



Development, standardization, and validation of a biofilm efficacy test: The single tube method

Authors: Darla M. Goeres, Diane K. Walker, Kelli Buckingham-Meyer, Lindsey Lorenz, Jennifer Summers, Blaine Fritz, Danielle Goveia, Grace Dickerman, Johanna Schultz, and Albert E Parker

©2019 This manuscript version is made available under the CC-BY-NC-ND 4.0 license <https://creativecommons.org/licenses/by-nc-nd/4.0/>

Goeres, Darla M., Diane K. Walker, Kelli Buckingham-Meyer, Lindsey Lorenz, Jennifer Summers, Blaine Fritz, Danielle Goveia, Grace Dickerman, Johanna Schultz, and Albert E. Parker. "Development, Standardization, and Validation of a Biofilm Efficacy Test: The Single Tube Method." *Journal of Microbiological Methods* 165 (October 2019): 105694. doi:10.1016/j.mimet.2019.105694.

**Development, standardization, and validation of a biofilm efficacy test:
the single tube method**

Darla M. Goeres^{1 2}, Diane K. Walker^{3 4}, Kelli Buckingham-Meyer³, Lindsey Lorenz^{3 5}, Jennifer Summers⁶, Blaine Fritz^{6 7}, Danielle Goveia⁶, Grace Dickerman⁴ and Johanna Schultz⁴, Albert E. Parker^{2 8}

Methods validated by a standard setting organization enable public, industry and regulatory stakeholders to make decisions on the acceptability of products, devices and processes. This is because standard methods are demonstrably reproducible when performed in different laboratories by different researchers, responsive to different products, and rugged when small (usually inadvertent) variations from the standard procedure occur. The Single Tube Method (ASTM E2871) is a standard method that measures the efficacy of antimicrobials against biofilm bacteria that has been shown to be reproducible, responsive and rugged. In support of the reproducibility assessment, a six-laboratory study was performed using three antimicrobials: a sodium hypochlorite, a phenolic and a quaternary/alcohol blend, each tested at low and high efficacy levels. The mean log reduction in viable bacteria in this study ranged from 2.32 to 4.58 and the associated reproducibility standard deviations ranged from 0.89 to 1.67. Independent follow-up testing showed that the method was rugged with respect to deviations in sonication duration and sonication power but slightly sensitive to sonicator reservoir degassing and tube location within the sonicator bath. It was also demonstrated that when a coupon was dropped into a test tube, bacteria can splash out of reach of the applied antimicrobials, resulting in substantial bias when estimating log reductions for the products tested. Bias can also result when testing products that hinder the harvesting of microbes from test surfaces. The culmination of this work provided recommended changes to the early version of the standard method E2871-13 (ASTM 2013b) including use of splashguards and microscopy checks. These changes have been incorporated into a revised ASTM

¹ ASTM study director, director of all follow-up investigations

² study design

³ Implementation of all studies

⁴ Splashguard bias study

⁵ Imaging

⁶ Ruggedness study

⁷ Harvesting bias study

⁸ Statistical analysis

method E2871-19 (ASTM 2019) that is the basis for the first regulatory method (ATMP-MB-20) to substantiate “kills biofilm” claims for antimicrobials registered and sold in the US.

Introduction

Antimicrobial products must pass performance criteria specific for different use sites before entering the marketplace. In the US, antimicrobial products for use on hard, non-porous surfaces have historically been tested using microorganisms dried onto a surface using semi- quantitative methods like the Use Dilution Method (Tomasino et al. 2012). In the last decade, a considerable effort has been made to update these historical tests with more quantitative test methods (Tomasino 2013). Validated quantitative methods are needed that test antimicrobials against planktonic bacteria in suspension, planktonic bacteria dried onto a surface, loosely surface-associated bacteria and biofilm bacteria. (Tomasino 2013).

