



A statistical model for assessing performance standards for quantitative and semi-quantitative disinfectant test methods

Authors: Albert E. Parker, Martin A. Hamilton, & Stephen F. Tomasino

NOTICE: This is a postprint of an article that originally appeared in Journal of AOAC International on January 2014. DOI: <http://dx.doi.org/10.5740/jaoacint.13-105>.

Parker A, Hamilton M, Tomasino S, "A statistical model for assessing performance standards for quantitative and semi-quantitative disinfectant test methods," Journal of AOAC International, January 2014 97(1): 58-67.

A Statistical Model for Assessing Performance Standards for Quantitative and Semiquantitative Disinfectant Test Methods

ALBERT E. PARKER

Montana State University, Center for Biofilm Engineering, Department of Mathematical Sciences, Bozeman, MT 59715

MARTIN A. HAMILTON

Big Sky Statistical Analysts LLC, Bozeman, MT 59715

STEPHEN F. TOMASINO¹

U.S. Environmental Protection Agency, Office of Pesticide Programs, Biological and Economic Analysis Division, Microbiology Laboratory Branch, Environmental Science Center, Fort Meade, MD 20755-5350

A performance standard for a disinfectant test method can be evaluated by quantifying the (Type I) pass-error rate for ineffective products and the (Type II) fail-error rate for highly effective products. This paper shows how to calculate these error rates for test methods where the log reduction in a microbial population is used as a measure of antimicrobial efficacy. The calculations can be used to assess performance standards that may require multiple tests of multiple microbes at multiple laboratories. Notably, the error rates account for among-laboratory variance of the log reductions estimated from a multilaboratory data set and the correlation among tests of different microbes conducted in the same laboratory. Performance standards that require that a disinfectant product pass all tests or multiple tests on average, are considered. The proposed statistical methodology is flexible and allows for a different acceptable outcome for each microbe tested, since, for example, variability may be different for different microbes. The approach can also be applied to semiquantitative methods for which product efficacy is reported as the number of positive carriers out of a treated set and the density of the microbes on control carriers is quantified, thereby allowing a log reduction to be calculated. Therefore, using the approach described in this paper, the error rates can also be calculated for semiquantitative method performance standards specified solely in terms of the maximum allowable number of positive carriers per test. The calculations are demonstrated in a case study of the current performance standard for the semiquantitative AOAC Use-Dilution Methods for *Pseudomonas aeruginosa* (964.02) and *Staphylococcus aureus* (955.15), which allow up to one positive carrier out of a set of 60 inoculated and treated carriers in each test. A simulation study was also conducted to verify the validity of the model's assumptions and accuracy. Our approach, easily implemented using the computer code provided, offers a quantitative decision-making tool for assessing a performance standard for any disinfectant test method for which log reductions can be calculated.

Introduction

This paper presents a general statistical approach for assessing a performance standard for any disinfectant test method for which a log reduction (LR) in a microbial population can be calculated. The assessment is based on a conventional hypothesis testing framework and multivariate t probability calculations that account for multiple tests of multiple microbes conducted in (possibly) multiple laboratories.

A disinfectant test method quantifies the antimicrobial efficacy of a disinfectant; efficacy is typically determined by comparing a measure of the viable microbes on inoculated carriers that have been treated with the disinfectant to the measure of viable microbes on untreated control carriers. A performance standard (PS) for a disinfectant test method defines an acceptable outcome for a product being tested. The acceptable outcome can be specified per test, with the requirement that the product passes all of multiple tests, or stipulated as an average over multiple tests. The PS may also specify the number of tests that must be performed, the number of laboratories required to conduct the tests, and the test microbes used. Throughout this paper, there are two different uses of the word microbe(s) that will be made clear from context; in the previous sentence, microbes indicates the different microbial species, such as *Pseudomonas aeruginosa* and *Staphylococcus aureus*, to be used in the disinfectant test method, while viable microbes indicates individual living cells of a certain microbial species. It is important to emphasize that while a PS may require that a method be applied in multiple tests against multiple microbes, we assume that each individual test is against a single microbial species.

The statistical evaluation reported here focuses on the pass-error and fail-error rates associated with a PS for a method (1). In this context, rate is a synonym for probability. Errors occur when application of the PS to the test results for a disinfectant product lead to the wrong conclusion. A pass-error, which corresponds to a Type I error in statistical hypothesis testing, occurs when the disinfectant being tested truly has unacceptably low efficacy but is deemed a pass by the PS. A fail-error, corresponding to a Type II error in statistical hypothesis testing, occurs when the disinfectant truly is of acceptable efficacy, but is deemed a failure.

The error rate evaluations pertain to PSs for quantitative methods, such as the AOAC Three Step Method (2), and to semiquantitative methods, such as the AOAC Use-Dilution Method (UDM; 3–5) and Hard Surface Carrier Test (HSCT; 6). For a quantitative method, all inoculated carriers, both untreated control carriers and those treated with disinfectant, are enumerated for viable microbes. Thus, the measure of efficacy for quantitative methods is the LR (7). For a semiquantitative method, the density of viable microbes is measured only on the control carriers, while each treated carrier is recorded as either a positive or negative for the presence or absence of viable microbes, as determined by turbidity of the broth culture containing the carrier. Although test outcomes for semiquantitative methods, such as the UDM and HSCT, have historically been reported using the number of positive carriers observed among a set of inoculated and treated carriers, the availability of microbial densities on the control carriers makes it possible to calculate a LR (7). Thus, the existing PSs for semiquantitative methods, specified with respect to a maximal number of positive carriers per test, can also be evaluated using the techniques presented in this paper.

The error rates can be calculated for PSs that require multiple tests of each required microbe. While other researchers have considered disinfectant method error rates for a single-test PS of each microbe separately (8–11), our approach is to calculate the error rates across all required tests of all microbes by accounting for the correlation among tests of different microbes conducted in the same laboratory. The error rate calculations provided here also account for among-laboratory sources of variability, a critical aspect of method performance evaluations. In other words, it is desirable to understand the reproducibility of a method across multiple laboratories prior to its use for product registration. If cost were not an issue, producers of a disinfectant product could conduct an interlaboratory study for product registration, directly estimate the among-laboratory variance, and then conduct a hypothesis test or construct a confidence interval using the newly generated study data to decide whether the product attains some target LR with a high level of confidence. Unfortunately, an interlaboratory study is not a practical or cost-effective approach. In this paper, techniques that use existing interlaboratory data of a method are demonstrated.

The goals of this paper are to describe the hypothesis-testing framework and the subsequent statistical techniques used for calculating error rates. As an example, the error rates for the PS associated with a widely used disinfectant method, the UDM, and two microbes, *P. aeruginosa* (5) and *S. aureus* (4), will be estimated based on data from a recent five-laboratory collaborative study (12). The error rates, easily calculated using the computer code provided in an appendix of this paper, offer a quantitative decision-making tool for assessing a PS for any disinfectant test method for which LRs can be calculated.

