# Testing Surface Disinfectants

This series of knowledge sharing articles is a project of the Standardized Biofilm Methods Laboratory in the CBE

KSA-SM-03

## Testing surface disinfectants: Desirable attributes of a standardized method

[*Key words*: reasonableness, relevancy, validity, ruggedness, resemblance, responsiveness, repeatability, reproducibility]

A standardized disinfectant test should possess an acceptable level of each of the attributes: ***reasonableness, relevancy, validity, ruggedness, resemblance, responsiveness, repeatability, and reproducibility***. These attributes provide the framework for evaluating disinfectant test methods. The eight attributes will be defined and described in the context of a quantitative or semi-quantitative surface disinfectant test, with special emphasis on the log reduction (*LR*) measure of efficacy. The evaluation of a new standardized method typically requires a sequence of experiments, beginning with one laboratory and culminating with a collaborative (multi-laboratory) study.

### Reasonableness

A reasonable test method can be conducted within practical limitations on time, materials, and labor. It must be relatively easy to learn. It should require only conventional or inexpensive laboratory equipment. Standard methods guidance documents (e.g., AOAC Appendix D, 2005) discuss the principles involved in determining whether a method is reasonable enough to merit evaluation by a collaborative study. Reasonableness usually is established during the early phases of method development. However, even a collaborative study will provide information about reasonableness if the study is followed by a survey in which a questionnaire or structured interview captures the laboratory cost and laboratory specialists' critical evaluations.

### Relevancy

A relevant test method emulates the real-world environment where the disinfectant will be applied. The need for relevancy has motivated the increased emphasis on tests against surface-associated microbes, in contrast to tests against planktonic microbes. Perfect relevancy is unachievable because it is impossible to create a laboratory test system that is a scaled-down version of the field application conditions. With a surface disinfection test, it is particularly difficult to match the surface area to volume ratio, microbial species and density, biofilm thickness, and biofilm topography. Many laboratory tests are accelerated, as compared to the time span over which surfaces are fouled and treated in the field. For these reasons, it is informative to conduct parallel laboratory and field experiments to check the relevancy of the laboratory protocol (e.g., Zelver et al. 1999), but such parallel studies seldom are performed.

Standard methods guidance documents (e.g., AOAC Appendix D, 2005) recommend that the protocol specifies the probable use of the method and delineates the range of applicability. Such specifications usually are determined during first phases of methods development. However, additional experience with the test sometimes uncovers influential relevance factors not considered

in the original protocol. If those factors are measured during subsequent studies, then the actual observed levels of the factors can be used to characterize more completely the conditions for which the test method is demonstrably applicable.

### Validity

The disinfectant test method is valid if the *LR* is unbiased; that is, over many tests of the same disinfectant treatment, the mean of the observed *LR* values equals the true *LR* for that treatment. A method is invalid if it is biased in that the observed *LR* values are consistently too high or consistently too low. In other fields of science, a bias assessment is conducted by comparing observed measurements to the known, true value. For example, an instrument is calibrated against known values to eliminate measurement bias. However, for disinfectant tests, the true LR is unknowable. For that reason, standard methods guidance documents usually point out that, for microbiological methods, there is no way to measure bias directly (AOAC Appendix D, 2005). Thus, it is impossible to calibrate a disinfectant test method and directly establish its validity.

An alternative, indirect approach is required. The prevailing strategy is to evaluate separately each key step in the method. In principle, if each step is unbiased then the overall method is unbiased. In practice, some steps are evaluated by expert judgment and some steps are evaluated empirically using focused experiments. As an example of the former evaluation, consider a step where the laboratory technician's subjective decisions could potentially bias the result. The expert will judge whether the protocol should be revised to remove that potential bias. To take a specific case, if half of the inoculated carriers will be treated and the other half will be untreated controls, potential selection bias can be averted if the half to be treated is chosen randomly. By requiring randomization, the technician's subjective decision is replaced by an unbiased selection procedure.

As an example of an empirical validity evaluation, consider the neutralization step, which is conducted to stop the action of the disinfectant when the designated contact time has been reached. The *LR* result would be biased upward if the neutralizer itself killed some microbes or if the neutralizer failed to stop disinfection activity completely. To provide an empirical demonstration that such bias does not affect the outcome, many disinfectant test protocols require a neutralization check experiment (e.g., ASTM E1054-08, 2008).

### Ruggedness

A rugged method is one for which the *LR* outcome is insensitive to minor perturbations of operational factors or environmental conditions. A ruggedness investigation can highlight the critical components of a laboratory method so that practitioners know which steps or conditions require special attention or which operational parameters must be optimized. There is no conventional quantitative measure for ruggedness, although several have been suggested (Thompson et al., 2002; Youden, 1975, pp 33-36; ASTM E1169-02, 2002; Goeres et al., 2005). AOAC Appendix D (2005) lists ruggedness as one of the criteria for deciding that a method is suitable for evaluation by a collaborative study, and it specifically advocates the ruggedness testing methods of Youden (1975).