Fundamental research has demonstrated that biofilm bacteria are phenotypically different than planktonic bacteria of the same genotype (Sauer et al. 2002, Mah and O'Toole 2001). Biofilm not only consists of the bacteria, but also contains extracellular matrix (EPS) that glues the cells together and provides a boundary layer of protection. Depending upon where the biofilm is living, the EPS may also contain a variety of inorganic and organic components. Because the bacteria are in close proximity to each other, they can share genetic material and communicate via quorum sensing (Lewandowski and Beyenal 2013). A consequence of these differences from the planktonic state, biofilm bacteria are more tolerant to antimicrobials and antibiotics (Stewart 2015, Davies 2003, Buckingham-Meyer, Goeres, and Hamilton 2007, Behnke et al. 2011) and more variable in their responses (Parker, Hamilton, and Goeres 2018). A practical outcome of this understanding of biofilms is that researchers, antimicrobial manufacturers, pharma industries, and regulators must use newer biofilm methods, not conventional planktonic methods, to measure the efficacy of antimicrobials against biofilm bacteria (EPA 2016b). Since the manner in which a biofilm is grown impacts its tolerance (Buckingham-Meyer, Goeres, and Hamilton 2007, Stewart

2015, Manner et al. 2017, Stewart and Parker 2019), then standard biofilm methods using different reactor systems that represent a variety of environmental conditions are necessary to predict how an antimicrobial will perform.

Standard(ized) methods, *i.e.*, those taken through the standardization process with a consensus-setting organization such as ASTM International, are the tools researchers, industry and regulators use to evaluate antimicrobial product performance. Standard methods development is the creation of laboratory protocols for the purpose of comparison, both within a single laboratory and among laboratories. The impetus for the development of many microbial standard methods is efficacy testing for product registration with a regulatory agency such as the US Environmental Protection Agency (EPA) or the US Food and Drug Administration (Tomasino 2013). To protect public health, regulatory agencies require efficacy data when a product is registered. Table 1 gives the definition of the “statistical R’s” that encapsulate the desirable attributes of an antimicrobial test method. For product registration, standard methods that are repeatable, reproducible, rugged and responsive are absolutely required (Hamilton et al. 2013). A standard method should also be reasonable, meaning it should utilize equipment that is “typical” for a laboratory and it should not require an excessive amount of time, supplies or highly specialized training. Most importantly, standard methods, unlike research methods, undergo rigorous multi-laboratory statistical evaluations of these attributes.

Table 1. Required characteristics for a standard method. SD denotes standard deviation.

Statistical	
Characteristic	Definition
Repeatable	<p>Independent repeats of the same experiment in the same laboratory produce nearly the same result, as indicated by a small <i>repeatability SD</i>. An ANOVA provides the SD and determines how the variance partitions into within experiment, among-experiment, and among-technician components. A single lab study is required.</p>
Rugged	<p>Slight changes to the SOP for the method do not result in a substantially different method response. Ruggedness is assessed by the coefficients in a regression analysis. A fractional factorial design is required to assess interaction effects.</p>
Responsive	<p>A method is able to distinguish between treatments of high and low efficacy. A responsive method is also called sensitive. A regression or ANOVA is used to evaluate responsiveness. A study with antimicrobials with a range of efficacy is required.</p>
Reproducible	<p>Repeats of the same experiment run independently by different researchers in different laboratories produce nearly the same result, as indicated by a small <i>reproducibility SD</i>. An ANOVA provides the SD and determines how the variance partitions into the repeatability component and an among-laboratory component. A multi-lab study is required.</p>

An effective strategy in biofilm methods development is to partition the methods into a sequence of 4 discrete steps: Grow, Treat, Sample and then Analyze. The first step is to grow a reproducible and relevant biofilm. The second step is to expose the mature biofilm to an antimicrobial product. The treated biofilm is then sampled (harvested) and then the effects of the antimicrobial are analyzed by quantifying the biofilm bacteria that survived the treatment.