Using Error Rates to Assess a Performance Standard

In what follows, the statistical parameter of interest for a disinfectant product is the true mean LR, μ_{LR} , over multiple tests conducted using the method (other percentiles are considered in Appendix A; see appendixes on *J. AOAC Int.* website, <http://aoac.publisher.ingentaconnect.com/content/aoac/jaoac>). The disinfectant product is deemed effective if and only if

$\mu_{LR} > LR_{target}$, where the target LR, LR_{target} , is a numerical value set by a stakeholder (e.g., a regulatory agency). An ineffective disinfectant product has $\mu_{LR} \leq LR_{target}$. To calculate the pass-error and fail-error rates for a given PS, one can use conventional statistical hypothesis testing techniques where the null and alternative hypotheses are

$$H_0: \mu_{LR} \leq LR_{target}; H_a: \mu_{LR} > LR_{target} \quad (1)$$

The null hypothesis H_0 states that the disinfectant product is ineffective. The manufacturer of the disinfectant product anticipates that testing and application of the PS will lead to rejection of H_0 in favor of H_a , which states that the disinfectant is effective.

It is important to emphasize that a specified PS determines the rejection region for the hypothesis test (13). That is, if the testing outcome passes the PS, the outcome is in the rejection region, H_0 is rejected, and one concludes that the disinfectant product is effective. That relationship between the PS and the hypothesis test rejection region justifies the hypothesis testing framework that we use for calculating the error rates for a PS.

The error rates for a PS are the probabilities of Type I and Type II errors (13) when using the PS to test the hypotheses in Equation 1. The Type I error probability, α , also known as the significance level, is the pass-error rate of a PS. In other words, the pass-error rate is the probability that an ineffective product passes the PS. The Type II error probability, β , can be calculated for any μ_{LR} under H_a . The fail-error rate of a PS is the probability of a Type II error for a highly effective product, which is a product having $\mu_{LR} \geq LR_{high}$, where LR_{high} is a numerical value specified by the technical authority. As is conventional in statistical hypothesis testing, the pass-error rate calculations will pertain to a product with $\mu_{LR} = LR_{target}$ for which the Type I error rate is largest among all ineffective products. The fail-error rate calculations will pertain to a product with $\mu_{LR} = LR_{high}$ for which the Type II error rate is largest among all highly effective products.

The error rate calculations depend on the values chosen for LR_{target} and LR_{high} . We demonstrate how to perform the error rate calculations within the hypothesis testing framework, assuming that LR_{target} and LR_{high} have been specified. For semiquantitative tests, we demonstrate how to evaluate traditional PSs based only on numbers of positive carriers by showing how to determine the appropriate values of LR_{target} and LR_{high} .

Steps for Error Rate Calculations

The process for calculating the error rates is separated into six steps. Steps 1–4 illustrate how to evaluate the error rates for a PS that requires only a single test. Step 5 illustrates how to evaluate a PS that requires multiple tests with a single microbe; and the disinfectant must pass all individual tests. Step 6 illustrates how to calculate the error rates for a PS that requires multiple tests with each of the multiple microbes, and the disinfectant must pass all individual tests. The details for evaluating an alternative PS which requires that a disinfectant passes multiple tests on the average are provided in Appendix A.

Step 1: Calculate the LR for each test.—For a quantitative method, every inoculated carrier, both untreated controls and those treated with the disinfectant, are enumerated for viable microbes. The density of viable microbes is commonly

quantified by enumerating CFUs per carrier or per cm² of carrier surface area. The log₁₀-transformed density is called the log density (LD) of microbes. The mean of the LDs across the control carriers in a test is:

$$\bar{LD}_{control}$$

where the LD for the i^{th} control carrier in a test is denoted by $LD_{control,i}$ and the total number of control carriers is given by $n_{control}$. The LR for each test is found by subtracting the mean of the LDs for the treated carriers from the *TestLD*

$$\bar{LD}_{treated}$$

where the LD for the i^{th} treated carrier in a test has been denoted by $LD_{treated,i}$ and the total number of treated carriers is given by $n_{treated}$. For a semiquantitative method, the *TestLD* is still calculated for the control carriers, while the number of positive carriers, denoted by N , out of the total number of inoculated and treated carriers, $n_{treated}$, is recorded for each test. Interestingly, from N and $n_{treated}$, the most probable number (MPN) of viable microbes on the typical treated carrier (14–19) can be found by

$$\text{MPN}(N) = \frac{N}{n_{treated}} \quad (2)$$

The LR for a semiquantitative test can now be calculated as a function of N and *TestLD*; $n_{treated}$ is presumed fixed (15, 18):

$$LR(N; \text{TestLD}) = \text{TestLD} - \log_{10}[\text{MPN}(N)] \quad (3)$$

The LR disinfectant efficacy measure has been used for various semiquantitative disinfectant test methods, such as the UDM (12), HSCT (11), and a sporicide activity test (20). In many studies, including (11, 12, 20), the LR values were approximately normally distributed.

Step 2: Estimate the among-laboratory variability of the LRs.—In order to calculate the appropriate test statistic to evaluate the hypotheses in Equation 1, it is necessary to generate an estimate of the variability of the LRs across multiple laboratories for each microbe separately. Thus, a multilaboratory study specifically designed to generate method performance data is highly desirable. These data are often associated with the validation of new test methods or studies designed to improve existing test methods. Unfortunately, the number of controlled multilaboratory studies for disinfectant test methods is somewhat limited. Thus, it is important to use whatever previously completed multilaboratory study of a test method is already available, especially if the study was well-documented and conducted with a high degree of quality control. Given LRs from an existing study, an estimate of the among-laboratory variance, S_{lab}^2 , is found by fitting, for each microbe separately, a one-way analysis of variance (ANOVA) with a random effect due to laboratory (7, 21). This ANOVA also provides an estimate of the within-laboratory variance, S_{test}^2 , also known as the repeatability variance. The reproducibility SD, S_R , is calculated from the variance components (7, 21) by

$$S_R = [S_{lab}^2 + S_{test}^2]^{1/2} \quad (4)$$

The SD S_R is used in all subsequent calculations of the pass-error

and fail-error rates. The Satterthwaite approximate degrees of freedom associated with S_R^2 is:

$$\frac{df_R}{S_R^2} = \frac{df_{lab}}{S_{lab}^2} + \frac{df_{test}}{S_{test}^2} \quad (5)$$

where I is the number of laboratories that took part in the study used to generate the estimates of the variance components, J the number of repeated tests in each laboratory, and $H = S_{lab}^2/S_{test}^2$ (22). The Satterthwaite approximation is based on the fact that $df_R S_R^2/E(S_R^2)$ is approximately a $\chi^2(df_R)$ random variable when the LRs are normally distributed (23, p. 971), where $E(\cdot)$ denotes the expectation operator. Nijhuis & Van den Heuvel (24) showed that the χ^2 approximation is particularly good and that the use of df_R leads to reliable probability calculations.

Step 3: Performance standard for a single test.—Consider a PS that passes a disinfectant if and only if the observed LR for a single test is at least as large as some specified numerical value LR_{PS} . The corresponding rejection region for the hypotheses in Equation 1 is:

$$RR_1 = \{\mathbf{LR} \geq LR_{PS}\} \quad (6)$$

where the **bold** font indicates random variables. The value of LR_{PS} is commonly chosen to ensure that the Type I pass-error rate, α , is no larger than some specified value (e.g., 0.05). The subscript 1 indicates that the rejection region pertains to a single test PS.

