Testing Surface Disinfectants

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

KSA-SM-05

Testing surface disinfectants: how the differences between disinfectant tests and chemical assays affect method evaluation criteria

[*Key words*: collaborative study, biological variability, bias, repeatability conditions, biochallenge, efficacy-response, outliers, equivalence testing, performance standards]

Various official guidelines are available to steer the statistical analysis of collaborative (multi-laboratory) studies of chemical assay methods (e.g., AOAC Appendix D, 2005; ISO 5725-Parts 1-3, 1994, which also have been accepted by ASTM International). The guidance documents are written in generic terminology and conform to established scientific practice; consequently, they are widely used for method evaluation in fields of science other than chemistry. To our knowledge, there are no similar publications that pertain specifically to disinfectant tests. At present, the chemical assay documents may provide the best available guidance for evaluating a disinfectant test method.

This situation is far from ideal. There are a number of important ways in which disinfectant tests differ from chemical assays (Niemi and Niemela, 2001; Tillett and Sartory, 2004). Some important disinfectant test topics do not appear in existing guidance documents because those topics are not relevant to chemical assays. On the other hand, there is much information in the chemical study guidelines that does not pertain to disinfectant tests. In this article we describe the primary differences between surface disinfectant tests and chemical assays, focusing on differences that affect method evaluation criteria.

Differences between chemical assays and disinfectant tests

- *Disinfectant tests are affected by biological factors*

Springthorpe and Sattar (2005) review many important biological factors that influence the results of a disinfectant test. Those factors force researchers to use different design and analysis strategies than are employed for chemical studies. A disinfectant test is a bioassay in which the test subject is an individual microbe, and each microbe has a tolerance to the disinfectant treatment, where it is tolerant if it survives exposure to the treatment. That tolerance depends on the genotype and phenotype of the microbe as well as on the disinfectant treatment (formulation, concentration, contact time, etc.). Although biological factors are not completely controllable, the protocol for a disinfectant test will include extensive microbiological instructions, designed to create relatively consistent microbial tolerances across the carriers. Nevertheless, microbial tolerance to a disinfectant will necessarily vary among carriers, both within and among tests (Bloomfield and Looney, 1992).

To explain why unavoidable, inherent biological variability affects disinfectant efficacy, consider a disinfectant test against a bacterial biofilm. A biofilm is a self-organized community of bacteria that is heterogeneous (with respect to cell density, biofilm depth, chemical concentrations, diffusion rates, etc.) and expresses unique physiochemical and biological characteristics each time it is grown in the laboratory (Donlan, 2002; Boles et al., 2004). Moreover, those characteristics are dynamic in that they change with time. Even the genotypic and phenotypic states of a biofilm bacterium are not necessarily stable. Each individual bacterium's physiological state is determined by the interaction between its genetic makeup and the microenvironment surrounding it. Micro-scale (0.1-10 μm) environmental changes can cause a bacterium to experience phenotypic changes. The microenvironment surrounding a bacterium is affected by the topography and chemistry of the film and by other nearby bacteria. In fact, bacteria can exchange genetic components and can engage in cell-to-cell signaling that triggers individual phenotypic changes. The biofilm system is further complicated by the fact that bacteria and bacterial products can affect local chemistry. The bacteria capture and metabolize chemical molecules and secrete chemical products, thereby altering the chemistry in a system. The bacteria-created EPS that binds the bacteria to each other and holds the biofilm to the carrier surface may react with or impede the disinfectant, essentially neutralizing it. As a result, the aggregate tolerance of the biofilm community to a disinfectant treatment varies considerably among tests and among carriers in a test.

Because most microbes are invisible to the naked eye, it is difficult to see a microbe and to determine whether it is alive or dead. It is therefore challenging to measure antimicrobial effects and the measurement system itself contributes more variability than is typical for a chemical assay measurement. Overall, the results for disinfectant tests usually exhibit much more variability than is typical for chemical assay results (Tillett and Sartory, 2004; Tillett and Lightfoot, 1995).