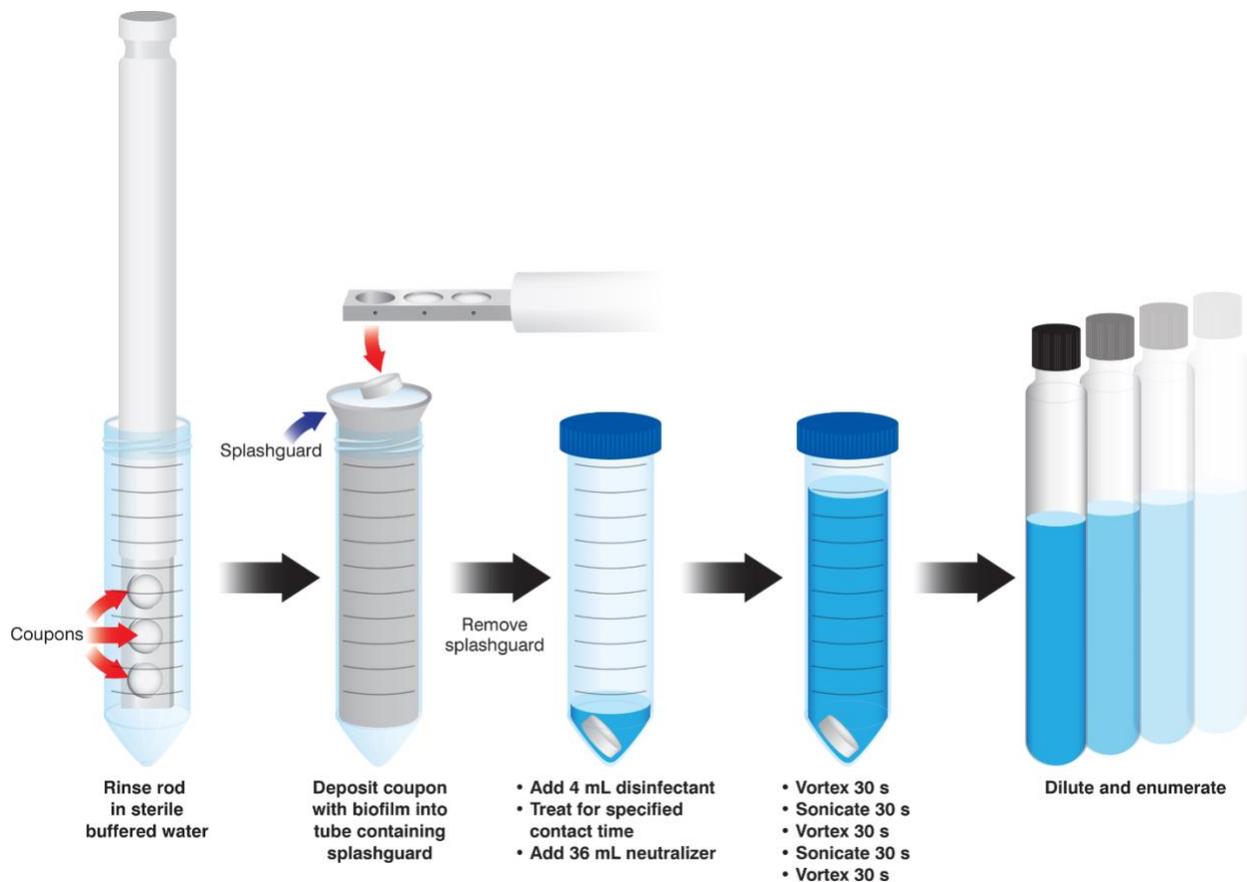
This paper describes the development and validation of the Single Tube Method (STM, ASTM E2871), a standard laboratory method that tests the efficacy of antimicrobial products against biofilms. The STM was developed with the interconnected but discrete steps of Grow, Treat, Sample and Analyze in mind. This means, for example, that although the STM was validated using a robust *P. aeruginosa* biofilm grown in the high shear environment of the CDC reactor (Goeres et al. 2009, ASTM 2007), the STM treatment efficacy step was intentionally designed to accommodate biofilm grown with any microorganisms of interest (e.g., the CDC reactor has been used to grow biofilms of *Staphylococcus aureus* (Lineback et al. 2018), *Candida tropicalis* (Fernandez-Rivero et al. 2017), or a consortia (Yoon and Lee 2017)) in any reactor system. The STM provides a quantitative measurement of the log reduction in viable biofilm bacteria that results from exposing a mature biofilm to an antimicrobial for a specified contact time. The STM was designed to treat and sample biofilm in a single tube and volume; hence, the STM does not differentiate between cells killed on the surface and cells that are removed from the surface, for example by a surfactant in the antimicrobial product formulation, then killed by active agents in the bulk liquid. A method that exposes a biofilm-coated surface (coupon) to an antimicrobial treatment in one vessel and then transfers that coupon into a second vessel for harvesting can provide data separately on removal of the biofilm from the surface (from viable cell counts from vessel 1) and kill (remaining viable cells attached to the coupon from vessel 2). The STM measures kill only.

