all continents through partnerships with local standards organizations. All isolates are to be stored at −70 ± 10°C in appropriate media before use. A sufficient number of vials of stock culture should be prepared and stored, with minimal serial transfer, to permit preparation of an actively growing culture from a separate vial each time bacteria are prepared for experimental exposure throughout the trial. Although the test organisms recommended above are not known to be pathogenic to humans, it is prudent to use care in handling all bacterial cultures. Unused bacterial suspensions should be autoclaved before disposal, and personnel should be instructed to wear disposable gloves and sanitize hands with antiseptic soap after contact with any bacterial suspensions or milk from infected cows.

Preparation of *Staph. aureus*. One 6-mL tube of sterile tryptic soy broth (TSB) is inoculated from a vial of stored stock culture or from 2 to 3 colonies of
plated stock culture, and incubated at 37 ± 2°C for 5 to 7 h. The culture is streaked onto a tryptic soy agar plate containing 5% calf or sheep blood as a purity check. The suitability of blood for this purpose should be determined as described previously (Harmon et al., 1990). At the same time as blood agar plates are plated and cultured for a purity check, a 500-mL flask of TSB is inoculated with the entire contents of the 6-mL TSB culture tube and incubated for 16 to 18 h at 37 ± 2°C on a gyroratory shaker. After incubation, bacterial cells are pelleted by centrifugation (3,000 to 4,000 × g for approximately 15 min), washed twice with 0.1% proteose-peptone No. 3, and resuspended in 50 mL of 0.1% proteose-peptone. A standard plate count is conducted on the stock suspension and stored at 5 to 7°C. This plate count is used to determine the dilution required to prepare an initial bacterial suspension containing approximately 5 × 10^7 cfu/mL in TSB or skim milk broth media. When the colony count is 1 log under or over this goal value, the principal investigator should document the occurrence and evaluate the effect on the trial. When the colony count is 1 log under or over this goal value, the principal investigator should document the occurrence and evaluate the effect on the trial. Alter- 

Preparation of Strep. agalactiae. Cultures of Strep. agalactiae are prepared daily of experimental exposure by inoculating four 6-mL vials of TSB with stored stock culture or from 2 to 3 colonies of plated stock culture. The 6-mL culture tubes are incubated for approximately 15 h at 37 ± 2°C, and then the entire contents of 4 tubes are inoculated into a 500-mL flask of TSB, which is incubated for 6 to 7 h at 37 ± 2°C on a gyratory shaker. The culture is diluted using TSB or skim milk broth media to produce a culture concentration of approximately 5 × 10^7 cfu/mL as determined by optical density using a spectrophotometer (620 nm). When the colony count is 1 log under or over this goal value, the principal investigator should document the occurrence and evaluate the effect on the trial. Alternative preparation techniques that repeatedly produce bacterial suspensions of the desired bacterial concentration and result in the desired rate of IMI following experimental exposure of teats may be used but should be precisely documented in the trial report.

When both Staph. aureus and Strep. agalactiae are used for experimental exposure, as is usually the case, a portion of the Staph. aureus stock suspension is incorporated into the Strep. agalactiae TSB suspension and diluted with skim milk or broth medium to achieve approximately 5 × 10^7 cfu/mL of each pathogen.

Counts of the bacterial solution should be documented weekly. Trial staff may decide to perform experimental exposure with both organisms or with a single organism. The latter would become more relevant when several cows have experienced infections with 1 of the 2 organisms. The organism with many previous infections would be omitted from the experimental exposure. Experimental exposure procedures need to be documented.