Step 4: Error rates for a single test performance standard.—The previous steps provide sufficient information to derive the test statistic, its sampling distribution, and subsequently, the error rate equations. After some algebra, the rejection region in Equation 6 can be rewritten as

$$RR_1 = \{(\mathbf{LR} - LR_{target})/S_R \geq (LR_{PS} - LR_{target})/S_R\} \\ = \{\mathbf{T} \geq t_1\} \quad (7)$$

The test statistic is

$$\mathbf{T} = (\mathbf{LR} - LR_{target})/S_R \quad (8)$$

and the value that defines the one-sided rejection region is:

$$t_1 = (LR_{PS} - LR_{target})/S_R \quad (9)$$

If the LRs are normally distributed (*see Discussion*), then \mathbf{T} approximately follows a t distribution with degrees of freedom df_R given by Equation 5. The pass-error rate can now be calculated (13, section 8.3.1) using

$$\alpha_1 = Pr(\text{Rejecting } H_0 \mid \mu_{LR} = LR_{target}) \\ = Pr(\mathbf{T} \geq t_1; df_R, \lambda = 0) \quad (10)$$

where the probability function $Pr(\cdot)$ is evaluated using the t -distribution with degrees of freedom df_R and non-centrality parameter $\lambda = 0$. The fail-error rate of a highly effective treatment using a single test PS is

$$\beta_1 = Pr(\text{Failing to reject } H_0 \mid \mu_{LR} = LR_{high}) \\ = Pr(\mathbf{T} < t_1; df_R, \lambda = \lambda_1) \quad (11)$$

where the non-centrality parameter is given by

$$\lambda_1 = (LR_{\text{high}} - LR_{\text{target}})/S_R \quad (12)$$

Equations 4, 5, and 9–12 make clear how the variance components S_{lab}^2 and S_{test}^2 from the existing multilaboratory study are used, in conjunction with LR_{PS} , LR_{target} , and LR_{high} to calculate the error rates α_1 and β_1 .

Step 5: Error rates for a performance standard that requires that a disinfectant passes all tests for a single microbe.—Steps 1–4 show how to evaluate a PS for a single test per microbe; this may be used to monitor product efficacy claims following registration. However, for other purposes where multiple tests are required (e.g., for product registration), $K > 1$ tests per microbe are required in each of L laboratories, and a product must pass all $K \cdot L$ tests. This multiple testing scenario corresponds to simultaneously testing $K \cdot L$ hypotheses of the form shown in Equation 1. It is important to emphasize that I and J denote the number of laboratories and tests in a previously completed multilaboratory study, while L and K denote the number of laboratories and tests in the testing scheme required by the PS.

The pass-error rate for the “pass-all” PS that requires conducting K tests in each of L laboratories is given by

$$\alpha_{KL} = Pr(\text{Rejecting } H_0 \text{ in all } K \cdot L \text{ tests} \mid \mu_{LR} = LR_{\text{target}}) \quad (13)$$

$$= Pr(T \geq t_1 \text{ in all } K \cdot L \text{ tests}; df_R, \lambda = 0 \mid \mu_{LR} = LR_{\text{target}})$$

and the probability that a highly effective product fails to pass-all $K \cdot L$ tests is

$$\beta_{KL} = Pr(\text{Failing to reject } H_0 \text{ in at least one test} \mid \mu_{LR} = LR_{\text{high}})$$

$$= 1 - Pr(\text{Rejecting } H_0 \text{ in all } K \cdot L \text{ tests} \mid \mu_{LR} = LR_{\text{high}})$$

$$= 1 - Pr(T \geq t_1 \text{ in all } K \cdot L \text{ tests}; df_R, \lambda = \lambda_1 \mid \mu_{LR} = LR_{\text{high}}) \quad (14)$$

In the special case where the $K \cdot L$ tests are statistically independent, these error rates are easily calculated using

$$\alpha_{KL} = (\alpha_1)^{K \cdot L} \quad (15)$$

and

$$\beta_{KL} = 1 - (1 - \beta_1)^{K \cdot L} \quad (16)$$

However, in practice, tests conducted in the same laboratory are usually not statistically independent. Therefore, Equations 15 and 16 are seldom appropriate. A measure of this dependence is the intralaboratory correlation coefficient, estimated by (ref. 23, section 24.1.)

$$r = S_{\text{lab}}^2 / (S_{\text{lab}}^2 + S_{\text{test}}^2) \quad (17)$$

To properly account for the correlation among tests of the same microbe in the same laboratory, a multivariate t -distribution can be used to calculate α_{KL} and β_{KL} (25). The multivariate t -distribution requires the same inputs, t_1 , df_R , and λ , that were required for Equations 10 and 11. An additional input is the block diagonal matrix of correlations among the $K \cdot L$ tests (ref. 23, p. 961–962). Statistical software for these calculations is readily available (Appendix B).

Step 6: Error rates for a performance standard that requires that a disinfectant passes all tests for each of multiple microbes.—Step 5 considers the error rates for each microbe

separately; however, PSs commonly require that a disinfectant pass all tests for a set of microbes to meet regulatory labeling requirements. To calculate the pass-error and fail-error rates for this scenario, additional notation is required. To ease notation, only two microbes are considered, although the equations can be extended for application to any number of microbes. To fit the case study presented later in this paper, let the two microbes be *P. aeruginosa* and *S. aureus*. Consider a PS that requires K_{Pa} tests against *P. aeruginosa* and K_{Sa} tests against *S. aureus* at each of L laboratories, for a total of $K = K_{Pa} + K_{Sa}$ tests at each of L laboratories. The hypotheses in Equation 1 must now be extended to account for different population means ($\mu_{LR,Pa}$ and $\mu_{LR,Sa}$) and different target LRs ($LR_{\text{target},Pa}$ and $LR_{\text{target},Sa}$) for *P. aeruginosa* and *S. aureus*, respectively, leading to the hypotheses

$$H_0: \{\mu_{LR,Pa} \leq LR_{\text{target},Pa} \text{ or } \mu_{LR,Sa} \leq LR_{\text{target},Sa}\};$$

$$H_a: \{\mu_{LR,Pa} > LR_{\text{target},Pa} \text{ and } \mu_{LR,Sa} > LR_{\text{target},Sa}\} \quad (18)$$

The pass-error rate α_{KL} for a PS requiring that a disinfectant pass all $K = K_{Pa} + K_{Sa}$ tests at each of L laboratories is calculated by

$$\alpha_{KL} = Pr(\text{Rejecting } H_0 \text{ in all } K \cdot L \text{ tests} \mid \mu_{LR,Pa} = LR_{\text{target},Pa}$$

$$\text{and } \mu_{LR,Sa} = LR_{\text{target},Sa})$$

$$= Pr(T \geq t_{1,Pa} \text{ in all } K_{Pa} \cdot L \text{ tests and } T \geq t_{1,Sa} \text{ in all } K_{Sa} \cdot L \text{ tests} \mid$$

$$\mu_{LR,Pa} = LR_{\text{target},Pa} \text{ and } \mu_{LR,Sa} = LR_{\text{target},Sa}) \quad (19)$$

In Equation 19, the notation $t_{1,Pa}$ and $t_{1,Sa}$ indicate values calculated via Equation 9, with $LR_{\text{PS},Pa}$ and $LR_{\text{PS},Sa}$ specified for each microbe separately.

