



Use of statistical modeling to reassess the performance standard for the AOAC use-dilution methods (955.15 and 964.02)

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

NOTICE: This is a postprint of an article that originally appeared in Journal of AOAC International on June 2014. DOI: <http://dx.doi.org/10.5740/jaoacint.13-216>.

Tomasino S, Parker A, Hamilton M, "Use of statistical modeling to reassess the performance standard for the AOAC use-dilution methods (955.15 and 964.02)," Journal of AOAC International, January 2014 97(1): 68-77.

Made available through Montana State University's [ScholarWorks](http://scholarworks.montana.edu)
scholarworks.montana.edu

Use of Statistical Modeling to Reassess the Performance Standard for the AOAC Use-Dilution Methods (955.15 and 964.02)

STEPHEN F. TOMASINO

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

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

The AOAC Use-dilution methods (UDM) 955.15

(*Staphylococcus aureus*) and 964.02 (*Pseudomonas aeruginosa*) are laboratory methods used to substantiate antimicrobial efficacy claims for liquid disinfectants on inanimate surfaces. The UDM is accepted by the U.S. Environmental Protection Agency for the purpose of product registration.

To attain a hospital-level claim, testing against

S. aureus and *P. aeruginosa* is required, and the product must pass against both microbes. Currently, the UDM's performance standard for a single

60-carrier test is the same for both microbes, and allows up to one positive carrier for the product to be considered as a pass. In this paper, the performance standards for these methods are reassessed using data from a 2009 five-laboratory collaborative study and a recently published statistical model. The reassessment focuses on the pass-error rate for ineffective products and the fail-error rate for highly effective products. The calculations indicate that the pass-error rate is between 9 and 24% and the fail-error rate between 18 and 23% when the current performance standard is used for a single test.

For product registration, a smaller pass-error rate

(1%) historically has been maintained by requiring that a disinfectant pass three UDM tests for each of the two microbes; however, the calculations

also indicate that the fail-error rate is between 42 and 45%. This large fail-error rate is a compelling reason to consider a new performance standard

for the two UDM methods, 955.15 (*S. aureus*) and 964.02 (*P. aeruginosa*). One alternative performance standard allows no more than six positive carriers out of 60 as a pass when using *P. aeruginosa* and no more than three positive carriers out of 60 when using *S. aureus*. In addition, the new performance standard requires that three UDM tests be performed with each of the two microbes, and the disinfectant must pass all six tests to be considered efficacious. The statistical calculations for this alternative performance standard indicate that the pass-error rate is no more than 3%, while the fail-error rate is as small as 5%. Based on these error rate calculations, proposed revisions to the performance standards for AOAC Methods 955.15 and 964.02 are provided.

Introduction

The AOAC Use-dilution methods (UDMs) 955.14 for *Salmonella enterica* (1), 955.15 for *Staphylococcus aureus* (2), and 964.02 for *Pseudomonas aeruginosa* (3) are accepted by the U.S. Environmental Protection Agency (EPA) to support the registration of liquid antimicrobial products that claim to control microorganisms that pose a threat to human health on inanimate surfaces; these are known as public health products. Specifically, public health disinfectants, such as those for use in the healthcare environment (i.e., hospital disinfectants), must be effective against *S. aureus* and *P. aeruginosa*. The methods for *S. aureus* and *P. aeruginosa* are the focus of this paper.

The UDM requires the use of stainless steel cylinder carriers to simulate an environmental surface. The carriers are submerged in a test culture and dried. The inoculated carriers are divided into two groups: treated (60 carriers per test), and untreated (6 control carriers per test). The treated carriers are placed into 10 mL of disinfectant and allowed to stand for a defined contact time and temperature. Following the contact time, each carrier is removed aseptically, placed in a tube containing a liquid subculture medium for neutralization and recovery of viable bacteria, then incubated.

Product efficacy is determined qualitatively by assessing the presence (growth) or absence (no growth) of the test microbe in the subculture medium for each of the 60 treated carriers. The untreated (control) carriers are used to quantitatively measure microbial loading, and are expressed as the number of CFUs per carrier. Historically, the UDM has been considered a qualitative procedure; however, due to recent improvements (4) to the method that require assessing the bacterial population on inoculated carriers (control carrier counts) and associated minimum and maximum control carrier counts for a valid test, the current UDM may be considered a semiquantitative method that allows a log reduction (LR) measure of disinfectant efficacy to be calculated. For decades, the UDM's performance standard (PS) has been

defined by the maximum number of positive carriers out of 60 per test per microbe to achieve an acceptable outcome for a product. The PS requirement for each UDM test is currently 0–1 positive carriers out of 60 for each microbe. Thus, a product with two or more positive carriers in any single test is considered a failure

In practice, the results from one or more UDM tests are classified as a pass or fail outcome, where a pass means the product exhibited satisfactory disinfection efficacy and a fail means it did not. The current testing required by EPA for the registration of liquid disinfectants specifies the evaluation of three representative production lots (or batches) of the product tested once against each of the required microbes.

The statistical evaluation used in this paper focuses on the pass-error and fail-error rates associated with a PS for a method (5). The errors occur when application of the PS to the test results for a disinfectant product lead to the wrong conclusion. A pass-error occurs when the disinfectant being tested truly has unacceptably low efficacy, but is considered a pass by the PS. A fail-error occurs when the disinfectant truly is of acceptable efficacy, but is considered a failure. See the *Methodology* section below for technical details.

The statistical justification of the current UDM PS was published in 1961 by Ortenzio and Stuart (6). Their work relied on a probit analysis (7) of unpublished data from replicate product applications and a wide range of product use-dilutions. The statistical analysis that led to the current PS did not account for variability among laboratories; however, it is now believed that variability among laboratories is a critical aspect of method performance evaluations.