Milking System Evaluation and Premilking Udder Preparation

A milking system evaluation should be performed before trial initiation and at least once during the trial to ensure that all machinery performs within the desirable range as defined by ISO 6690:2007, “Milking machine installations—Mechanical tests” (ISO, 2007). Any deviation of the milking equipment from the desirable range should be documented and evaluated with regard to its importance for the trial by the principal investigator. Premilking udder preparation methods should be decided before the start of the trial and recorded for inclusion in the final report. For trials requiring no premilking disinfection, premilking udder preparation should consist of the use of single-service towels that are free of sanitizer to clean the teats before forestripping. Forestripping is accomplished by expressing 2 squirts of milk. For trials requiring that a premilking teat disinfectant be used, each quarter should be forestripped by expressing 2 squirts of milk followed by dipping all teats with the designated premilking disinfectant. Approximately a 30-s contact time, or the time prescribed by the product manufacturer, should be allowed, and the teats should be wiped dry using a single-service towel before attaching the milking unit. The premilking disinfectant used should have been previously demonstrated to be effective through NMC recommended testing and listed in the NMC summary.

Experimental Exposure of Teats and Teat Disinfection

Each teat is experimentally exposed by immersion to a depth of approximately 25 mm in a beaker or
other suitable vessel of a freshly prepared TSB suspension of the test organism(s). Teats should be exposed to the bacterial suspension immediately after removal of the milking machine. Fresh TSB suspension of test microbes should be added as needed to maintain a depth of approximately 25 mm of suspension in each vessel. Experimental exposure should be performed at least once daily for a minimum of 5 d each week. Teat disinfectant (test and control) should be applied gently to the entire teat surface immediately after exposure to the bacterial suspension. Teat disinfectant is applied at every milking. Negative control teats remain not disinfected; no placebo product or water is used. Usage of teat disinfectant during the trial should be collected and reported. Teat disinfectant use information is valuable for comparison between products and as an aid in validating proper teat disinfectant usage during the trial. Duration of the trial is defined at a maximum of 12 wk unless otherwise specified before the start of the trial. In experimental exposure trials, the risk of new IMI tapers off over time and reaches low values after 10 to 12 wk into the trial (e.g., Leslie et al., 2006). Early evaluation of efficacy is acceptable at predefined time points before the end of the trial. These early evaluation time points should be defined before the start of the trial. Ending the trial upon achieving desired efficacy at predefined time points is acceptable.

**Sampling Schedules and Procedures**

Duplicate milk samples are collected from each lactating quarter to determine the bacteriological status of quarters within 7 d before initiation of the trial. A third sample is collected when results of the first 2 samples differ. Quarter foremilk samples are collected weekly during the trial. If any quarter secretes abnormal milk or develops signs of clinical mastitis, a milk sample must be obtained before treatment. All samples should be collected immediately before a regular milking. Sampling procedures are as described (Dodd et al., 1969). Briefly, 3 or 4 streams of foremilk are discarded from each quarter before sanitizing teat ends with cotton swabs and collecting samples. Sterile swabs should be soaked in 70% alcohol from which excess fluid has been expressed; no more than 2 teats should be sanitized per swab. The far teats should be sanitized first and the near teats last to avoid contaminating the near teats with the sampler’s hands. When collecting milk samples, the near teats should be sampled first and the far teats last to avoid contaminating the near teats with the sampler’s hands. This sanitization and collection procedure reduces the number of the bacteriologically false-positive or contaminated samples.

**Criteria for Diagnosing Infections**

All milk samples should be examined microbiologically according to standard procedures (Harmon et al., 1990). A new IMI in a quarter is diagnosed (Andersen et al. 2010) when the same bacterial species is isolated from (1) a single sample during the trial with $\geq 1,000$ cfu/mL; (2) 2 out of 3 consecutive samples during the trial with $\geq 100$ cfu/mL; or (3) a single sample from a quarter with clinical mastitis with $\geq 100$ cfu/mL. An individual quarter is eligible for only one new IMI per bacterial species during the trial. A quarter may receive treatment after a new IMI is confirmed, but the quarter is not eligible for another new IMI by the same pathogen. The bacterial strain in the IMI may be confirmed as the challenge strain by molecular typing (Zadoks et al., 2011); however, this is not essential because any IMI with the bacterial species involved in the study would be considered a valid IMI under the indicated definitions.