To calculate the probability that a highly effective product fails to pass all $K \cdot L$ tests, suppose that a highly effective treatment has the following characteristics: $\mu_{LR,Pa} = LR_{\text{high},Pa}$ and $\mu_{LR,Sa} = LR_{\text{high},Sa}$. Now the fail-error rate is

$$\beta_{KL} = Pr(\text{Failing to reject } H_0 \text{ in at least one test} \mid$$

$$\mu_{LR,Pa} = LR_{\text{high},Pa} \text{ and } \mu_{LR,Sa} = LR_{\text{high},Sa})$$

$$= 1 - Pr(\text{Rejecting } H_0 \text{ in all } K \cdot L \text{ tests} \mid$$

$$\mu_{LR,Pa} = LR_{\text{high},Pa} \text{ and } \mu_{LR,Sa} = LR_{\text{high},Sa})$$

$$= 1 - Pr(T \geq t_{1,Pa} \text{ in all } K_{Pa} \cdot L \text{ tests and } T \geq t_{1,Sa} \text{ in all } K_{Sa} \cdot L \text{ tests} \mid$$

$$\mu_{LR,Pa} = LR_{\text{high},Pa} \text{ and } \mu_{LR,Sa} = LR_{\text{high},Sa}) \quad (20)$$

Consider the special case where the *P. aeruginosa* test results are statistically independent of the *S. aureus* test results within each laboratory. Let $\alpha_{K_{Pa}}$ and $\alpha_{K_{Sa}}$ denote the pass-error rate for K_{Pa} *P. aeruginosa* tests and K_{Sa} *S. aureus* tests, respectively, conducted in each of L laboratories, both calculated using Equation 13. Similarly, let $\beta_{K_{Pa}}$ and $\beta_{K_{Sa}}$ denote the fail-error rates for the *P. aeruginosa* tests and *S. aureus* tests, respectively, both calculated by Equation 14. Then the pass-error rate in Equation 19 can be simplified to:

and the fail-error rate in Equation 20 can be simplified to:

)

It is highly plausible that the preceding independent microbe

scenario is not appropriate because the *P. aeruginosa* test results are correlated with the *S. aureus* test results in the same laboratory. To incorporate the between-microbe correlation into the probability model, a multivariate *t*-distribution (25) can be used. Inputs for the calculations using the multivariate *t* include the values for t_1 and λ_1 , calculated for each microbe separately. Given these inputs, as well as the block diagonal matrix of correlations among the tests, statistical software is readily available to perform the calculations (Appendix B).

An Approach to Choosing LR_{target} , LR_{PS} , and LR_{high} for Semiquantitative Methods

The responsibility for choosing the numerical values of LR_{target} , LR_{PS} , and LR_{high} for any disinfectant test method lies with the stakeholder, such as a regulatory agency. For standard semiquantitative methods, such as the UDM and HSCT, disinfectant efficacy has historically been reported as the number of positive carriers, not as an LR value. These standard methods have traditionally specified a PS for a single test by specifying the maximum allowable number of positive carriers N_{PS} in order to pass the test. The rejection region for this PS is

$$RR_1 = \{N \leq N_{PS}\} \quad (21)$$

To help stakeholders of semiquantitative methods transition to the LR-based evaluations described in this paper, we show how the LR framework that we have presented is related to N , the number of positive carriers observed in a semiquantitative test.

First, a stakeholder may wish to specify the value for LR_{target} in Equation 1, for example, by choosing a target number of positive carriers, N_{target} . To calculate an LR using Equation 3, a value for $TestLD$, τ , must also be specified. For these choices, the hypotheses in Equation 1 can be rewritten in a mathematically equivalent form for a semiquantitative test as

$$\begin{aligned} H_0: \mu_{LR} &\leq LR(N_{target}; \tau); \\ H_a: \mu_{LR} &> LR(N_{target}; \tau) \end{aligned} \quad (22)$$

Next, to evaluate these hypotheses, a stakeholder may specify the PS in Equation 6 by choosing the maximum number of positive carriers, N_{PS} , allowed in order to pass a single test. Using Equation 3 with $TestLD = \tau$, the corresponding rejection region is:

$$\begin{aligned} RR_1 &= \{LR(N; TestLD) \geq LR(N_{PS}; \tau)\} \\ &= \frac{\dots - L}{\dots - L} \\ &= \dots \\ &= \{T \geq t_1\}. \end{aligned} \quad (23)$$

Equation 23 makes clear how to calculate t_1 , and now straightforward calculations yield the Type I pass-error rate for a PS that requires a disinfectant to either pass a single test (cf. Equation 10); pass all of multiple tests of a single microbe (cf. Equation 13), or pass all of multiple tests of multiple microbes (cf. Equation 19).

The stakeholder may also define a highly effective product as one for which the mean LR is $LR_{high} = LR(N_{high}; \tau)$ for a small number of positive carriers, N_{high} . Given this specification, the Type II fail-error rate calculations via Equations 11, 14, or 20

are straightforward because the non-centrality parameter is (cf. Equation 12)

$$\lambda_1 = \frac{\dots - L}{\dots - L}. \quad (24)$$

Equations 23 and 24 show that t_1 and λ_1 depend on a difference of LRs, both evaluated at τ . Thus, the error rate calculations when using Equation 23 to evaluate Equation 22 depend on neither the magnitude of the fixed value of $TestLD = \tau$, nor on the magnitude of the LR values. Instead, the error rates depend on the variability of the $TestLDs$ and the LRs via S_R .

When evaluating the hypotheses in Equation 22, it is useful to relate the error rates for the LR-based PS in Equation 23 to the error rates when using the rejection region $RR_1 = \{N \leq N_{PS}\}$ in Equation 21, traditionally used for semiquantitative methods. After some algebra, the LR-based rejection region in Equation 23 can be expressed equivalently as:

$$RR_1 = \left\{ N \leq \text{floor} \left[\left(n_{tr} + \frac{1}{2} \right) - (n_{tr} + 1) \left(\frac{(n_{tr} + \frac{1}{2}) - N_{PS}}{(n_{tr} + 1)} \right)^{10^{TestLD - \tau}} \right] \right\} \quad (25)$$

where $\text{floor}[\cdot]$ is the floor function, also called the integer part function. Because N is a random variable with support on the integers in the interval $[0, n_{treated}]$, then as long as $0 \leq TestLD - \tau < \epsilon$, for a small positive deviation ϵ that can be calculated for a specific PS (as in the example below), then the rejection region $RR_1 = \{N \leq N_{PS}\}$ in Equation 21 is equivalent to the region given in Equation 25. Thus, in order to evaluate the PS in Equation 21, consider specifying LR_{target} , LR_{PS} , and LR_{high} with respect to $\tau = TestLD_{min}$, where $TestLD_{min}$ is the lower bound for $TestLD$. Using this choice in Equation 25 shows that when the realized $TestLD$ in each test is close to its lower bound $TestLD_{min}$, the error rates associated with using Equation 23 to evaluate Equation 22 are the same as when using $RR_1 = \{N \leq N_{PS}\}$ in Equation 21 to evaluate Equation 22. For larger realized values of $TestLD$, the rejection region in Equation 25 is larger than and includes the rejection region in Equation 21. That is, $RR_1 = \{N \leq N_{PS}\}$ is more stringent than Equation 23 for large values of $TestLD$. Thus, the pass-error rate for Equation 23 is an upper bound for the pass-error rate when $RR_1 = \{N \leq N_{PS}\}$ is used. Additionally, the fail-error rate when using Equation 23 is a lower bound for the fail-error rate associated with using Equation 21.