Relevant Laboratory Studies

Multilaboratory studies are critical to evaluation of method performance, and the resulting data may be used to support revisions designed to reduce variability in the method's results. For example, multilaboratory studies conducted by Cole et al. (8) and Rutala and Cole (9) demonstrated that UDM results may vary considerably among laboratories; they identified major aspects of the method that were likely sources of the variability. Unfortunately, control carrier counts were not evaluated or required in these early studies. An initiative designed to resolve the sources of variability led to the development and validation of AOAC Method 991.47, the hard surface carrier test (HSCT; 10), which is very similar to the UDM. It includes use of the same microbes, and specifies the standardization of test cultures, assessment of control carrier counts, and minimum and maximum control count levels. In 1992, Rubino et al. (11) published a statistical analysis of a 10-laboratory, six-disinfectant collaborative study of the HSCT, including a proposed PS. The HSCT is accepted by EPA to

support the registration of liquid disinfectants; however, the HSCT is rarely used by the regulated community.

More recently, a five-laboratory collaborative study was conducted in 2009 by the EPA and an industry-based technical work group to assess a set of proposed modifications to the UDM (12). This collaborative effort was not designed specifically to address the reassessment of the UDM PS. In the study, a single batch of an EPA-registered quaternary ammonium liquid disinfection product was tested at two distinct dilutions for each microbe, *S. aureus* and *P. aeruginosa*. EPA's Standard Operating Procedure (SOP) for conducting the UDM, SOP MB-05 (13), based on a strict interpretation of the published UDM, was followed by the participating laboratories. The collaborative study results indicated that the LRs for the current UDM for *S. aureus* varied relatively little within and among laboratories. Unlike the results for the HSCT (11, 14), the *P. aeruginosa* test LRs were more variable among laboratories than were those for *S. aureus*. For both *P. aeruginosa* and *S. aureus*, the LR variability exhibited by the current UDM was no larger than that observed for other accepted disinfectant test methods (15). Based on the data and the use of the same SOP across all five laboratories, the current UDM exhibited quite good repeatability and reproducibility (Table 1).

To improve the decision-making process (e.g., regulatory decisions) that relies on the outcome of the UDM, the EPA is interested in utilizing the best available statistical methodology to analyze the method's variability. Data from the 2009 collaborative study were determined to be very suitable for reassessing the PS, given its level of quality control, standardization, and replication. In conjunction with the collaborative data, a statistical model for assessing PS for quantitative and semiquantitative disinfectant test methods was recently published in the *J. AOAC Int.* (16). Thus, the initiative of this paper is directly supported by data published on the 2009 collaborative study (12) and a recently published report that describes the statistical model (16). The repeatability and reproducibility SDs of the LRs in the 2009 collaborative study (12) are used to form the basis for reassessing the UDM PSs (Table 1). Based on the statistical evaluation, proposed revisions to the PSs for AOAC Methods 955.15 and 964.02 are provided in this paper.

Methodology

The recently published model (16) provides a statistical approach for assessing a PS for any disinfectant test method for which an LR in a microbial population can be calculated. For the UDM, the combination of the number of positive carriers observed among a set of treated and inoculated carriers and the associated control carrier counts make it possible to calculate an LR for each test (16, 17). The PS assessment is based on a quantification of the pass-error rate for ineffective products and

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

Microbe	Mean LR	Mean <i>TestLD</i>	Median <i>N</i>	Among-lab variance (S^2_{lab})	Within-lab variance (S^2_{test})	Intra-lab 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

the fail-error rate for highly effective products. Pass-errors for the UDM were previously considered by Ortenzio and Stuart (6) and Rubino et al. (11), while fail-errors were of importance in the UDM assessment conducted by Arlea et al. (18). The process for calculating the error rates for a semiquantitative method, such as the UDM, is presented and explained in great detail in the statistical paper (16), and thus, is not repeated here. The published model provides examples and the computer code necessary to assess error rates for PSs that require passing only a single test, passing all of multiple tests, or passing multiple tests on average. The discussion below presents the relevant notation from the modeling paper (16).

For the UDM, the density of viable microbes on the untreated control carriers is measured as CFUs per carrier. The \log_{10} -transformed density is called the log density (LD). The mean of the LDs across the control carriers in a single test is called the *TestLD*:

$$TestLD = \frac{1}{n_{control}} \sum_{i=1}^{n_{control}} LD_{control, i} \quad (1)$$

where the LD for the i^{th} control carrier in the test has been denoted by $LD_{control, i}$ and the number of control carriers is given by $n_{control}$.

For a semiquantitative method like the UDM, due to the high efficacy treatments being tested, CFUs are not enumerated on the treated carriers. Rather, the number of positive carriers, denoted by N , out of 60 inoculated and treated carriers, is recorded for each test. Interestingly, the most probable number (MPN) of viable microbes per treated carrier is a simple function of N (7, 19–23):

$$most\ probable\ number = -\ln\left(\frac{60-N}{60}\right) \quad (2)$$

Unfortunately, Equation 2 is undefined when the number of positive carriers is either $N=0$ or $N=60$. It is necessary, therefore, to modify Equation 2 so that the viable cell density is calculable even for these extreme outcomes. By applying a modification that is commonly used for binomial proportions (24, 25), we calculate the viable cell density for treated carriers by:

$$MPN(N) = -\ln\left(\frac{60.5-N}{61}\right) \quad (3)$$

The LR for the UDM can now be calculated as a function of N and *TestLD* by:

$$LR(N; TestLD) = TestLD - \log_{10}[MPN(N)] \quad (4)$$

The LR disinfectant efficacy measure has also been used for other semiquantitative disinfectant test methods, such as the HSCT (14) and a sporicidal activity test (26).

When performing the UDM, we assume that the statistical parameter of interest for a disinfectant product is the true mean LR, μ_{LR} , over multiple tests (arbitrary LR percentiles are considered in Appendix A of ref. 16). The disinfectant product is considered effective if and only if $\mu_{LR} > LR_{target}$, and ineffective if and only if $\mu_{LR} \leq LR_{target}$. The target LR, $LR_{target} = LR(N_{target}; \tau)$ (cf. Equation 4), is a numerical value that can be set by the user (e.g., a regulatory agency) by choosing a target number of positive carriers, N_{target} , and a fixed value for *TestLD*, τ . For all the calculations presented later, N_{target} is varied for each PS considered, while $\tau = 6$, which is the minimum allowable *TestLD* for each UDM test. To calculate the pass-error and fail-error rates for a given UDM PS, one can use

conventional statistical hypothesis testing techniques where the null and alternative hypotheses are:

$$H_0: \mu_{LR} \leq LR(N_{target}; 6); H_a: \mu_{LR} > LR(N_{target}; 6) \quad (5)$$

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.