**Statistical Analysis**

The unit of analysis in experimental exposure trials is the individual quarter of a cow. It is recognized that quarters within cow are correlated and statistical methods will need to account for this potential correlation (Barkema et al., 1997). Data on efficacy of a teat disinfectant formulation are examined in terms of reduction in rate of new IMI achieved among quarters disinfected with the test product compared with the rate among control quarters. The precision of rate reduction is used to decide statistical significance. Data summarized from the trial must express the quarters becoming infected in quarters dipped with the test disinfectant and in control quarters (whether this is a positive or negative control). Differences between the rate of quarters becoming infected between treatments can be tested using several statistical methods. For each quarter, the presence or absence of infection in combination with the time to infection should be incorporated in the statistical analysis. Two separate models, one for *Staph. aureus* and one for *Strep. agalactiae*, will need to be run. Statistical methods should be capable of controlling for known risk factors for infection, such as teat location and age of the cow. Valid methods for analyses include time-to-event methods such as Cox proportional hazards models and parametric time-to-event models. Generalized linear models that take the time at risk for infection into account are also valid for statistical analysis of the data. In all analyses, days should be used as the unit of time. Statistical significance should be evaluated at the type I error level of 5%.
As an example, a general formulation of the generalized linear mixed model is proposed:

\[ f(Y_{ijk}) = \text{intcpt} + b_1 \times \text{treatment} + b_2 \times \text{exposure} \]
\[ + \ b_k \times \text{covariates}_k + \text{cow}_j(\text{random}) + \text{error}_{ijk}, \]

where \( f(Y) \) is the link function (usually a log or logit link is chosen), \( Y_{ijk} \) is the occurrence of a new infection (yes vs. no) in a quarter \( i \), \( \text{intcpt} \) is the intercept, \( \text{treatment} \) is the variable indicating whether a quarter is in the treatment or the control group, \( b_1 \) refers to the regression coefficient for treatment, \( \text{exposure} \) is the total number of quarters shedding in a given week in the trial, and \( \text{covariates} \) is any additional covariate of interest. Covariates could include the quarter location (front vs. hind), age of the cow, week of the trial, and other potential factors affecting the rate of new infections. Finally, \( \text{cow}_j \) is an indicator for each individual cow that is used in a random cow effect, and \( \text{error} \) refers to the residual term in the model. Generally, a Poisson-distributed error term should be used although a binomial error term may also be appropriate (Schukken et al., 2003). In general, the canonical link function should be used. To estimate risk difference in the case of a positive control trial, a linear link will be necessary. Because the time period at risk for each observation is the same (7 d), there is no need to include an offset term in the model. When different times at risk are present between observations, than the natural log of the time-at-risk should be used as the offset in a Poisson regression model (Schukken et al., 2003). Because the standard sample interval is 7 d, the at-risk period can be divided piece-wise into 7-d periods. Coding of the data is on a 7-d basis, where for every quarter every 7 d, there is a new line of data. The data file would be developed as shown in Table 1. In this table, a variable named \( \text{At-riskSau} \) or \( \text{At-riskAg} \) indicates whether the quarter should be included in the analysis or not. A value of 0 for \( \text{At-riskSau} \) as shown for a quarter previously infected with \( \text{Staph. aureus} \) would eliminate an observation from the analysis.

When treatment is coded such that the new product is coded as 1 and the control is coded as 0, the value of \( \beta_1 \) is the log of the ratio of rate of new infections between the new product and control. To obtain the actual ratio, the value of \( \beta \) is exponentiated. The estimate for new product efficacy relative to control is defined as

\[ \text{Efficacy (new product vs. control)} = 1 - \exp(b_1). \]

A confidence interval of the efficacy can be obtained by

\[ 1 - \exp[b_1 \pm 1.96 \times \text{SE}(b_1)], \]

where \( \text{SE}(b_1) \) is the standard error of the \( b_1 \) estimate. The estimated incidence rate of new infections in the treatment and control group can be obtained using the inverse link function and, in the linear predictor, the intercept, treatment, and the remainder of the covariates at their mean value.