A Case Study: Reassessing the UDM Performance Standard

The UDM is accepted by the U.S. Environmental Protection Agency (EPA) for the purpose of product registration. Stainless steel cylinder carriers are used to simulate an environmental surface. Testing of *S. aureus* and *P. aeruginosa* is required to support a hospital disinfectant claim. For the UDM, six untreated control carriers and 60 treated carriers are evaluated per test. Product efficacy is determined by the presence or absence (based on turbidity of the subculture medium) of viable microbes on each of the 60 inoculated carriers following exposure to the product. A positive carrier is a carrier associated with the presence of viable microbes. To meet the current PS for the UDM in a single test, an acceptable outcome, or pass, occurs

Table 1. Summary statistics for the LR for each of the two microbes in the UDM collaborative study (ref. 12) for the current method and high treatment

Microbe	Mean LR	Mean <i>TestLD</i>	Median <i>N</i>	Among-laboratory variance (S^2_{lab})	Within-laboratory variance (S^2_{test})	Intralaboratory correlation (<i>r</i>)	Reproducibility SD (S_R)
<i>P. aeruginosa</i>	8.26	6.71	1	0.175	0.111	0.6119	0.5348
<i>S. aureus</i>	8.48	6.64	0	0.000	0.100	0.0000	0.3162

if no more than $N_{PS} = 1$ positive carrier is observed out of the 60 carriers treated with the product. The UDM is considered a semiquantitative test, as the control carriers are used to assess microbial loading via *TestLD*. For a valid test, the *TestLD* must fall between 6 and 7 logs (26). The current PS required by EPA for the registration of liquid disinfectants specifies that three separate tests (one test for each of three production lots) must be conducted using *S. aureus* and *P. aeruginosa* (i.e., $K_{Pa} \cdot L = K_{Sa} \cdot L = 3$) for a total of six tests [i.e., $K \cdot L = (K_{Pa} + K_{Sa}) \cdot L = 6$], all of which may be conducted in $L = 1$ laboratory (4, 5). The disinfectant must pass all six tests in order to be considered efficacious.

For the purpose of this example, assume that the target number of positive carriers is $N_{target} = 2$. In this case, the hypotheses in Equations 1 and 22 can be rewritten as:

$$\begin{aligned} H_0: \mu_{LR} &\leq LR(N_{target} = 2; \tau = 6) = 7.38; \\ H_a: \mu_{LR} &> LR(N_{target} = 2; \tau = 6) = 7.38 \end{aligned} \quad (26)$$

Note that $\tau = 6$ is the lower bound for the *TestLDs* for a valid UDM test (26). For this example, a disinfectant will pass a single test if the realized LR is larger than $LR(N_{PS} = 1; \tau = 6) = 7.60$. Furthermore, we assume that a highly effective treatment has $\mu_{LR} \geq LR_{high} = LR(N_{high} = 0; \tau = 6) = 8.08$.

The error rate calculations for this example are based on a recent multilaboratory study of the UDM (12) in which $J = 3$ UDM tests, against *P. aeruginosa* and *S. aureus*, were conducted at each of $I = 5$ laboratories. It is informative to revisit the calculations in Steps 1–6 from the perspective of testing the hypotheses in Equation 26 for the UDM using the UDM collaborative study data (12). The R code used to implement the following six steps is provided in Appendix B.

Step 1 example: Calculate the LR for each UDM test.—For each of the two microbes, there were a total of $I \cdot J = 15$ tests in the UDM collaborative study (12). Using Equations 2 and 3, an LR was calculated for each test.

Step 2 example: Estimate the among-laboratory variability of the UDM LRs.—For *P. aeruginosa* and *S. aureus* separately, the LRs from the UDM collaborative study (12) were analyzed to find the variance components; the results are provided in Table 1. In the UDM collaborative study, $n_{control} = 3$ control carriers were used in each UDM test, not six as prescribed by the current UDM standard (3–5). Thus, we used $n_{control} = 3$ in all calculations to be consistent with the results observed in the study. The use of three control carriers provides a conservative (i.e., larger) measure of the *TestLD* and LR variability, and was used in the final calculations for estimating the error rates. Using Equation 5, the degrees of freedom $df_R = 6.9$ and 13.8 were calculated for *P. aeruginosa* and *S. aureus*, respectively.

Step 3 example: Performance standard for a single UDM test.—To evaluate the hypotheses in Equation 26, the rejection region (cf. Equations 7 and 23 for $N_{PS} = 1$ is

$$RR_1 = \{LR(N; TestLD) \geq LR(N_{PS} = 1; \tau = 6) = 7.60\} \quad (27)$$

By Equation 25, this single test PS is identical to the current, historical UDM PS, $RR_1 = \{N \leq 1\}$, in Equation 21 as long as $6 \leq TestLD < 6.23$; i.e., if the realized *TestLD* for the test is within $\varepsilon = 0.23$ logs of the minimum of 6. Otherwise, if the realized *TestLD* for the test is larger than 6.23, then $RR_1 = \{N \leq 1\}$ is a more stringent PS than Equation 27, in which case the example calculations in Steps 4–6 below yield upper bounds for the pass-error rates for the current UDM PS, and lower bounds of the fail-error rates.

Step 4 example: Calculating the error rates for a single UDM test.—Using Table 1, for *P. aeruginosa*, the critical value in Equation 27 (cf. Equation 9 is:

$$\begin{aligned} t_1 &= [LR(1; \tau = 6) - LR(2; \tau = 6)]/S_R \\ &= 0.2253/0.5348 \\ &= 0.4217 \end{aligned} \quad (28)$$

Now application of Equation 10 with $df_R = 6.9$ shows that the pass-error rate for a single UDM test using *P. aeruginosa* is $\alpha_1 = 0.343$. For the PS in Equation 27 with $N_{PS} = 1$, the regulatory authority is $100(1 - \alpha_1)\% = 66\%$ confident that the disinfectant is effective (i.e., that $\mu_{LR} > 7.38$).

To calculate the fail-error rate β_1 , Equation 12 shows that the non-centrality parameter is

$$\begin{aligned} \lambda_1 &= [LR(0; \tau = 6) - LR(2; \tau = 6)]/S_R \\ &= 0.7063/0.5348 \\ &= 1.32 \end{aligned} \quad (29)$$

and now by Equation 11 the fail-error rate for a single test is $\beta_1 = 0.182$. Consequently, 82% of UDM tests against *P. aeruginosa* will correctly pass highly effective disinfectant products. Similar calculations for *S. aureus* yield $\alpha_1 = 0.244$ and $\beta_1 = 0.064$ based on t -distributions with $df_R = 13.8$ and $\lambda_1 = 2.23$. These results for a single UDM test PS are summarized in Table 2.

Step 5 example: Error rates for a UDM performance standard that requires that a disinfectant passes all tests for a single microbe.—For product registration, a product must pass all three UDM tests, with $N_{PS} = 1$ allowable positive carrier per test, for both *P. aeruginosa* and *S. aureus*. Similar to the single test case above, the following example calculations utilize $LR_{target} = LR(2; 6) = 7.38$, $LR_{PS} = LR(1; 6) = 7.60$, and $LR_{high} = LR(0; 6) = 8.08$.