While a PS may require multiple tests, we consider PSs that require that the product passes all tests. One possible PS passes a disinfectant in each individual UDM test if the LR in each test is larger than some minimum value:

$$LR \geq LR(N_{PS}; 6) \quad (6)$$

where N_{PS} is a number of positive carriers set by the user. The traditional way to specify a PS for the UDM is to specify the maximum number of positive carriers allowed in each test:

$$N \leq N_{PS} \quad (7)$$

When the LRs are approximately normally distributed, it is straightforward to calculate the error rates for different UDM PSs by using Equation 6 to evaluate the hypotheses in Equation 5. We focus instead on the requirement in Equation 7 since it is the traditional way to specify a UDM PS. Thus, while all error rates presented in this paper were calculated using LRs via Equation 6, these results are always presented in terms of the PS based on numbers of positive carriers via Equation 7. The key relationship between the two PSs is that using the numbers of positive carriers to implement a PS (Equation 7) is more stringent (i.e., will fail products more often) than the requirement in Equation 6 based on LRs, with equivalence if the *TestLD* in each UDM test is close to 6 (cf. Equation 25 in ref. 16). Intuitively, this is because the former PS ignores the realized value of the *TestLD* in each UDM test. Hence, the pass-error rate for a PS based on numbers of positive carriers, will be smaller than the pass-error rates presented in this paper, and the fail-error rate for a PS that uses Equation 7 will be larger than the fail-error rates reported (16). The error rates are the same for both PSs when the realized *TestLD* is 6 in each test (16).

When using a PS to test the hypotheses in Equation 5, the Type I error probability, α , is the pass-error rate of a PS (27). In other words, the pass-error rate is the probability that an ineffective product with $\mu_{LR} = LR(N_{target}; 6)$ passes the PS. The Type II error probability, β , is the fail-error rate of a highly effective product, which is a product having $\mu_{LR} = LR(N_{high}; 6)$, where N_{high} is the number of positive carriers to be attained by a highly effective treatment specified by the technical authority.

In order to calculate the error rates, one must have information about the statistical uncertainty of the LRs. More specifically, for each microbe separately, it is necessary to have an estimate of the among-laboratory variance of the LRs (S_{lab}^2) and an estimate of the within-lab variance (S_{test}^2 , also known as the repeatability variance). An estimate of the reproducibility SD (S_R) can then be calculated (17, 28) by:

$$S_R = [S_{lab}^2 + S_{test}^2]^{1/2} \quad (8)$$

The estimates for S_{lab}^2 , S_{test}^2 and S_R that are used in all

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

Microbe	PASS ALL of this number of tests	PS required to PASS each test ($N_{PS} = 1$)	Max pass-error rate ($\alpha \leq$) (for $N_{target} = 2$), %	Min fail-error rate ($\beta \geq$) (for $N_{high} = 0$), %
<i>Pa</i>	1 test	$N \leq 1$	34.30	18.19
<i>Sa</i>	1 test	$N \leq 1$	24.38	6.44
<i>Pa</i>	3 tests in 1 lab	$N \leq 1$	15.50	33.45
<i>Sa</i>	3 tests in 1 lab	$N \leq 1$	1.59	18.01

^a The current UDM PS allows up to $N_{PS} = 1$ positive carrier out of a set of 60 inoculated and treated carriers for each microbe separately. Three tests in one laboratory is the current registration PS; one test is the current post-registration PS (ref. 16, Table 2).

subsequent calculations are from the 2009 UDM collaborative study (12). See Table 1.

As explained in detail elsewhere (Equations 6–12 in ref. 16), one may calculate the error rates for a PS that requires only a single test by calculating a univariate t test statistic ($T = (LR - LR_{target})/S_R$). For PSs that require multiple tests, calculations from a multivariate t distribution are used to calculate the error rates (details are provided in Equations 13–20 of ref. 16; computer codes for the calculations are provided in Appendix B of ref. 16).

While Table 1 provides all of the inputs required when calculating the error rates for a PS that requires tests against a single microbe, an additional input is required to evaluate a PS that requires tests of multiple microbes. This additional input is an estimate of the correlation among tests of the different microbes when conducted in the same laboratory. Unfortunately, it is not possible to reliably estimate this correlation from the 2009 UDM collaborative study (12). Thus, the error rates were calculated for a wide range of hypothetical correlation values: 0.00, 0.25, 0.5, and 0.95. In some instances, a between-microbe correlation of 0.5 or greater was not attainable given the observed variability of the two microbes (Table 1). In these cases, the error rates were not calculable (reported as NC in the tables).

In the collaborative study, three control carriers were used in each UDM test, not six as prescribed by the current UDM standard (1–3). The use of three control carriers provides a conservative (i.e., larger) measure of the *TestLD* and LR

variability, and thus was used in the final calculations for estimating the error rates.

Assessment of the Current UDM Performance Standard

The current UDM PS sets the maximum allowable number of positive carriers per test to $N_{PS} = 1$ in order for a disinfectant to be considered a pass. This requires that the observed LR for the disinfectant in each test must be larger than $LR(N_{PS} = 1; 6) = 7.60$ (cf. Equations 4 and 6) in order for the disinfectant product to be considered efficacious. Four PS testing regimens are assessed next: a single test PS for a single microbe; a PS that requires three tests of a single microbe; a PS that requires a single test of each of two microbes; and a PS that requires three tests of each of two microbes.