Ruggedness evaluations of new disinfectant test methods are seldom published because this testing is conducted during the method development phase, prior to the method going to a collaborative study. At any stage of test development, a study could be conducted to measure potentially influential operational parameters or environmental factors that are not set by the test protocol. Then statistical analysis can elucidate the effects of those parameters and factors on the *LR* value (Goeres et al. 2005), thereby providing a quantitative ruggedness evaluation.

## Resemblance

Inoculated carriers are the experimental units in a surface disinfectant test. It is desirable for the carriers to resemble each other, both within a test and across separate tests; consequently, it is important to check for resemblance. For quantitative and semi-quantitative disinfectant test methods, resemblance assessment is based on viable microbe enumerations for control carriers. Such data are usually available because the *LR* measure of efficacy is calculated by subtracting the mean log density for the treated carriers from the mean log density for the control carriers (KSA-SM-02). In order for the control carrier data to be representative of the treated carriers, the control carriers usually should be prepared and observed side-by-side with the treated carriers.

Let LD denote the $\log_{10}$-transformed density for a control carrier, $n_C$ denote the number of control carriers required by the test method protocol, and *Test LD* denote the mean of the LD values averaged across the $n_C$ control carriers in the test. Variances and standard deviations are used to assess the extent to which carriers resemble each other. For a set of independent tests within a laboratory, an analysis of variance can be applied to the LD values to calculate the variance within a test day and the variance among test days. The resemblance summary typically includes the repeatability standard deviation (denoted here by $CS_r$) of the *Test LD* (e.g., the repeatability results in Goeres et al. 2005). If control carrier LD data are available from multiple laboratories, the analysis of variance applied to the combined data also can produce the variance among laboratories and the reproducibility standard deviation (denoted here by $CS_R$) for the *Test LD* values (e.g., the control count results in Tomasino et al. 2008). In our experience, $CS_r \leq 0.5$ indicates acceptable repeatability and the $CS_R \leq 0.7$ indicates acceptable reproducibility. (These guidelines are based on our experience with many resemblance assessment projects; see the Appendix for a summary table). It may be necessary to increase $n_C$ to achieve sufficiently small standard deviations for *Test LD*.

The protocol for a disinfectant test sometimes will require that the *Test LD* exceed a specified value; too small a *Test LD* nullifies the test. Sometimes the protocol also specifies a maximum value that the *Test LD* must not exceed; the maximum prevents the test from being irrelevant to the anticipated field application of the treatment. The resemblance assessment should include a comparison of the distribution of observed *Test LD* values to the specified range of acceptability (e.g., Tomasino et al. 2008 & 2009).

Resemblance assessment usually can be based on data pooled across tests of different disinfectant treatments because the control carriers should be inoculated and manipulated according to the same protocol regardless of the treatment. However, if the test employs neutralization procedures that differ among the treatments and the neutralization procedures are applied to control carriers as well as to treated carriers, then it may prove inappropriate to pool control data from tests of different treatments.

## Responsiveness

A responsive disinfectant test method is sensitive enough that it can detect an important efficacy–response effect and specific enough that sham treatments are not shown falsely to be effective. In a responsiveness evaluation study, the disinfectant test is applied at two or more efficacy levels of an established disinfectant, usually in side-by-side (parallel) testing. One can adjust the efficacy by altering the concentration of the active ingredient, by changing the contact time, or by setting influential variables such as temperature or pH. The observed *LR* of the presumably higher efficacy treatment should be discernibly greater than the observed *LR* of the presumably lower efficacy treatment. A collaborative study can show the extent to which different laboratories observe the

same increases in *LR* values when testing the same pair of presumably lower and higher efficacy levels (e.g., Fig. 2 in Tomasino et al. 2008).

### Repeatability

A repeatable disinfectant test method will produce nearly the same *LR* values in independent tests within a laboratory. In a repeatability study, the same disinfectant treatment is tested with the test method on different days in the same laboratory. The quantitative summary is the standard deviation of *LR* values, which is denoted by $S_r$ and is called the repeatability standard deviation. A small $S_r$ indicates good repeatability. In the disinfectant testing context, the repeatability variance ($S_r^2$) is the sum of two variance components, the within-test variance and the among-tests variance of *LR*. The within-test variance comprises both the variance among control carrier viable microbe measures and the variance among treated carrier viable microbe measures. In a review of the literature on standardized disinfectant tests, it was found that the observed $S_r$ ranged from 0.2 to 1.2, with a median of 0.5 (Tilt and Hamilton, 1999). We believe $S_r \leq 1$ indicates acceptable repeatability for most practical purposes.