In the UDM collaborative study, the among-laboratory variance for *S. aureus* was not statistically significant and was estimated to be zero (Table 1). Thus, *S. aureus* tests conducted within the same laboratory appear to be uncorrelated (cf. Equation 17), so Equation 15 yields the pass-error rate of ineffective disinfectants that are required to pass all three *S. aureus* tests, $\alpha_{KL} = (0.2438)^3 = 0.015$. The fail-error rate for a highly effective product applied to three independent

Table 2. Estimated error rates, expressed as percentages, for the current UDM PS^a

Microbe	Pass-all of this number of tests	Pass-error rate, α (for $N_{\text{target}} = 2$), %	Fail-error rate, β (for $N_{\text{high}} = 0$), %
Pa	1 test	34.30	18.19
Sa	1 test	24.38	6.44
Pa	3 tests in 1 laboratory	15.50	33.45
Sa	3 tests in 1 laboratory	1.59	18.01
Pa	1 test in each of 3 laboratories	4.26	44.92
Sa	1 test in each of 3 laboratories	1.59	18.01

^a UDM PS allows up to $N_{\text{PS}} = 1$ positive carrier out of a set of 60 inoculated and treated carriers for each microbe separately; “three tests in one lab” is the current UDM registration PS; one test is a potential test scenario for a post-registration PS.

S. aureus UDM tests, calculated via Equation 16, is $\beta_{KL} = 1 - (1 - 0.0644)^3 = 0.181$. These values are very close to those presented in Table 2, for which the calculations were done via Equations 13 and 14 using the multivariate t distribution.

Similar calculations assuming independent tests for *P. aeruginosa* would provide incorrect error rates because the UDM collaborative study found a statistically significantly large among-laboratory variance (Table 1), and consequently, a correlation of $r = 0.61$ for *P. aeruginosa* tests conducted in the same laboratory. Thus, the error rates for *P. aeruginosa* must be found via Equations 13 and 14 using a multivariate t probability calculation with $t_1 = 0.4217$ from Equation 28 and non-centrality parameter $\lambda_1 = 1.32$ from Equation 29. The results of these calculations, considering both single-laboratory and multilaboratory testing scenarios, are provided in Table 2.

Step 6 example: Error rates for a UDM performance standard that requires a disinfectant to pass all tests for each of multiple microbes.—Currently, EPA’s registration program requires three separate tests per microbe; fewer tests may be used for monitoring product performance post-registration. In either case, the PS requires that all tests for all microbes pass in order for the disinfectant to be deemed efficacious. Table 3 provides the UDM PS error rates for both the registration and post-registration single test scenario for *P. aeruginosa* and *S. aureus* (calculated via Equations 19 and 20). In the former case, the error rates were calculated two ways: first with all tests conducted in the same laboratory; and second, with one test conducted in each of three laboratories. The calculation of the error rates in Table 3 requires an estimate of the correlation between microbes. Unfortunately, it was not possible to reliably estimate this correlation from the UDM collaborative study. Thus, the error rates were calculated for a wide range of hypothetical correlation values: 0.00, 0.25, 0.5, and 0.95. In two instances, when the between-microbe correlation was either 0.5 or 0.95, the correlation matrix was no longer positive-definite, so the error rates were not calculable (reported as NC in Table 3). An intuitive explanation of this failure is that the observed variability of the two microbes is inconsistent with these high hypothetical correlations. In the calculations that generated Table 3, LRs corresponding to $\tau = 6$, $N_{\text{PS}} = 1$, $N_{\text{target}} = 2$, and $N_{\text{high}} = 0$ were used for both microbes.

Results and Discussion

The case study presented in this paper utilizes results from available multilaboratory data (12) to ensure that among-laboratory variance is appropriately addressed when the UDM PS is assessed. When the *TestLD* in each test is close to six, estimates of the error rates for the single-microbe UDM PS currently used in EPA’s registration process are presented in Table 2. These rates can be compared directly to the UDM and HSCT PS error rates suggested by other authors (8–10). The statistical justification for the current UDM PS was published in 1961 by Ortenzio and Stuart (8). They proposed the current PS that allows no more than one positive carrier out of a 60-carrier set in each test, and claimed that this PS allowed a pass-error rate of 0.05 when challenged against *S. aureus*. The results in Table 2, however, suggest that for a PS based only on *S. aureus*, it is crucial to require multiple tests in order to maintain a pass-error rate less than 0.05. Due to the significant among-laboratory variance, a PS based only on *P. aeruginosa* tests would require testing at multiple laboratories in order to maintain a pass-error rate less than 0.05 (Table 2). The results in Table 3 underscore the importance of the multi-microbe requirement in the current UDM PS, which has an estimated pass-error rate between 0.003 and 0.01. From the point of view of manufacturers of highly effective treatments, it is unfortunate that this small pass-error rate is accompanied by fail-error rates as large as 0.451.

It may not be practical or cost-effective to require a disinfectant manufacturer to perform multiple tests at multiple laboratories to meet registration requirements for a disinfectant product. A multilaboratory requirement would presume that appropriate laboratories with expertise in conducting the procedure are available. Nonetheless, a multilaboratory study is desirable to provide a high level of confidence in the product’s performance from laboratory to laboratory.

We have shown how to calculate the error rates for PSs based solely on numbers of positive carriers, traditionally used for semiquantitative methods, when *TestLD* is close to the lower bound for the controls for the method. Quantification of “close” depends on the number of inoculated and treated carriers used (n_{treated}) in the method (Equation 25). When the *TestLD* is much larger than the lower bound, the error rates found via the techniques in this paper are bounds on the error rates for semiquantitative method PSs based solely on numbers of positive carriers. For example, for the current registration and possible post-registration (i.e., one-test scenario) PSs for the semiquantitative UDM, Tables 2 and 3 provide point estimates for the corresponding pass-error and fail-error rates as long as the *TestLD* in each test is within 0.23 of the minimum of $\text{TestLD}_{\text{min}} = 6$. Those experienced with the UDM might anticipate that for most tests, the realized *TestLD* will be between 6 and 6.23. For tests with larger *TestLD*s, the pass-error rates in Tables 2 and 3 are upper bounds and the fail-error rates are lower bounds.