Tables 2 and 3 provide bounds on the error rates for these four PSs given two key assumptions. First, for the pass-error calculations, it is assumed that effective products have a true mean LR greater than $LR(N_{target} = N_{PS} + 1 = 2; 6) = 7.38$. In the context of Equations 5–7, each of these PSs require that $LR \geq 7.60$ in each test to provide evidence for the claim that $\mu_{LR} > 7.38$. We choose the target LR value to correspond to $N_{target} = N_{PS} + 1$ to ensure there is no “gray area” between the PS used to make decisions (N_{PS}) and the target (N_{target}), and thus provides a clear decision-making process regarding product efficacy. Second, the fail-error calculations further depend on the assumption that high efficacy products have true mean $LR(N_{high} = 0; 6) = 8.08$. The EPA is charged with using a PS

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

Correlation between microbes	No. of labs	No. of <i>Pa</i> tests at each lab	No. of <i>Sa</i> tests at each lab	PS required to PASS each test ($N_{PS} = 1$)	Max pass-error rate ($\alpha \leq$), %	Min fail-error rate ($\beta \geq$), %
0	1	1	1	$N \leq 1$	8.9	23.3
0.25	1	1	1	$N \leq 1$	11.9	22.4
0.5	1	1	1	$N \leq 1$	15.2	21.1
0.95	1	1	1	$N \leq 1$	23.9	18.2
0	1	3	3	$N \leq 1$	0.3	45.1
0.25	1	3	3	$N \leq 1$	1.0	41.6
0.5	1	3	3	$N \leq 1$	NC	NC
0.95	1	3	3	$N \leq 1$	NC	NC

^a The current UDM PS allows up to $N_{PS} = 1$ 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. The hypothetical correlation between the LRs of tests of the different microbes, and the numbers of tests for each microbe at each laboratory is varied (ref. 16, Table 3). NC = Not calculable.

Table 4. Estimated error rate bounds, expressed as percentages, for hypothetical PSs for *P. aeruginosa*^a

No. of labs	No. of tests at each lab	PASS ALL of this No. of tests	PS to PASS each test ($N \leq N_{PS}$)	Max pass-error rate ($\alpha \leq$) (for $N_{target} = N_{PS} + 1$), %	Min fail-error rate ($\beta \geq$) (for $N_{high} = 0$), %
1	1	1	$N \leq 1$	34.30	18.19
1	1	1	$N \leq 2$	39.38	9.23
1	1	1	$N \leq 3$	41.94	5.42
1	1	1	$N \leq 4$	43.49	3.47
1	1	1	$N \leq 5$	44.52	2.35
1	1	1	$N \leq 6$	45.26	1.67
1	3	3	$N \leq 1$	15.50	33.44
1	3	3	$N \leq 2$	19.20	18.88
1	3	3	$N \leq 3$	21.23	11.81
1	3	3	$N \leq 4$	22.49	7.89
1	3	3	$N \leq 5$	23.36	5.53
1	3	3	$N \leq 6$	23.97	4.02
2	1	2	$N \leq 1$	12.02	32.91
2	1	2	$N \leq 2$	15.64	17.56
2	1	2	$N \leq 3$	17.68	10.64
2	1	2	$N \leq 4$	18.98	6.80
2	1	2	$N \leq 5$	19.87	4.66
2	1	2	$N \leq 6$	20.53	3.30
2	2	4	$N \leq 1$	4.62	47.10
2	2	4	$N \leq 2$	6.58	27.43
2	2	4	$N \leq 3$	7.78	17.25
2	2	4	$N \leq 4$	8.57	11.52
2	2	4	$N \leq 5$	9.14	8.04
2	2	4	$N \leq 6$	9.55	5.81
3	1	3	$N \leq 1$	4.26	44.92
3	1	3	$N \leq 2$	6.24	25.12
3	1	3	$N \leq 3$	7.47	15.35
3	1	3	$N \leq 4$	8.29	10.03
3	1	3	$N \leq 5$	8.88	6.89
3	1	3	$N \leq 6$	9.31	4.91

^a A disinfectant must pass all multiple tests as the numbers of laboratories, numbers of tests at each laboratory, and allowable number of positive carriers (N) out of a 60 carrier set per test is varied. The number of laboratories is critical due to the large laboratory-to-laboratory variance for *P. aeruginosa*.

that ensures protection of the public's health by correctly failing ineffective products with a high probability ($1 - \alpha$). At the same time, EPA recognizes that the PS should fail highly effective products with a small probability (β) to avoid removing highly effective products from the marketplace. For the purpose of this exercise, the LR for a highly effective product is defined with respect to $N_{high} = 0$.

Tables 2 and 3 suggest high fail-error rates for the current UDM PSs. This confirms the high fail-error rate observed in the 2009 UDM collaborative study (12). The results also reveal that when only a single test per microbe is used to assess product efficacy, one should anticipate a high pass-error rate.

As is the case for scientific hypothesis testing, a PS should address both pass-errors and fail-errors effectively (27). Due to natural variability among microbial cultures, product

formulations, microbiological techniques, analyst proficiency, etc., it is impossible to eliminate both of these errors; however, the error rates can be reduced to acceptable levels, for example, by conducting multiple UDM tests. Setting acceptable bounds for the pass-error rate and the fail-error rate is not a fixed aspect in modeling, but rather is set by experts, policy makers, and/or scientific advisors. The error rates 0.01, 0.05, and 0.10 are commonly used in scientific hypothesis testing. For this initiative, a level of 0.05, for both error rates, was used as the baseline for the decision-making exercise.

Determining a New UDM Performance Standard

In order to find a PS that keeps both the pass-error and fail-error rates as low as possible while requiring a minimum

Table 5. Estimated error rate bounds, expressed as percentages, for hypothetical PSs for *S. aureus*^a

PASS ALL of this No. of tests	PS required to PASS each test ($N \leq N_{PS}$)	Max pass-error rate ($\alpha \leq$) (for $N_{target} = N_{PS} + 1$), %	Min fail-error rate ($\beta \geq$) (for $N_{high} = 0$), %
1	$N \leq 1$	24.38	6.44
1	$N \leq 2$	32.15	1.28
1	$N \leq 3$	36.32	0.34
1	$N \leq 4$	38.89	0.11
2	$N \leq 1$	6.15	12.43
2	$N \leq 2$	10.46	2.54
2	$N \leq 3$	13.27	0.68
2	$N \leq 4$	15.18	0.22
3	$N \leq 1$	1.59	18.01
3	$N \leq 2$	3.43	3.78
3	$N \leq 3$	4.87	1.01
3	$N \leq 4$	5.94	0.33