### Reproducibility

A reproducible disinfectant test method will produce nearly the same *LR* value when the same disinfectant treatment is retested in a different laboratory. In a collaborative (or reproducibility) study, the same disinfectant treatment is tested by the same method in different laboratories. The quantitative summary is the standard deviation of the *LR* values, which is denoted by $S_R$ and is called the reproducibility standard deviation. A small $S_R$ indicates good reproducibility. The reproducibility variance ($S_R^2$) is the sum of $S_r^2$ and the variance of *LR* among laboratories. In a review of the literature on standardized disinfectant tests, typically 50% of $S_R^2$ was attributable to the variance among laboratories. Also, the literature $S_R$ values ranged from 0.3 to 1.5, with a median of 0.9 (Tilt and Hamilton, 1999). We believe $S_R \leq 1.3$ indicates acceptable reproducibility for most practical purposes.

## Appendix

**Table 1.** Median and range of *Test LD* repeatability and reproducibility standard deviations observed in 14 separate surface test evaluation projects that we have analyzed since 1999. For each of $S_r$ and $S_R$, N indicates the number of data sets underlying the summary and $n_C$ is the number of control carriers sampled. In some of the projects, more than one test method was evaluated. Some of the projects provided $CS_r$ and $CS_R$ for multiple choices of $n_C$. The large maximum values (> 0.9) are for test methods that standard methods organizations have not evaluated and approved at this time.

| | $S_r$ | | | | $S_R$ | | | |
|---|---|---|---|---|---|---|---|---|
| $n_C$ | Median | Min | Max | N | Median | Min | Max | N |
| 1 | 0.25 | 0.09 | 0.92 | 14 | 0.29 | 0.18 | 0.92 | 7 |
| 2 | 0.23 | 0.10 | 0.79 | 22 | 0.27 | 0.17 | 0.28 | 3 |
| 3 | 0.19 | 0.08 | 0.48 | 14 | 0.28 | 0.05 | 0.91 | 13 |
| 6 | 0.24 | 0.20 | 0.27 | 4 | 0.30 | 0.25 | 0.31 | 4 |

## References


AOAC International (2005) *Official Methods of Analysis 18th Ed. – Appendix D: Guidelines for collaborative study procedures to validate characteristics of a method of analysis*. AOAC International, Gaithersburg, MD.

ASTM Standard E1169-07 (2007): *Standard Practice for Conducting Ruggedness Tests*. ASTM International, West Conshohocken, PA.

ASTM Standard E1054-08 (2008): *Standard Test Methods for Evaluation of Inactivators of Antimicrobial Agents*. ASTM International, West Conshohocken, PA.

Goeres, D. M., L. R. Loetterle, M. A. Hamilton, R. Murga, D. W. Kirby, and R. M. Donlan (2005) Statistical assessment of a laboratory method for growing biofilms. *Microbiology* 151:757-762.

KSA-SM-02 (2010) Testing surface disinfectants: quantitative, semi-quantitative, qualitative, and alternative methods. KSA-SM-02

Thompson, M., S. Ellison, and R. Wood (2002). Harmonized guidelines for single-laboratory validation of methods of analysis. *Pure Appl Chem* 74:835–855.

Tilt, N. and M. A. Hamilton (1999) Repeatability and reproducibility of germicide tests: a literature review. *J.AOAC Int.* **82:**384 – 389.

Tomasino, S.F., R. M. Pines, M. P. Cottrill, and M. A. Hamilton (2008) Determining the efficacy of liquid sporicides against spores of *Bacillus subtilis* on a hard nonporous surface using the quantitative three step method: collaborative study. *J. AOAC Int*. 91(4):833-852.

Tomasino, S. F., R. M. Pines, and M. A. Hamilton (2009) Improving the AOAC use-dilution method by the establishment of minimum log density values for test microbes on inoculated carriers. *J AOAC Int.* 92(5):1531-1540.

Youden, W. J. (1975) *Statistical Manual of the AOAC – Statistical Techniques for Collaborative Tests,* AOAC International, Gaithersburg, MD.

Zelver, N., M. Hamilton, B. Pitts, D. Goeres, D. Walker, P. Sturman, and J. Heersink (1999) Methods for measuring antimicrobial effects on biofilm bacteria: from laboratory to field, Chapt. 45 in *Methods in Enzymology - Biofilms*, R.J. Doyle, editor, 310:608-628.



Version date: 10 June 2010



Lead Author: Martin A. Hamilton, Professor Emeritus of Statistics
mhamilton@biofilm.montana.edu