**Trial Size**

Sample size estimations for regression analyses are not straightforward and may require extensive input of information on covariates. To keep sample size calculation somewhat straightforward here, a difference in the proportion of infected quarters between a test product and a control was utilized. In general, the estimated sample sizes calculated using simple proportions are somewhat larger than estimates obtained using a method that is specifically suitable for multiple regression.

**Negative Control Trial.** In a negative control trial, the objective is to show superiority of the teat disinfec-
tant compared with a nondisinfected control. The null ($H_0$) and alternative ($H_A$) hypotheses can be defined as follows:

$$H_0: p_1 = p_2$$

and

$$H_A: p_1 < p_2,$$

where $p_1$ and $p_2$ are the risk of new infection in the teat disinfection group and the control group, respectively. Note that the definition of the null and alternative hypothesis calls for one-sided testing. Sample size per treatment group can be approximately calculated by the following formula proposed by Casagrande and Pike (1978) and Wickramaratne (1995):

$$E_i = \left(\frac{Z_{1-\alpha/2} + Z_{1-\beta}}{2} \left(\sqrt{R} - 1\right)\right)^2,$$

where $E_i$ is the number of events, $E$ (e.g., new IMI), in each treatment group, $Z_{1-\alpha/2}$ is the value of the standard normal distribution evaluated at $1 - \alpha/2$, $Z_{1-\beta}$ is the value of the standard normal distribution evaluated at $1 - \beta$, and $R$ is the risk of new infection in the disinfected group/risk of new infection in control group.

For teat disinfection experimental exposure trials, the sample size in the control group is always equal to the sample size in the treated group. In this notation, $E_1$ and $E_2$ are the number of new IMI in the 2 groups required to detect a relative risk of size $R$ with power $1 - \beta$. The number of subjects per group, $N$, is then equal to $E_1/p_1$, where $p_1$ is the proportion infected in the controls. Examples of trial sizes calculated using this formula are shown in Table 2.

For the purpose of evaluation of teat disinfectants, sample size is based on finding a ratio of infections in the treatment group and control group of 0.6 or smaller. This is assuming that an acceptable efficacy is at least 40% ($= 1 - 0.6$). The sample size formula leads to the suggested sample sizes per group that are shown in Table 2. Typically, an infection risk of approximately 20% may be expected in the control group (i.e., Drechsler et al., 1993). With an anticipated efficacy of 60 to 70%, the sample size per group would be between 76 and 115 quarters per group, or a total of 40 to 60 cows in the trial. The greater the expected efficacy of the teat disinfectant against the negative control, the smaller the trial size needed to determine efficacy. Also, with a higher proportion of quarters infected in the control group and the same teat disinfectant efficacy, a smaller sample size will be necessary.

**Positive Control Trial.** The objective of a positive control trial is to show noninferiority of a test product versus a control product. Figure 1 shows the principle of a noninferiority trial, depicting the difference between 2 products along the horizontal axes. When the observed difference between 2 products is equal to zero (or the relative efficacy is equal to zero) the products may be considered equivalent. Even if the test product is slightly worse than the positive control, the products may still be considered equivalent. Only when the difference is larger than a predetermined critical value ($d$, in Figure 1) is it concluded that the test product is inferior to the control (Piaggio et al., 2006). The null and alternative hypotheses can be defined as follows:

$$H_0: (p_1 - p_3) \geq d \quad \text{(the interpretation is that } p_1 \text{ is inferior to } p_3)$$

and

$$H_A: (p_1 - p_3) < d \quad \text{(the interpretation is that } p_1 \text{ is not inferior to } p_3),$$

where $p_1$ is the risk of infection in the test disinfectant group, and $p_3$ is the risk of infection in the positive control disinfectant group. Again, the definitions of null and alternative hypotheses are based on one-sided testing.