^a A disinfectant must pass all multiple tests as the numbers of tests and the allowable number of positive carriers (N) out of a 60 carrier set per test is varied. The error rates are representative of either single- or multiple-laboratory scenarios (due to the insignificant laboratory-to-laboratory variance for *S. aureus*).

amount of testing, a set of possible testing scenarios and PSs for the UDM were investigated. By considering Equations 5 and 6, sets of error rates were generated by varying the target number of positive carriers for each test (N_{target}), the number of tests that must be passed at each lab (K), and the number of labs performing the tests (L). In every case, a highly effective disinfectant was defined with respect to $N_{high} = 0$, and the allowable number of positive carriers in each test was $N_{PS} = N_{target} - 1$ (as in the previous section). Error rates are presented for two PS scenarios: the first involves tests with a single microbe; the second involves tests with both microbes. In addition, a PS that required averaging LRs across tests was considered (16); however, calculations showed that the testing requirements for this PS would be too resource-intensive to achieve acceptable error rates. Thus, this approach is not presented in this report.

Scenario 1.—Tests with a single microbe.—For a PS that requires that all tests for a single microbe passed bounds on the error rates are provided in Table 4 for *P. aeruginosa* and in Table 5 for *S. aureus*. This information allows a direct comparison to the UDM and HSCT PS error rates suggested by other authors (6, 11, 18). The results in Table 4 suggest that for a PS based only on tests of *P. aeruginosa*, a multilaboratory testing scenario is necessary to maintain a pass-error rate of 0.10 or less, and a fail-error rate as small as 0.05. For *S. aureus*, the error rates, especially the fail-error rate, are much lower than for *P. aeruginosa*, even when all tests are performed in the same laboratory. This is due to the negligible among-laboratory variance for *S. aureus* in the 2009 UDM collaborative study.

Scenario 2.—Tests with two microbes—Table 6 provides bounds on the error rates for a PS that requires that all tests for both *P. aeruginosa* and *S. aureus* pass. These results may be used as a tool to develop a practical and informed decision on

the best approach to selecting a practical PS and the associated testing requirements. The model's output suggests that in order to simultaneously maintain a pass-error rate less than 0.05 and a fail-error rate as small as 0.05, three tests (e.g., testing three production lots or testing a single lot three times) must be conducted for each microbe, all possibly performed in the same laboratory. Using this approach, a disinfectant would pass each *P. aeruginosa* UDM test if $N \leq 6$ [i.e., no more than six positive carriers out of 60 tested, corresponding to $LR \geq LR(6; 6) = 6.95$] and would pass each *S. aureus* UDM test if $N \leq 3$ [i.e., no more than three positive carriers out of 60 tested, corresponding to $LR \geq LR(3; 6) = 7.23$], in which case, the pass-error rate, α , is at most 0.028 and the fail-error rate, β , is as small as 0.049 (see the bolded error rates in Table 6). This pass-error rate is a slight increase compared to the current PS, for which α is at most 0.010, while the fail-error rate decrease is much larger compared to the current PS, for which β is as small as 0.416.

The error rates in Table 6 are based on an estimate of zero for the among-laboratory variance (i.e., $S_{lab}^2 = 0$) for *S. aureus*. This negligible among-laboratory variance is due to the large number of *S. aureus* tests (10 of the 15 total tests) in the 2009 collaborative study that exhibited zero positive carriers. Among the dozens of collaborative studies of disinfectant test methods that we have analyzed, only one other study produced $S_{lab}^2 = 0$. Typically, S_{lab}^2 was equal to S_{test}^2 , i.e., the median intralaboratory correlation coefficient among these collaborative studies was 0.5.

Using the computer code provided in ref. 16, one can check the error rates for other scenarios of concern. For example, to investigate the influence that a positive among-laboratory variance for *S. aureus* would have on the error rates, the model used to generate Table 6 was reconsidered, but this time with an among-laboratory *S. aureus* variance equal to the observed among-experiment variance, i.e., $S_{lab}^2 = S_{test}^2 = 0.100$, which results in a larger value for S_R via Equation 8. The results are presented in Table 7. Over the range of hypothetical correlations, the maximum pass-error rate for the proposed PS increased from 0.028 to 0.104, and the minimum fail-error rate from 0.049 to 0.092. Thus, in the presence of a moderate level of among-laboratory variability for *S. aureus*, the proposed PS maintains a pass-error rate of at most 0.104 and a fail-error rate as small as 0.092. In this scenario, to maintain a pass error rate of approximately 0.10 or less, multiple tests (e.g., three tests) are required (Table 7). Testing at multiple laboratories further reduces the pass-error rate.

Discussion

An important feature of the statistical model used in this assessment is transforming the number of positive carriers to an LR value. This is possible because the untreated control carriers are used to quantitatively measure microbial populations in the UDM (via the *TestLD*). The advantages of using the LR instead of analyzing the number of positive carriers some other way, e.g., other authors have used probit analysis (6) or logistic regression (18), are:

(a) the resulting value is directly interpretable as the LR of microbes over the entire 60 carrier set;

(b) 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;

Table 6. Estimated error rate bounds, expressed as percentages, for hypothetical PSs that require a disinfectant to pass all of multiple tests for both microbes^a