For semiquantitative methods, a practical feature of the PS assessment presented here is transforming the number of positive carriers into a more conventional measure of disinfectant efficacy, LR, thereby properly taking account of the mean microbial loading on the carriers. The advantages of using the LR instead of analyzing the numbers of positive carriers some other way [e.g., for the UDM, others have used a probit analysis (8) or logistic regression (10)] are (1) the resulting value is directly interpretable as the LR of microbes

Table 3. Estimated error rates, expressed as percentages, for the current UDM PS^a

Correlation between microbes	No. of laboratories	Number <i>P.a.</i> tests at each laboratory	Number <i>S.a.</i> tests at each laboratory	Pass-error rate, α (for $N_{\text{target}} = 2$), %	Fail-error rate, β (for $N_{\text{high}} = 0$), %
0	1	1	1	8.9	23.3
0.25	1	1	1	11.9	22.4
0.5	1	1	1	15.2	21.1
0.95	1	1	1	23.9	18.2
0	1	3	3	0.3	45.1
0.25	1	3	3	1.0	41.6
0.5	1	3	3	NC	NC
0.95	1	3	3	NC	NC
0	3	1	1	0.1	54.4
0.25	3	1	1	0.2	52.7
0.5	3	1	1	0.4	50.5
0.95	3	1	1	1.6	45.1

^a UDM PS allows up to $N_{\text{PS}} = 1$ positive carrier out of a set of 60 inoculated and treated carriers that require a disinfectant to pass all of multiple tests for both microbes. Scenarios are presented for one and three tests per microbe, and one and three laboratories.

over the entire 60-carrier set; (2) the results of the variability of the method can be directly compared to other studies of disinfectant tests, which are usually analyzed with respect to the LR; (3) analyses, such as ANOVA and *t*-tests, appropriate for normally distributed LRs, can be directly applied; and (4) the normality and homogeneity of variance assumptions are easily verified with well-known diagnostic tools (7).

Thus, when applying the approach outlined in this paper, it is necessary to verify the model's assumptions. Experience with both quantitative and semiquantitative methods indicates that LRs satisfy the assumptions of normality and homogeneity of variance across a variety of different test microbes and standard methods (11, 12, 20, 27–31). To verify the validity of the model's assumptions and accuracy of the model's predictions, a numerical simulation study was conducted in which LRs were simulated by generating CFU data for each carrier via a Poisson process, with random effects due to laboratory, tests within each laboratory, and carriers within each test. For the case study of the semiquantitative UDM, one Poisson process was utilized for the three control carriers, and a separate Poisson process was utilized for the 60 treated carriers so that the Poisson rates and the variances for the random effects for the control and treated carriers could be set independently and consistently with the statistics reported in Table 1. Each simulated UDM treated carrier was next recorded as either positive or negative, with a positive outcome for a carrier whenever a non-zero CFU was simulated for that carrier. From these simulated control CFU and treated positive carrier data, an LR for each UDM test was calculated via Equation 3. A range of PSs was considered requiring that a disinfectant pass all $K=1$ to 3 tests at each of $L=1$ to 3 laboratories. For each PS, results from 10^4 simulated trials corroborated the predicted error rates reported in Tables 2 and 3.

When simulating the LRs for a highly effective product with $\mu_{\text{LR}} = \text{LR}_{\text{high}} = \text{LR}(0; \tau = 6) = 8.08$, due to the high frequency of tests with zero positive treated carriers, the reproducibility SD was about the same as the reproducibility SD for the control *TestLD*s. This is because, according to Equation 3, the variance of the LRs can be decomposed into contributions from the *TestLD* and the mean of the treated carrier LDs

(mean $\text{LD}_{\text{treated}}$) minus two times the covariance (32). For example, in the UDM collaborative study (12), because the *P. aeruginosa* control and treated data appear to be uncorrelated (32), the variance of the LRs for *P. aeruginosa* is simply the sum of the variance of the *TestLD* and the variance of the treated carriers. That is, $\text{Var}(\text{LR}) = 0.53^2 = \text{Var}(\text{TestLD}) + \text{Var}(\text{mean}(\text{LD}_{\text{treated}})) = 0.38^2 + 0.37^2$, because the reproducibility SD for the *TestLD* was 0.38 and the LR reproducibility SD 0.53 (Table 1). However, for a highly effective treatment with $\mu_{\text{LR}} = \text{LR}_{\text{high}} = 8.08$, there will be a large frequency of zeros which forces $\text{Var}(\text{mean}(\text{LD}_{\text{treated}})) \approx 0$. That is, the variance of the LRs for a highly effective product is equal to the variance for the control *TestLD*s. Since the fail-error rate calculations for the UDM are based on the larger reproducibility SDs observed in the collaborative study (Table 1), the fail-error rates in Tables 2 and 3 are larger than the fail-error rates for a highly effective treatment that exhibits LRs with smaller reproducibility SDs. For a quantitative method, a similar decrease in the variability of the LRs occurs when many treated carriers in a single test exhibit zero CFUs at all plated dilutions, with the zeros at the smallest plated dilution substituted by a small value (e.g., 0.5 or 1).

While the pass-all approach has been the focus in this paper, an alternative approach was to consider a PS on the average LR across multiple tests (Appendix A). For *P. aeruginosa*, we found that for a UDM PS that requires the mean LR across three tests to be greater than $\text{LR}(N_{\text{PS}} = 1; \text{TestLD} = 6)$, while efficacy is defined by the target $\text{LR}(N_{\text{PS}} = 2; \text{TestLD} = 6)$, and a highly efficacious disinfectant has a mean LR of $\text{LR}(N_{\text{high}} = 0; \text{TestLD} = 6)$, then $\alpha = 0.33$ and $\beta = 0.145$ when all three tests are conducted in the same laboratory, and $\alpha = 0.24$ and $\beta = 0.060$ when one test is conducted in each of three laboratories. Comparing these values to the results in Table 2 suggests that given the same number of tests, a pass-all PS for the UDM has smaller pass-error rates and larger fail-error rates than an averaging PS does. This illustrates that, in general, both the pass-all and averaging scenarios should be considered when devising PSs, with the choice of PS depending on the balance between the error rates and the required experimental effort.

When LRs are symmetrically distributed, the mean LR, μ_{LR} , is also the 50th percentile. Instead of the 50th percentile (in the hypotheses in Equation 1), it is conceivable that a stakeholder would rather choose to define efficacy with respect to a different LR percentile (Appendix A). To investigate such a scenario, an alternative definition for efficacy based on the 25th percentile was considered. Such a definition implies that 75% of the LRs from individual tests are at least as large as the target, LR_{target} . For the same number of K tests at each of L laboratories, and fixed values of LR_{PS} and LR_{target} , when the parameter of interest is the 25th LR percentile, the pass-error rate is larger and the fail-error rate smaller than when the parameter is the 50th percentile. For example, whereas the error rates for a single *P. aeruginosa* UDM test PS were $\alpha_1 = 0.343$ and $\beta_1 = 0.182$ (Table 2) when evaluating the hypotheses in Equation 26 for the 50th percentile using the rejection region in Equation 27, the error rates were $\alpha_1 = 0.605$ and $\beta_1 = 0.057$ when efficacy was defined by the 25th LR percentile. For a PS that requires passing one test in each of three laboratories, $\alpha_{KL} = 0.043$ and $\beta_{KL} = 0.449$ when evaluating the hypotheses with respect to the 50th percentile (Table 2), while the error rates were $\alpha_{KL} = 0.226$ and $\beta_{KL} = 0.161$ when efficacy was defined by the 25th LR percentile. This illustrates that a more stringent PS is required in order to maintain small pass-error rates when the parameter of interest is the p^{th} percentile with $p < 50$.

For LR-based PSs for either quantitative or semiquantitative methods, the framework in this paper provides point estimates for the error rates. One could also provide, via computer simulations, a margin of error and confidence interval for each error rate. Such calculations are beyond the scope of this paper.