In Figure 2, this translates to a point estimate of the difference between 2 products and the 95% CI of this difference. From Figure 2, it can be seen that noninferiority is shown when the right-hand, or upper, limit of the confidence interval of the difference is <d. This is irrespective of the actual observed difference (i.e., d could be smaller or larger than 0).

<table>
<thead>
<tr>
<th>Negative-control IMI risk (p2)</th>
<th>Efficacy of test product</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>40%</td>
</tr>
<tr>
<td>30%</td>
<td>203</td>
</tr>
<tr>
<td>25%</td>
<td>244</td>
</tr>
<tr>
<td>20%</td>
<td>265</td>
</tr>
<tr>
<td>15%</td>
<td>247</td>
</tr>
</tbody>
</table>

![Table 2. Number of eligible quarters per treatment group needed to compare 2 proportions with a one-tailed test, assuming 80% power and one-sided 5% significance level in a negative control trial](image)

**Figure 1.** The principle of a noninferiority trial. The difference, $d$, between incidence rate of IMI in test product versus control is depicted. When the difference in the rate of IMI is larger than $d$, the null hypothesis of inferiority of the test product cannot be rejected.

Journal of Dairy Science Vol. 96 No. 4, 2013
For teat disinfection trials, the positive control product should have a published efficacy against the negative control of at least 70% for the challenge organisms of interest. Efficacy of several teat disinfectants meeting this 70% efficacy requirement for both Staph. aureus and Strep. agalactiae have been reported in the NMC summary (e.g., Boddie et al., 1998; Foret et al., 2003). In general, a teat disinfection product is considered to be adequate if the efficacy is at least 40%. Hence, compared with a positive control, the difference in efficacy should be 30% or less for the test product to still meet the NMC efficacy criteria. This can be shown mathematically by defining \( p_i \) as the risk of IMI in group \( i \), and \( p_1 = \) risk in test disinfectant, \( p_2 = \) risk in negative control, \( p_3 = \) risk in positive control disinfectant. Therefore, \( 1 - (p_1/p_2) > 0.4 \) and we define \( 1 - (p_3/p_2) = 0.7 \). Then, \( p_1/p_2 = (p_1/p_3) \times (p_3/p_2) \), so, \( 1 - (0.3 \times p_1/p_3) > 0.4 \) or \( p_1/p_3 < 2 \). This means that the risk of infection in the test product divided by the risk of infection in the positive control group should be <2. To relate this relative effect (efficacy is based on a ratio of 2 infection risks) to the critical value of the difference in infection risk in treatment and positive control (\( d \)), the following formula applies: \( d = p_1 - p_3 \); for \( p_1 \leq 2 \times p_3 \), this becomes \( d = 2 \times p_3 - p_3 = p_3 \). For \( n = n_1 = n_3 \), the sample size per group, \( N \), is therefore given by the following formula (Christensen, 2007): 

\[
N = (Z_{1-\alpha/2} + Z_{1-\beta})^2 \times \frac{[p_1(1 - p_1) + p_3(1 - p_3)]}{d^2},
\]

where \( Z_{1-\alpha} \) is the value of the standard normal distribution evaluated at \( 1 - \alpha/2 \) and \( Z_{1-\beta} \) is the value of the standard normal distribution evaluated at \( 1 - \beta \). Using an expected IMI risk in the positive control between 5% and 7.5% (e.g., Leslie et al., 2005, 2006), the trial size for a noninferiority trial would have to be between 111 and 171 quarters per treatment group, or between 56 and 86 cows in the trial, approximately 20 cows larger than in a negative control superiority trial (Table 3). In Figure 3, the power of the trial is calculated for a noninferiority trial when the risk of infection in the positive control group was estimated at 8.5%. This IMI risk estimate of 8.5% for disinfected quarters was obtained from the data shown in the example below.