Correlation between Microbes	No. of labs	No. of Pa tests at each lab	No. of Sa tests at each lab	Current PS		Proposed PS	
				Max pass-error rate ($\alpha \leq$), %	Min fail-error rate ($\beta \geq$), %	Max pass-error rate ($\alpha \leq$), %	Min fail-error rate ($\beta \geq$), %
0	1	1	1	8.9	23.3	16.7	2.0
0.25	1	1	1	11.9	22.4	20.4	2.0
0.5	1	1	1	15.2	21.1	24.5	1.9
0.95	1	1	1	23.9	18.2	34.8	1.7
0	1	2	2	1.5	36.2	4.2	3.6
0.25	1	2	2	3.1	33.9	7.1	3.5
0.5	1	2	2	5.1	31.2	10.5	3.4
0.95	1	2	2	NC	NC	NC	NC
0	1	3	3	0.3	45.1	1.2	5.0
0.25	1	3	3	1.0	41.6	2.8	4.9
0.5	1	3	3	NC	NC	NC	NC
0.95	1	3	3	NC	NC	NC	NC
0	2	1	1	0.9	41.0	2.8	4.0
0.25	2	1	1	1.6	39.5	4.2	3.9
0.5	2	1	1	2.5	37.6	6.0	3.8
0.95	2	1	1	6.0	33.0	12.2	3.3
0	2	2	2	0.0	58.8	0.2	7.1
0.25	2	2	2	0.1	55.9	0.5	6.9
0.5	2	2	2	0.3	52.4	1.1	6.6
0.95	2	2	2	NC	NC	NC	NC
0	3	1	1	0.1	54.4	0.5	5.9
0.25	3	1	1	0.2	52.7	0.9	5.8
0.5	3	1	1	0.4	50.5	1.5	5.7
0.95	3	1	1	1.6	45.1	4.3	4.9

^a The hypothetical correlation between the LRs of tests of the different microbes, numbers of laboratories, and numbers of tests for each microbe at each laboratory is varied. For the current PS, the allowable number of positive carriers out of a 60 carrier set per test is $N_{PS} = 1$ for *P. aeruginosa* and $N_{PS} = 1$ for *S. aureus*; for the proposed PS, the allowable number of positive carriers is $N_{PS} = 6$ for *P. aeruginosa* and $N_{PS} = 3$ for *S. aureus*. NC = Not calculable.

(c) analyses, such as analysis of variance (ANOVA) and *t*-tests, appropriate for samples of approximately normally distributed LRs, can be directly applied; and

(d) the normality and homogeneity of variance assumptions are easily checked with well-known diagnostic tools (17). Experience indicates that the LR values satisfy the assumptions of normality and homogeneity of variance across a variety of different test microbes and standard methods (12, 14, 25, 26, 29–31).

The analyses depend on estimates of the variability of the LRs from the one and only collaborative study of the UDM in which control carrier counts were observed. This five-laboratory study included a single representative disinfectant product with a high level of efficacy. Although conventional guidelines for collaborative studies recommend the inclusion of multiple disinfectant formulations (28), a single-disinfectant study can yield results of practical value. If available resources are limited,

as was the case for the UDM collaborative study, it is advisable to use a single disinfectant in all tests. Otherwise, to reduce cost, one would be forced to use fewer laboratories, in which case, the reproducibility estimates S_R would be less reliable (i.e., S_R^2 would have lower degrees of freedom).

Experience shows that efficacious products having similar LR values also have similar reproducibility SDs. Highly efficacious treatments produce small S_R values, completely ineffective treatments produce small S_R values, and partially efficacious treatments produce medium or high S_R values (32–35). Therefore, we believe that the reproducibility SD, S_R , produced by the high efficacy treatment in the single disinfectant study of the UDM is representative of the reproducibility for most other high efficacy disinfectants and that the observed S_R can be used when crafting PSs. Support for this claim is found in the multilaboratory collaborative study results for the HSCT. The reproducibility SD, S_R , for the single-disinfectant UDM study

Table 7. Assuming an among-laboratory variance for *S. aureus* of the same magnitude as the observed within-laboratory variance, these are the estimated error rate bounds, expressed as percentages, for hypothetical PSs that require a disinfectant to pass all of multiple tests for both microbes^a

Correlation between microbes	No. of labs	No. of Pa tests at each lab	No. of Sa tests at each lab	Current PS		Proposed PS	
				Max pass-error rate ($\alpha \leq$), %	Min fail-error rate ($\beta \geq$), %	Max pass-error rate ($\alpha \leq$), %	Min fail-error rate ($\beta \geq$), %
0	1	1	1	11.08	29.44	18.4	4.4
0.25	1	1	1	14.42	27.80	22.2	4.3
0.5	1	1	1	18.11	25.75	26.4	4.0
0.95	1	1	1	28.31	19.53	37.5	3.0
0	1	2	2	3.76	43.56	7.6	7.8
0.25	1	2	2	6.17	40.51	11.0	7.5
0.5	1	2	2	9.22	36.87	15.0	6.9
0.95	1	2	2	NC	NC	NC	NC
0	1	3	3	1.83	52.30	4.1	10.6
0.25	1	3	3	3.57	48.36	6.9	10.1
0.5	1	3	3	6.01	43.72	10.4	9.2
0.95	1	3	3	NC	NC	NC	NC
0	2	1	1	1.32	49.92	3.4	8.5
0.25	2	1	1	2.19	47.62	5.0	8.3
0.5	2	1	1	3.42	44.66	7.0	7.9
0.95	2	1	1	8.21	35.12	14.1	5.9
0	2	2	2	0.16	67.72	0.6	14.9
0.25	2	2	2	0.42	64.25	1.2	14.4
0.5	2	2	2	0.91	59.84	2.3	13.4
0.95	2	2	2	NC	NC	NC	NC
0	3	1	1	0.17	64.27	0.6	12.5
0.25	3	1	1	0.35	61.84	1.1	12.2
0.5	3	1	1	0.67	58.61	1.9	11.6
0.95	3	1	1	2.43	47.60	5.3	8.7

^a The hypothetical correlation between the LRs of tests of the different microbes, the numbers of laboratories, and the numbers of tests for each microbe at each laboratory is varied. For the current PS, the allowable number of positive carriers out of a 60 carrier set per test is $N_{PS} = 1$ for *P. aeruginosa* and $N_{PS} = 1$ for *S. aureus*; for the proposed PS, the allowable number of positive carriers is $N_{PS} = 6$ for *P. aeruginosa* and $N_{PS} = 3$ for *S. aureus*.

are consistent with the S_R values for two earlier multidisinfectant HSCT collaborative studies for *P. aeruginosa* (14, 36).