- *The validity of a disinfectant test method cannot be evaluated empirically*

Chemical assays are usually conducted to determine the concentration of a target analyte (quantitative assay) or to detect the presence of a target analyte (qualitative assay). A method is invalid if it is biased; that is, if the measurements are consistently too high or consistently too low. When evaluating a quantitative chemical assay, usually there is a well-defined target value, at least for controlled environmental conditions. The target is known because the true concentration can be created artificially by spiking the sample with a known amount of the analyte. Alternatively, the concentration in the sample sometimes can be determined by analyzing it with a gold standard method. In either case, the bias of the chemical assay method can be estimated empirically by comparing the assay results to the true value.

The situation is quite different for disinfectant tests. Consider a quantitative disinfectant test method for measuring efficacy. The true log reduction (LR) value (as determined by the disinfectant test for a specific disinfectant treatment) can neither be artificially constructed nor precisely measured. Instead it is a conceptual quantity – the mean LR for an infinite number of independent tests. Because the true mean is unknowable, an empirical estimate of bias is impossible (Tillett and Sartory, 2004; AOAC Appendix D, 2005).

Existing statistical guidelines for evaluating chemical assays contain extensive information about the empirical assessment of bias, but that guidance is not applicable to disinfectant test studies. Validity for a disinfectant test method must be demonstrated prior to the collaborative study and based on evidence provided by pre-collaborative study evaluations and expert reviews. In a collaborative study, it is conventional practice to trust that the method is valid (Forster, 2009).

- *Repeatability conditions are different for disinfectant tests*

For a chemical assay, it is usually possible for the same technician to conduct several analyses of the same test sample, at essentially the same time, using the same laboratory instruments. For a chemical assay method, those several analyses are called repeats. Disinfectant tests, however, are conducted over longer time periods; for example, several weeks are required to grow, treat, sample, and analyze a mycobacterial biofilm. Between runs of a disinfectant test, the laboratory apparatus must be disassembled and sterilized. The apparatus may comprise some single-use components that must be discarded and replaced. When the apparatus is reassembled, it seldom is exactly the same as previously used. Because of the time interval between runs, fresh preparations of culture media and other chemicals, including the disinfectant treatment, may be required. A unique population of microbes is prepared for each independent test. For a disinfectant test, a "repeat" is an independent test observed in the same laboratory; however, it is at a later time, on different apparatus, perhaps by a different analyst, using a different population of microbes, and perhaps assessing a different preparation of the disinfectant (Forster, 2003). The repeatability assessment for a disinfectant test is based on these repeatability conditions.

- *The biochallenge in each disinfectant test must resemble the desired biochallenge*

Each disinfectant test method includes a protocol for preparing the initial population of microbes and placing them on the carriers (called the inoculation step or biofilm growth step). The group of microbes on a carrier can be viewed as the biochallenge to the disinfectant treatment. A good test protocol will create nearly the same biochallenge across the carriers in replicate tests, even if tests are done in different laboratories. It is not practical to measure fundamental microbial characteristics, such as the genotype/phenotype distribution. Instead, the biochallenge is measured by the density of viable microbes on a carrier. An important goal of a disinfectant test collaborative study is assessing the biochallenge resemblance of the carriers by a statistical analysis of the viable microbe densities on untreated carriers. Because resemblance is not an issue of concern for chemical assays, the existing statistical guidelines do not consider the assessment of resemblance.

- *The efficacy–response effect must be evaluated for a disinfectant test*

A collaborative study should determine whether a disinfectant test reliably discriminates between high efficacy and low efficacy disinfectants. The issue is important because the collaborative study potentially could show that the disinfectant test was too variable or too insensitive for practical use. The assessment of responsiveness is an important goal for the collaborative study of a disinfectant test method. However, statistical guidance for responsiveness assessment is not taken up in existing guideline documents because a concentration-response effect is not an issue in a chemical assay collaborative study. In fact, a quantitative chemical assay method that does not produce a consistent concentration-response calibration curve would be deemed unsuitable for collaborative study.