Here we describe the early life cycle of the STM including the results from a six-laboratory collaborative study that led to its initial standardization by ASTM as E2871-13 (ASTM 2013b), subsequent ruggedness testing on the harvesting and disaggregation steps, and investigations of potential bias that can result from inadequate harvesting and coupon splashing. This work culminated in a revised version of the STM as E2871-19 (ASTM 2019) and the first regulatory test, ATMP-MB-20, for substantiating efficacy claims against biofilms in the US (EPA 2016a). This evolution of the STM demonstrates that standard methods are living documents that are reviewed and updated to incorporate an improved understanding of the system under investigation and changes in technology.

Materials and Methods

The Single Tube Method

ASTM E2871-19, the most current version of the Single Tube Method (ASTM 2019), is summarized in Figure 1 and demonstrated in a video created by the Standardized Biofilm Methods Laboratory in the Center for Biofilm Engineering at Montana State University (<http://www.biofilm.montana.edu/standardized-biofilm-methods-training-videos.html>).



© 2016, MSU Center for Biofilm Engineering, J Parchen.

Figure 1. Diagram of the Single Tube Method process E2871-19 (ASTM 2019) which includes the splashguard).

Here, we describe E2871-13 and then the subsequent steps that led to E2871-19. First, a *Pseudomonas aeruginosa* biofilm (see Figure S1 in Supplementary Material for a confocal scanning laser microscope image of a typical biofilm) was grown in the high shear environment of a CDC Biofilm Reactor (Figure S2 in Supplementary Material) according to standard practice E3161 (ASTM 2018b), a modified version of ASTM E2562 (ASTM 2007). Modifications were the use of: *P. aeruginosa* ATCC 15442 instead of *P. aeruginosa* ATCC 700888, borosilicate glass coupons instead of polycarbonate coupons, and a protocol to clean coupons individually thereby minimizing scratches to the glass coupon surfaces. Assembly and operation of the CDC reactor were unmodified from the standard E2562. In brief, coupons

were loaded into rods, inserted into the reactor and the system was sterilized. Nutrient broth was added to the reactor and inoculated with *P. aeruginosa*. At the end of a 48 hour growth phase (stirring 24 hr in batch and 24 hr under a continuous flow of fresh nutrients), coupons with a mature biofilm were ready for the efficacy test. Coupons were rinsed by gently dipping each reactor rod in sterile buffered dilution water (Figure 1). A rod was held horizontally over a sterile tube so that there was no contact between the rod and the tube (according to ASTM E2871-13 without the splashguard in Figure 1). A set screw was loosened to allow a coupon to fall freely to the bottom of the tube. A coupon that did touch the tube wall as it fell was not used in the study as the biofilm would be disrupted and smeared on the tube wall leaving behind cells that would not end up in contact with the antimicrobial.

Each coupon was exposed to 4 mL of an antimicrobial treatment or the control (dilution water) for a contact time of 10 minutes at $20 \pm 1^\circ\text{C}$ with no mixing. Immediately after addition of the treatment or control solutions, each tube was tapped to remove air bubbles and checked to ensure that the coupons were angled for maximum surface contact with the antimicrobial.

At the end of the contact time, 36 mL sterile 2X D/E broth was added to each tube (neutralization was verified according using ASTM Method E1054 (ASTM 2002b) prior to the start of the study (results not included). The tubes were capped and shaken vigorously.

Biofilm sampling was performed via vortexing and sonication. The tubes underwent a vortex/sonicate series of 30 seconds each in a total of 5 steps, beginning with vortexing at 100% power. The tubes were then transferred to a test tube rack and submerged within a degassed sonicating water bath to a depth level with the tube contents. The sonicator was operated at 10% power and 45 kHz in sweep mode. The disaggregated samples were serially diluted, spread plated on R2A agar and incubated at $36 \pm 2^\circ\text{C}$ for 24 – 28 hours.

Multi-laboratory Study

Three antimicrobials were used in a 6-lab study of the STM for the purpose of determining reproducibility of the STM: NaOCl (6.15% sodium hypochlorite), phenol (10.5% o-benzyl-p-chlorophenol, 10.5% o-phenylphenol), and quat-alcohol (0.28% diisobutyl-phenoxyethoxy-ethyl dimethyl benzyl ammonium chloride, 17.20% isopropanol) (ASTM 2013a). The study began with 9 labs but due to deviations from the study protocol regarding plate counting and sonication techniques, three labs were excluded (ASTM 2013a), leaving data from the remaining 6 labs for the statistical analysis. On each test day, each laboratory applied each antimicrobial to 3 coupons. There were also 3 control coupons for a total of 12 coupons per test day. The multi-laboratory study required that the mean $\log_{10}(\text{CFU}/\text{cm}^2)$ for the control coupons for each test was between 8.4 and 9.7. Two efficacy levels (Low or High) for each antimicrobial were attained by diluting with DI water as follows: NaOCl (HIGH 3075 ppm, 1:20 dilution (1 + 19), LOW 123 ppm, 1:500 (1 + 499)); phenol (HIGH 1:16 (1 + 15), LOW 1:25 (1 + 24)); quat-alcohol (HIGH undiluted, LOW 1:2 (1 + 1)). Each efficacy level was tested on a separate day for a total of 6 test days.