For all the different high-efficacy disinfectant formulations tested in the UDM and HSCT collaborative studies, the range of S_R values was 0.09 to 0.66 and the median S_R was slightly less than 0.50. Thus, if an additional single-disinfectant collaborative study of the UDM was conducted with *P. aeruginosa* with a different high efficacy disinfectant, one would expect the resulting S_R to be about 0.50. Note that the statistical calculations for the UDM PSs proposed in this paper are based on $S_R = 0.53$ for *P. aeruginosa* from the UDM collaborative study. The pass-error and fail-error rates will be smaller for more effective disinfectants (since they are presumed to have smaller variability) and larger for less-effective disinfectants (that exhibit larger variability). The results for low efficacious

products have a greater likelihood of displaying variable results among different laboratories due to the variability associated with the treated carriers (i.e., varying levels of surviving bacteria).

Compared to *S. aureus*, the *P. aeruginosa* data in the 2009 collaborative study exhibited higher S_R values (Table 1). This may be due to the complexity of the *P. aeruginosa* test culture, which contains pellicle; the removal of the pellicle from the test culture is difficult to fully standardize. Small fragments of pellicle may lead to clumping of inoculum on inoculated carriers, which may result in more variable test results. However, this observation is based on a single collaborative study, and may not be evident for other methods or associated multilaboratory studies.

Due to the increased variability, the error rates for a PS based

solely on *P. aeruginosa* tests (Table 4) are larger than the same PS based solely on tests using *S. aureus* (Table 5). Thus, the number of positive carriers allowed for a passing outcome for each *P. aeruginosa* is predictably higher.

The approach in this paper may appear somewhat complex, since we present a framework for evaluating UDM PSs based on LRs (via Equation 6), and then use this framework to evaluate PSs based solely on numbers of positive carriers (via Equation 7) that have traditionally been used to evaluate product efficacy via the UDM. The critical difference in these two PSs is that the former explicitly accounts for the value of the *TestLD* for each test (i.e., the log-transform of the geometric mean of the control counts), while the latter does not. Consequently, a UDM PS based on numbers of positive carriers has a range of error rates depending on the realized value of the *TestLD* in each test. It was shown in ref. 16 that the error rates for the PS (6) based on LRs provide bounds on the error rates for the PS based on numbers of positive carriers, with equality if the realized *TestLD* in each test is 6. Thus, this paper reports estimates of the maximum possible pass-error rates and the minimum possible fail-error rates for PSs that specify only the allowable number of positive carriers per test.

An assessment of PSs for UDM tests that use a number of inoculated carriers different than 60 could also be performed. It is expected that increasing the number of carriers will have a negligible effect on the variability of UDM test outcomes since the among-laboratory variability of *P. aeruginosa* dominates (Table 1), and changing the number of inoculated carriers would only affect the within-laboratory variance component. Thus, all the results in this paper pertain to 60 carrier UDM tests.

Conclusions

Although liquid disinfectants have become more diverse over the years, and the UDM has been the subject of many studies and many revisions, the PS has not been reassessed and remains unchanged. The current investigation of the UDM PS utilizes the best available data and scientific practices, and is one of many aspects of the method that has been presented to AOAC for revision.

The multilaboratory UDM collaborative data set (12) provides a unique opportunity for reassessing the current UDM PS. Our investigation shows that the pass-error rate for a single UDM test, for either *P. aeruginosa* or *S. aureus*, is much higher than originally suggested by previous studies of the UDM (6) and the HSCT (11). The alternative PSs considered here reduce the error rates while maintaining a practical level of testing, and will provide reliable data for regulatory decisions. Based on the data analysis, the approach outlined in Scenario 2 of this paper provides an adequate and improved approach for the future use of the UDM as a regulatory procedure.

In conjunction with the number of required tests (three per microbe), use of $N \leq 6$ for *P. aeruginosa* and $N \leq 3$ for *S. aureus* as the passing requirement for each test should improve the overall confidence in the outcome of the testing scheme. Any of the six tests that fail to meet the PS would fail the product. When error rates for this proposed PS are calculated using the reproducibility SDs observed in the collaborative study, a low pass-error rate (of at most 0.028) is predicted for marginally ineffective products, and the fail-error rate for highly effective products is predicted to be lowered to approximately 0.05

(Table 6). While this suggested change may seem like a dramatic increase from the current PS that allows a single positive carrier per test, note that this current PS corresponds to the requirement that $LR \geq 7.60$ in each test, so that the allowable LR per test is being decreased by 0.37 for *S. aureus* (to $LR \geq 7.23$) and by 0.65 for *P. aeruginosa* (to $LR \geq 6.95$).

Recommendations

Based on the information provided in this report, the EPA is seeking AOAC's technical review of the proposed revisions. If approved by AOAC, the EPA will consider adoption of the revised performance standards for its antimicrobial regulatory program. AOAC is encouraged to publish the revised methods in the Official Methods of Analysis for distribution and reference. The following proposed revisions to AOAC Methods **955.15** and **964.02** are recommended:

The following revisions are proposed following section **C(f)** of each method:

(g) *Performance standard.*—For method **955.15** (*S. aureus*), Section **C(h)**: Performance standard: ≤ 3 positive (growth) carriers out of 60 tested. Conduct three separate UDM tests, one per day. All tests must meet the performance standard. Testing at a single laboratory is acceptable.

(g) *Performance standard.*—For method **964.02** (*P. aeruginosa*), Section **C(h)**: Performance standard: ≤ 6 positive (growth) carriers out of 60 tested. Conduct three separate UDM tests, one per day. All tests must meet the performance standard. Testing at a single laboratory is acceptable.

Note: Using the proposed Performance Standards, a hospital-level disinfectant (as defined by EPA) product must pass both *S. aureus* and *P. aeruginosa* tests to be considered effective (a total of six tests).

In addition to the proposed new PSs, meeting the minimum and maximum carrier count requirements (*TestLD* between 6 and 7). The use of the most current and revised version of the UDM further strengthens the reliability of the UDM and confidence in the resulting test data. Furthermore, if the new PSs are approved as First Action revisions, it would be possible to reassess them again following the generation of additional data. To monitor the impact of the proposed revisions, EPA would pursue additional analysis of data from the field if the new PSs are accepted and implemented.