**Presentation of Data**

Reporting guidelines for trials in food animals have been described in detail (O’Connor et al., 2010). These guidelines, summarized in the REFLECT (Reporting Guidelines for Randomized Control Trials) statement, pertain fully to the reporting of trials on the efficacy of postmilking teat disinfectants. The report of an efficacy trial should include information on the design considerations before starting the trial. This would include the choice of the treatment and control products, the
anticipated trial duration, the predetermined evaluation time points, the calculated sample size including assumptions about type I and type II error, and the statistical methods used. The herd and cows used for the trial should be described along with the results of equipment testing and the result of teat end scoring. Methods used for bacteriological analyses, bacterial growth methods, and colony counts of the challenge organisms during the trial should be documented. For each treatment group, the number of quarters initially eligible for infection, the total quarter-days at risk, and the number and rate of new IMI by each test organism should be reported. Usage of the volume of disinfectant (test and control products) should be reported. In the statistical analysis, the efficacy of the test product should be estimated in a properly controlled statistical procedure. The statistical procedures and the results of the analysis should be presented in sufficient detail to allow evaluation by the reader. For positive control trials, the estimated difference between treatment and control with the appropriate 95% CI should be reported (Piaggio et al., 2006).

### Example

A negative control experimental exposure trial was performed to evaluate the efficacy of an experimental teat disinfectant. The trial was performed in a university herd in a total of 171 eligible quarters in 43 cows during a period of 50 d. Equipment testing before and during the trial revealed that all parameters were within the desirable range. No quarters were excluded before or during the trial for teat abnormalities. One cow had one nonfunctional quarter. All quarters of all cows were challenged 5 d/wk at the morning milking. Colony counts in the challenge broth for each pathogen ranged between $1.3 \times 10^7$ and $8.1 \times 10^7$ cfu/mL. All left quarters were immersed in the postmilking teat disinfectant at all milkings, whereas the right quarters were left untreated. Over the duration of the trial, approximately 50 L of teat disinfectant was used, slightly more than 10 mL per cow per milking. Bacteriological analysis was done according to NMC standards (Harmon et al., 1990; NMC, 1999), and infection definition was as described in this protocol. No quarters infected with \textit{Staph. aureus} or \textit{Strep. agalactiae} were present before the start of the trial. For ease of presentation, only the \textit{Staph. aureus} data are shown. The number of new \textit{Staph. aureus} infections in 86 control quarters was 22 in 3,454 quarter-days at risk and in 85 disinfected quarters was 7 in 3,812 quarter-days at risk. These IMI numbers resulted in IMI risks of 0.026 and 0.082 for the control and disinfected quarters, and IMI rates of 0.0064 and 0.0018 per quarter-day at risk for the 2 groups, respectively. Comparing the IMI proportions provides an efficacy estimate of 67.8% with a Chi-squared value of 9.13 and a \(P\)-value of 0.0025. Comparing the IMI rates per day provides an efficacy estimate of 71.2% with a \(P\)-value of 0.0023.

Front quarters had 11 IMI and hind quarters had 18 \textit{Staph. aureus} IMI. For first-lactation animals, 2 out of 28 quarters (7.1%) had a \textit{Staph. aureus} IMI; in lactation \(\geq 2\) animals, 27 out 143 quarters (18.9%) had a \textit{Staph. aureus} IMI. The rate of infection in disinfected and control quarters over the duration of the trial is shown in Figure 4. Note the relatively high rate of IMI in control quarters in the first 15 d of the trial, resulting in an immediate reduction of the number of quarter-days at risk in this group.