Ruggedness to Sonicator Factors

The first independent study that followed the multi-laboratory study was to critically evaluate the use of the sonicator. Coupons with biofilm grown in the CDC reactor were processed according to the STM by exposure to either 3075 mg/L sodium hypochlorite (treated) or dilution water (control) for 10 minutes. At the end of the contact time, all samples were neutralized with 2X D/E broth. Biofilm sampling was then conducted with variations/deviations over a factorial experimental design of 16 independent runs that purposely varied degassing (yes or no), sonication power (20 or 200 Watts), sonication time (25, 30, 35 seconds) and placement (suspended in the bath or placed on the bottom of the bath, the latter of which being a practice *not* recommended by the water bath manufacturer, but known to occur in some labs). The intent of these experiments was to assess how these variations, which are what might be expected to vary amongst labs, affect the outcomes of the method.

Development & Evaluation of Splashguards

In addition to the observation that biofilm would smear and remain untreated if a coupon hit the inner wall as it was deposited into a tube, it was also observed that even if the coupon fell cleanly to the bottom, biofilm could splash above the treatment zone (*i.e.*, above the 4 mL volume of antimicrobial that would be added to the tube to cover the biofilm-coated coupon for treatment). Hence, the second independent study to come after the multi-laboratory results was to 1) evaluate the amount of biofilm that splashed out of treatment range and 2) compare different approaches for depositing the coupons into tubes. First, coupons were deposited as previously described and a sterile swab was used to wipe the inside wall of the tube down to the 4 mL mark. The swab was then placed in 10 mL sterile buffered dilution water and processed according to the STM vortex/sonicate series with serial dilution and plating on R2A agar. Next, three methods of coupon deposition were compared: (1) direct drop into tubes as previously described, (2) placement into tubes with forceps, and (3) using a splashguard. The splashguard (Figure S3) is a cylindrical tube of slightly smaller diameter that fits inside the 50 mL conical tube. To evaluate, a sterile splashguard was placed into each tube. Coupons were dropped into the tubes and the splashguards removed, taking any splashed droplets out of the system. After the coupons were dropped or placed into the tubes according to these three methods, the coupons were processed according to the STM and either treated with 5000 mg/L, pH 7 sodium hypochlorite or sterile buffered dilution water.

Harvesting Bias Checks

A third study was performed to evaluate the differential harvesting and disaggregation efficiencies between treated and control coupons processed (Hamilton, Buckingham-Meyer, and Goeres 2009) following the STM. These 3 antimicrobials, tested at 3 concentrations each, were suspected of hampering the harvest of biofilm bacteria surviving treatment: Divosan® (a hydrogen peroxide, acetic acid and peracetic acid blend) at 1:10, 1:100 and 1:1000, Diverflow® (a sodium hydroxide blend) at 1:10, 1:25 and 1:50; and TSA Acid Sanitizer (an acid cocktail) at 1:10, 1:100, 1:800. Viable plate counts, microscopy and crystal violet data were collected. For the microscopic evaluation, treated and control

coupons were stained for 20 minutes with BacLight Live/Dead stain (ThermoFisher Scientific) prepared according to the manufacturer's instructions. Coupons were rinsed with sterile filtered deionized water and imaged using an upright Leica TCS-SP5 confocal scanning laser microscope using the 488 and 561 nm laser excitation lines with a 1.25X extra-long working distance objective. Crystal violet assays were performed in two ways to evaluate the efficiency of the biofilm harvest and disaggregation processes; 3 control coupons were stained with crystal violet prior to the vortex/sonication cycle, while three control and three treated coupons were stained for 15 minutes with 5 mL of 1% crystal violet prepared in deionized water. After staining, coupons were rinsed by gentle dipping into three 30 mL volumes of dilution water then placed in 5 mL of 95% ethanol for 10 minutes. The absorbance of the eluted crystal violet was measured in a spectrophotometer (540 nm).