- *Automatic statistical outlier rejection methods are inappropriate for disinfectant test studies*

Guidelines for chemical studies often advocate the routine application of statistical outlier tests and the rejection of data identified by those tests. Usually a limit is set on the fraction of data that can be rejected. For example, it ordinarily is considered excessive to reject more than 2/9 of the data from each material in a study (AOAC Appendix D, 2005). Guidelines for the analysis of chemical studies usually list specific outlier tests and advocate the automatic elimination of outlying observations, repeats, or laboratories. For studies of disinfectant tests, however, the application of statistical rules for finding and removing outliers may well reject observations that are not errors, but represent the typical variability of the disinfectant testing method. For disinfectant tests, it often will be inappropriate to reject a data point simply because it is identified by a statistical outlier test.

Automatic outlier rejection rules potentially trim the natural extremes so that the estimated variances are systematically too small (Mandel 1998). Because a collaborative study of a disinfectant test method is expensive and time consuming, it is cost effective to utilize qualified analysts who can adapt the analysis to the statistical characteristics of the observations. Outlier detection and rejection decisions should not be determined automatically by prescribed statistical tests. All valid data should be included in the statistical analysis, where an observation is deemed invalid only if the Study Director has reason to believe that it is incorrect, but not correctable, or it is not statistically representative of the method under investigation (AOAC Appendix D, 2005). For example, data from a laboratory would be invalid if the Study Director discovered that the technicians in that laboratory did not follow the study protocol.

- *Equivalence tests are frequently required in disinfectant test collaborative studies*

As new technologies become available or efficient improvements are discovered, it may be advantageous to replace or modify the current disinfectant test (prevailing standard method). In a collaborative study, the modified method and the current method may be applied side-by-side to determine whether the efficacy results by the modified disinfectant test are equivalent to the results by the current test and whether the desired attributes are met at least as well by the modified test (Tomasino and Hamilton, 2006). Some guidance for establishing statistical equivalence between microbiological methods is available (Sartory, 2005), but at present, guidance is not available for disinfectant test studies.

- *Derivation of an efficient multiple test protocol depends on collaborative study information*

Crafting an efficient protocol is one of the final steps in preparing a new disinfectant test for standard use. The efficiency goal is to find a testing protocol that requires the least possible experimental effort to produce a precise efficacy estimate. Experience has shown that just increasing the number of carriers in a quantitative test does little to reduce the standard error. Instead, to achieve good precision, it usually is necessary to run multiple tests and average the efficacy values (Bloomfield and Looney, 1992; Bloomfield et al., 1993 and 1994). The goal is to determine the optimum numbers of untreated carriers and treated carriers per test, the number of tests in each laboratory, and the number of laboratories required to achieve an acceptably small standard error of the mean. The optimization can be refined to take into account the costs for each carrier, each repeat, and each laboratory (Marcuse, 1949; Johnson et al., 1993). Available collaborative study guidelines do not provide methods for estimating all the necessary variances, even though those variances are estimable from a collaborative study.

- *Derivation of a performance standard for the disinfectant test depends on collaborative study information*

If the disinfectant test potentially will be used for regulatory purposes, the collaborative study should provide information that pertains to performance standards (pass/fail criteria) for the test. The performance standard pass criteria will probably be constructed to provide confidence that the disinfectant truly achieves the target efficacy level that was specified by the regulatory authority. The performance standards will depend on the protocol for multiple testing and on the statistical characteristics of the test method, especially the variance components. The untreated carriers must contain enough viable microbes that the target efficacy can be attained with confidence. For collaborative studies of disinfectant tests, the guidance documents should cover the results needed to set performance standards; e.g., the variance components for the efficacy estimate, resemblance statistics, etc. Current guidelines do not discuss these issues.