### Table 3.

<table>
<thead>
<tr>
<th>IMI risk in control (p%</th>
<th>Acceptable difference, d</th>
<th>Sample size, n</th>
</tr>
</thead>
<tbody>
<tr>
<td>1%</td>
<td>0.01</td>
<td>890</td>
</tr>
<tr>
<td>2.5%</td>
<td>0.025</td>
<td>350</td>
</tr>
<tr>
<td>5.0%</td>
<td>0.05</td>
<td>171</td>
</tr>
<tr>
<td>7.5%</td>
<td>0.075</td>
<td>111</td>
</tr>
<tr>
<td>10.0%</td>
<td>0.10</td>
<td>81</td>
</tr>
<tr>
<td>12.5%</td>
<td>0.125</td>
<td>63</td>
</tr>
<tr>
<td>15.0%</td>
<td>0.15</td>
<td>51</td>
</tr>
</tbody>
</table>

Figure 3. The power of the trial is shown for increased sample size per group when the risk of IMI in the positive control group is estimated at 8.5%.
Statistical analysis of the data was done using Cox proportional hazards analysis with cow as a random effect (frailty) and by Poisson regression analysis using a piece-wise regression model as presented above. Covariates included in the regression models were hind versus front quarter, lactation 1 versus lactation 2+, and, for the Poisson regression, the number of quarters shedding *Staph. aureus* in each week. In the Poisson regression model, an intra-cow correlation was estimated to correct for lack of independence of quarters within cows. All analyses were done using SAS software (SAS Institute, 2009).

The results of the piece-wise Poisson regression analysis showed a significant within-cow clustering [variance inflation factor = 1.84 (0.08)]. The regression parameters are shown in Table 4; Cox proportional hazard model results are also shown in Table 4. When comparing the efficacy estimates of the teat disinfectant using the 3 presented methods, the Poisson regression yields an estimate of 72.6% (= 1 − exp^{−1.3}) with a 95% CI from 0.14 to 0.92, Cox proportional hazard 70.9% (= 1 − exp^{−1.24}), with a 95% CI from 0.01 to 0.92. The simple IMI proportions yield an estimate of 67.8%. The P-values evaluating the significance of the teat disinfectant using these 3 methods were, respectively, 0.028, 0.026, and 0.0024. The lower P-values for the multivariate methods reflect a more appropriate correction for confounding and within-cow correlation. From the regression models, it is obvious that first-lactation animals have a lower risk of IMI and hind quarters have a higher risk of IMI. Neither risk factor was statistically significant. In the piece-wise Poisson regression model, the time-dependent number of quarters shedding *Staph. aureus* was a significant predictor of new IMI. This would indicate that new IMI are not only due to experimental exposure to *Staph. aureus* but also due to contagious transmission of *Staph. aureus* between quarters and between cows. Our results indicate that correcting for this additional infection mechanism resulted in a somewhat higher estimate for teat disinfectant efficacy and a lower P-value. Given the correlated nature of IMI within cow, this result is expected.

### DISCUSSION

The experimental exposure trial as described here provides a guideline for performing trials to evaluate postmilking teat disinfectants. This updated version of the NMC protocol has attempted to provide a more precise description of a standardized experimental exposure trial and has been significantly changed with regard to statistical analyses and sample size calcula-
tions. The most important changes compared with the previous protocol result from years of experience in performing experimental exposure trials (e.g., Pankey et al., 1985; Leslie et al., 2005, 2006) and changing requirements with regard to the precision and repeatability of performing standardized trials (EMEA, 2009). Because of the inclusion of appropriate quality assurance methodologies in the protocol, a more definite description of trial requirement and trial procedures has now been provided. New insights into statistical analysis of complex data have resulted in a change in the description of the statistical analyses. The increased understanding of the correlated nature of quarter IMI data (Schukken et al., 2003) and the importance of correcting for contagious transmission has required the use of more complex regression models. Even though these changes in protocol design and statistical analysis will lead to more elaborate data collection and manipulation, the biological considerations underlying the need for correction for correlated data and external exposure are important enough to warrant such a change. Experimental exposures trials are large and expensive and for that reason alone deserve the best possible collection and analysis of complex data.

ACKNOWLEDGMENTS

Members of the NMC teat health committee and the NMC research committee are acknowledged for critical review of the proposed trial protocols. The protocols described in this manuscript were approved by both the teat health committee and the research committee of the NMC (Verona, WI).

REFERENCES