Conclusions

The appropriate use of a laboratory test method to assess the efficacy of disinfectants is critical to the overall goal of protecting public health. Collaborative studies are used to understand how test methods perform (e.g., repeatability and reproducibility), and the resulting data, as demonstrated in our case study, can be used to develop test criteria (e.g., number of tests and number of laboratories) and a set of associated PSs. Ultimately, the end-users of the test method, such as regulatory agencies, should have a high level of confidence in the test outcomes. This paper provides a statistical framework, including the computer code, to analyze such data to determine an appropriate PS specified in terms of LRs for both quantitative and semiquantitative disinfectant test methods. As others have proposed (1, 8–10), we believe that any PS used by regulatory agencies should be based on a quantitative assessment of the pass-error and fail-error rates. The model we have presented can be used to improve the decision-making process (i.e., regulatory decisions based on PSs) which rely on the outcome of the test method.

The statistical approach outlined in this paper is flexible, including the ability to assess PSs for different microbes within the same test method. This flexibility allows for inherent variability which exists among microbial species. Thus, those interested in assessing and standardizing product performance can set different targets (cf. Equation 18); define different acceptable outcomes (cf. Equation 19); set different expectations for highly efficacious treatments (cf. Equation 20); and investigate how these changes affect the error rate.

Acknowledgments

This work was performed under contract EP-W-08-014 between Montana State University and EPA OPP Microbiology Laboratory Branch. We gratefully acknowledge substantial contributions from Darla Goeres at the Center for Biofilm Engineering at Montana State University in the preparation of this manuscript.

References

- (1) Trout, J. (1985) *J. AOAC Int.* **68**, 763–765
- (2) *Official Methods of Analysis* (2008) AOAC INTERNATIONAL, Gaithersburg, MD, Method **2008.05**
- (3) *Official Methods of Analysis* (2012) 18th Ed., AOAC INTERNATIONAL, Gaithersburg, MD, Method **955.14**
- (4) *Official Methods of Analysis* (2012) 18th Ed., AOAC INTERNATIONAL, Gaithersburg, MD, Method **955.15**
- (5) *Official Methods of Analysis* (2012) 18th Ed., AOAC INTERNATIONAL, Gaithersburg, MD, Method **964.02**
- (6) *Official Methods of Analysis* (2009) 18th Ed., AOAC INTERNATIONAL, Gaithersburg, MD, Method **991.47**
- (7) Hamilton, M., Hamilton, G., Goeres, D., & Parker, A. (2013) *J. AOAC Int.* **96**, 1138–1147. <http://dx.doi.org/10.5740/jaoacint.12-217>
- (8) Ortenzio, L.F., & Stuart, L.S. (1961) *J. AOAC Int.* **44**, 416–421
- (9) Rubino, J.R., Bauer, J.M., Clarke, P.H., Woodward, B.B., Porter, F.C., & Hilton, H.G. (1992) *J. AOAC Int.* **75**, 635–645
- (10) Arlea, C., King, S., Bennie, B., Kemp, K., Mertz, E., & Staub, R. (2008) *J. AOAC Int.* **91**, 152–158
- (11) Hamilton, M.A., DeVries, T.A., & Rubino, R.R. (1995) *J. AOAC Int.* **78**, 1102–1109
- (12) Tomasino, S.F., Parker, A.E., Hamilton, M.A., & Hamilton, G.C. (2012) *J. AOAC Int.* **95**, 1618–1628. <http://dx.doi.org/10.5740/jaoacint.12-170>
- (13) Casella, G., & Berger, R.L. (1990) *Statistical Inference*, Duxbury Press, Belmont, CA
- (14) Hamilton, M.A. (2011) *The P/N Formula for the Log Reduction When Using a Semiquantitative Disinfectant Test of Type SQ1, KSA-SM-08*, Feb 1, 2011, Center for Biofilm Engineering at Montana State University, Bozeman, MT. http://www.biofilm.montana.edu/resources/knowledge_sharing_articles
- (15) Cochran, W.G. (1950) *Biometrics* **6**, 105–116. <http://dx.doi.org/10.2307/3001491>
- (16) McCrady, M.H. (1915) *J. Infect. Dis.* **17**, 183–212. <http://dx.doi.org/10.1093/infdis/17.1.183>
- (17) Collett, D. (1991) *Modelling Binary Data*, Chapman & Hall, New York, NY. <http://dx.doi.org/10.1007/978-1-4899-4475-7>
- (18) Blodgett, R.J. (2006) *Bacteriological Analytical Manual*, FDA. <http://www.fda.gov/Food/ScienceResearch/LaboratoryMethods/ucm2006949.htm> (accessed January 2, 2014)
- (19) Morgan, B.J. (1992) *Analysis of Quantal Response Data*, Chapman & Hall, London, UK
- (20) Tomasino, S.F., & Hamilton, M.A. (2006) *J. AOAC Int.* **89**, 1373–1397
- (21) *Official Methods of Analysis* (2005) 18th Ed., AOAC INTERNATIONAL, Gaithersburg, MD, Appendix D
- (22) Mee, R. (1984) *Technometrics* **26**, 251–254. <http://dx.doi.org/10.1080/00401706.1984.10487962>
- (23) Neter, J., Kutner, M.H., Wasserman, W., & Nachtsheim, C.J. (1996) *Applied Linear Statistical Models*, McGraw-Hill Co., Boston, MA
- (24) Nijhuis, M.B., & Van den Heuvel, E.R. (2007) *J. Biopharm. Stat.* **17**, 123–142. <http://dx.doi.org/10.1080/10543400601001519>
- (25) Genz, A., & Bretz, F. (2009) *Computation of Multivariate*

- Normal and t Probabilities* Springer-Verlage, Heidelberg, Germany. <http://dx.doi.org/10.1007/978-3-642-01689-9>
- (26) Tomasino, S.F., Pines, R.M., & Hamilton, G.C. (2012) *J. AOAC Int.* **95**, 1059–1063. <http://dx.doi.org/10.5740/jaoacint.12-170>
- (27) Tomasino, S.F., Pines, R.M., Cottrill, M.P., & Hamilton, M.A. (2008) *J. AOAC Int.* **91**, 833–852
- (28) Tomasino, S.F., Pines, R.M., & Hamilton, M.A. (2009) *J. AOAC Int.* **92**, 1531–1540
- (29) Hamilton, M.A., & DeVries, T.A. (1996) *Biometrics* **52**, 1112–1120. <http://dx.doi.org/10.2307/2533073>
- (30) Buckingham-Meyer, K., Goeres, D., & Hamilton, M. (2007) *J. Microbiol. Methods* **70**, 236–244. <http://dx.doi.org/10.1016/j.mimet.2007.04.010>
- (31) *Interlaboratory Study for ASTM Method E2799-11 Standard Test Method for Testing Disinfectant Efficacy Against Pseudomonas aeruginosa Biofilm Using the MBEC Assay*, ASTM International, research report RR:E35-1006, 2011
- (32) Hamilton, M.A. (2012) *The Importance of Concurrent Control Carriers in Laboratory Tests of Surface Disinfectants, KSA-SM-12*, Center for Biofilm Engineering, Montana State University, Bozeman, MT. http://www.biofilm.montana.edu/resources/knowledge_sharing_articles
- (33) R Development Core Team (2010) R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org>
- (34) Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F., & Hothorn, T. (2011) *mvtnorm: Multivariate Normal and t Distributions*, R package version 0.9-96. <http://CRAN.R-project.org/package=mvtnorm>