Conclusion

We believe there is a pressing need for statistical guidance that focuses specifically on studies of surface disinfectant test methods. Subsequent articles in this KSA series will contain our recommended guidance for the statistical design, analysis, and use of a collaborative study of a surface disinfectant test method.

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.

Bloomfield, S. F. and Looney, E. (1992) Evaluation of the repeatability and reproducibility of European suspension test methods for antimicrobial activity of disinfectants and antiseptics. *J. Applied Bacteriology* 73:87-93.

Bloomfield, S. F., Arthur, M., Begun, and Patel, H. (1993) Comparative testing of disinfectants using proposed European surface test methods. *Letters in Applied Microbiology* 17:119-125.

Bloomfield, S. F., Arthur, M., Van Klingeren, B., Pullen, W., Holah, J. T., and Elton, R. (1994) An evaluation of the repeatability and reproducibility of a surface test for the activity of disinfectants. *J. Applied Microbiology* 76:86-94.

Boles, B. R., Thoendel, M., and Singh, P. K. (2004) Self-generated diversity produces "insurance effects" in biofilm communities. *PNAS* 101:16630-16635.

Donlan, R. M. (2002) Biofilms: microbial life on surfaces. *Emerging Infectious Diseases* 8:881-890.

Forster, L. I. (2003) Measurement uncertainty in microbiology. *J. AOAC Int.* 86:1089-1094.

Forster, L. I. (2009) Conclusions on measurement uncertainty in microbiology. *J. AOAC Int.* 92:312-319.

ISO 5725-1 (1994) *Accuracy (trueness and precision) of measurement methods and results -Part 1: General principles and definitions*. International Organization for Standardization, Geneva, Switzerland. (Cor. 1:1998).

ISO 5725-2 (1994) *Accuracy (trueness and precision) of measurement methods and results - Part 2: Basic method for the determination of repeatability and reproducibility of a standard measurement method*. International Organization for Standardization, Geneva, Switzerland. (Cor. 1:2002).

ISO 5725-3 (1994) *Accuracy (trueness and precision) of measurement methods and results - Part 3: Intermediate measures of the precision of a standard measurement method*. International Organization for Standardization, Geneva, Switzerland. (Cor. 1:2001).

Johnson, K., Lundman, R., and Hamilton, M. (1993) Efficient sampling designs for microbial processes: a case study, *J. Microbiological Methods,* 18:69-81.

Mandel, J. (1998) Interlaboratory study, pp 2073- 2076 in *Encyclopedia of Biostatistics,* Armitage, P. and Colton, T. (Eds), Wiley, New York.

Marcuse, S. (1949) Optimum allocation and variance components in nested sampling with an application to chemical analysis. *Biometrics* 5:189-206.

Niemi, R. M. and Niemelä, S. I. (2001) Measurement uncertainty in microbiological cultivation methods. *Accreditation and Quality Assurance* 6:372-375.

Sartory, D. P. (2005) Validation, verification and comparison: adopting new methods in water microbiology. *Water SA* 31(3):393-396.

Springthorpe, S. and Sattar, S. (2005) Carrier tests to assess microbicidal activities of chemical disinfectants for use on medical devices and environmental surfaces. *J. AOAC Int.* 88:182-201.

Tomasino, S.F., and Hamilton, M.A. (2006) Modification to the AOAC sporicidal activity of disinfectants test (Method 966.04): collaborative study. *J. AOAC Int.* 89:1373-1397.

Tillett, H. E. and Lightfoot, N. F. (1995) Quality control in environmental microbiology compared with chemistry: what is homogeneous and what is random? *Water Science and Technology* 31:471-477.

Tillett, H. E. and Sartory, D. P. (2004) Measurement uncertainty in water microbiology may not estimate the laboratory's precision. *Accreditation and Quality Assurance* 9:629-632.

Version date: 19 August 2010

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