Acknowledgments

This work was performed under contract EP-W-08-014 between Montana State University and the 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) *Official Methods of Analysis* (2012) 18th Ed., AOAC INTERNATIONAL, Gaithersburg, MD, Method **955.14**
- (2) *Official Methods of Analysis* (2012) 18th Ed., AOAC INTERNATIONAL, Gaithersburg, MD, Method **955.15**
- (3) *Official Methods of Analysis* (2012) 18th Ed., AOAC INTERNATIONAL, Gaithersburg, MD, Method **964.02**

- (4) Tomasino, S.F., Pines, R.M., & Hamilton, G.C. (2012) *J. AOAC Int.* **95**, 1059–1063. <http://dx.doi.org/10.5740/jaoacint.12-171>
- (5) Trout, J.R. (1985) *J. AOAC Int.* **68**, 763–765
- (6) Ortenzio, L.F., & Stuart, L.S. (1961) *J. AOAC Int.* **44**, 416–421
- (7) Morgan, B.J.T. (1992) *Analysis of Quantal Response Data*, Chapman & Hall, New York, NY
- (8) Cole, E.C., Rutala, W.A., & Samsa, G.P. (1988) *J. AOAC Int.* **71**, 1187–1194
- (9) Rutala, W.A., & Cole, E.C. (1987) *Infect. Control* **8**, 501–506
- (10) *Official Methods of Analysis* (2009) 18th Ed., AOAC INTERNATIONAL, Gaithersburg, MD, Method **991.47**
- (11) 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
- (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) U.S. Environmental Protection Agency (2012) *Standard Operating Procedure for AOAC Use-Dilution Method for Testing Disinfectants (MB-05)*, EPA Office of Pesticid Programs, Microbiology Laboratory, Environmental Science Center, Ft. Meade, MD. <http://www.epa.gov/pesticides/methods/atmpa2z.htm>
- (14) Hamilton, M.A., DeVries, T.A., & Rubino, J.R. (1995) *J. AOAC Int.* **78**, 1102–1109
- (15) Tilt, N., & Hamilton, M.A. (1999) *J. AOAC Int.* **82**, 384–389
- (16) Parker, A.E., Hamilton, M.A., & Tomasino, S.F. (2014) A Statistical Model for Assessing Performance Standards for Quantitative and Semiquantitative Disinfectant Test Methods. *J. AOAC Int.* **97**
- (17) Hamilton, M.A., Hamilton, G.C., Goeres, D., & Parker, A.E. (2013) Guidelines for the Statistical Analysis of a Collaborative Study of a Laboratory Disinfectant Product Performance Test Method. *J. AOAC Int.* **96**, 1138–1151
- (18) Arlea, C., King, S., Bennie, B., Kemp, K., Mertz, E., & Staub, R. (2008) *J. AOAC Int.* **91**, 152–158
- (19) Hamilton, M.A. (2011) *The P/N Formula for the Log Reduction When Using a Semiquantitative Disinfectant Test of Type SQ1, KSA-SM-08*, Center for Biofilm Engineering at Montana State University, Bozeman, MT. http://www.biofilm.montana.edu/resources/knowledge_sharing_articles
- (20) Cochran, W.G. (1950) *Biometrics* **6**, 105–116. <http://dx.doi.org/10.2307/3001491>
- (21) McCrady, M.H. (1915) *J. Infect. Dis.* **17**, 183–212. <http://dx.doi.org/10.1093/infdis/17.1.183>
- (22) Collett, D. (1991) *Modelling Binary Data*, Chapman & Hall, London, UK. <http://dx.doi.org/10.1007/978-1-4899-4475-7>
- (23) Blodgett, R.J. (2006) *FDA Bacteriological Analytical Manual*, Appendix 2. <http://www.fda.gov/Food/FoodScienceResearch/LaboratoryMethods/default.htm>
- (24) Brown, L.D., Cai, T.T., & DasGupta, A. (2001) *Stat. Sci.* **16**, 101–117
- (25) Hamilton, M.A., & DeVries, T.A. (1996) *Biometrics* **52**, 1112–1120. <http://dx.doi.org/10.2307/2533073>
- (26) Tomasino, S.F., & Hamilton, M.A. (2006) *J. AOAC Int.* **89**, 1373–1397
- (27) Casella, G., & Berger, R.L. (1990) *Statistical Inference*, 2nd Ed., Duxbury Press, Belmont, CA
- (28) *Official Methods of Analysis* (2005) 18th Ed., AOAC INTERNATIONAL, Gaithersburg, MD, Appendix D
- (29) Tomasino, S.F., Pines, R.M., Cottrill, M.P., & Hamilton, M.A. (2008) *J. AOAC Int.* **91**, 833–852
- (30) Tomasino, S.F., Pines, R.M., & Hamilton, M.A. (2009) *J. AOAC Int.* **92**, 1531–1540
- (31) Buckingham-Meyer, K., Goeres, D., & Hamilton, M.A. (2007) *J. Microbiol. Methods* **70**, 236–244. <http://dx.doi.org/10.1016/j.mimet.2007.04.010>
- (32) Bloomfield, S. ., & Looney, E. (1992) *J. Appl. Bacteriol.* **73**, 87–93. <http://dx.doi.org/10.1111/j.1365-2672.1992.tb04975.x>
- (33) Bloomfield, S. ., Arthur, M., Van Klingerren, B., Pullen, W., Holah, J.T., & Elton, R. (1994) *J. Appl. Bacteriol.* **76**, 86–94. <http://dx.doi.org/10.1111/j.1365-2672.1994.tb04420.x>
- (34) Springthorpe, V.S., & Sattar, S.A. (2005) *J. AOAC Int.* **88**, 182–201
- (35) Parker, A.E. (2011) *Interlaboratory Study for ASTM Method E2799-11 Standard Test Method for Testing Disinfectant Efficacy Against Pseudomonas aeruginosa Biofilm using the MBEC Assay* (2011) ASTM International, research report RR:E35-1006
- (36) Hamilton, M.A. (2001) *Statistical Report: Preliminary Analysis of the Second Collaborative Study of the Hard Surface Carrier Test*, August 31, 2001, EPA contract No. 68-W-99-015