Statistical Methods

All CFUs were transformed to log densities (LD, $\log_{10}(\text{CFU}/\text{cm}^2)$). A log reduction (LR) was calculated for each treatment in each experiment by subtracting the mean LD of the treated coupons from the mean LD of the control coupons. A linear mixed effects model (lmm) was fit to the data from each study separately. For the multi-laboratory study, a lmm was fit to the control LDs with nested random effects for lab and experiment. A separate lmm was fit to the LRs for each antimicrobial and efficacy level combination with a random effect for lab. For each of the ruggedness and splashguard studies, the lmm also included a covariate or factor for each of the 4 operational factors under study. Splash bias of controls was assessed with statistical equivalence using an equivalency margin of $0.2 \log_{10}(\text{CFU}/\text{cm}^2)$, *i.e.*, differences on the average as large as $0.2 \log_{10}(\text{CFU}/\text{cm}^2)$ can be considered negligible and not of practical importance.

Results

Multi-laboratory study

The untreated control LDs ($\log_{10}(\text{CFU}/\text{cm}^2)$) from the inter-laboratory study of the STM were analyzed to assess the resemblance of the controls across labs, and across experiments within each lab. (Figure S4 in the Supplementary Material provides a graph of the LDs of the control data). The reproducibility SD = 0.2442 of the LDs indicates excellent reproducibility of the control biofilms grown in the high shear environment of the CDC reactor (ASTM E3161). Based on a review, (Parker and Hamilton 2011, Tilt and Hamilton 1999) suggest that, for microbial control data, an acceptable reproducibility SD is less than 0.7. Most of the reproducibility variance of the STM controls was due to among-lab sources (78%) with the rest due to within-lab (repeatability) sources (22%). The repeatability SD = 0.22, aggregated across the 6 labs, indicates excellent repeatability of the control biofilms. The review presented by Parker and Hamilton (2011), Tilt and Hamilton (1999) suggest that, for microbial control data, an acceptable repeatability SD is less than 0.5.

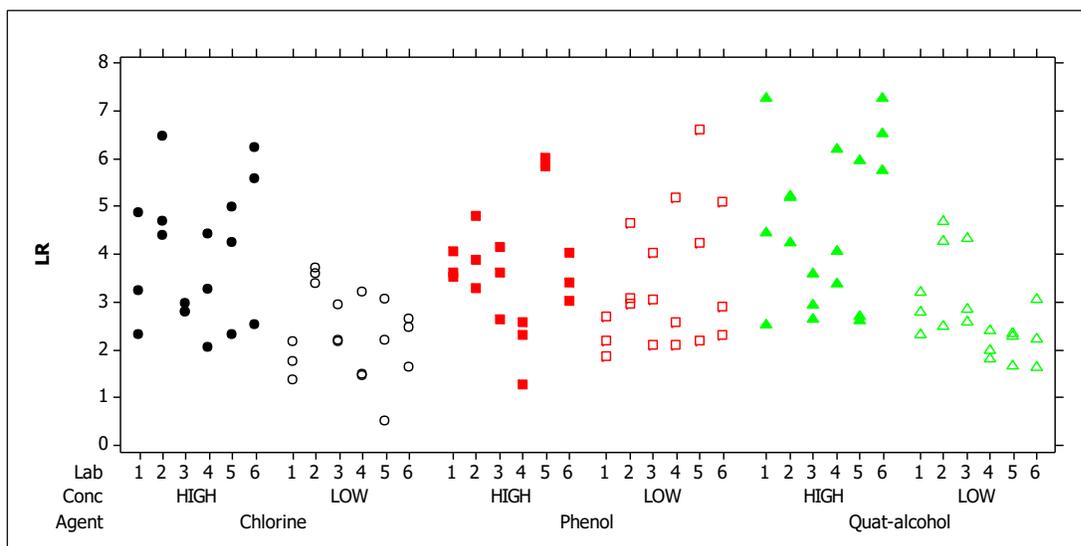


Figure 2. Each point in the figure is the LR for a single antimicrobial agent in a single experiment of the 6-lab collaborative study. The horizontal axis lists all 3 agents, the 2 concentrations tested for each antimicrobial, and the 6 labs (numbered 1-6).

Figure 2 presents the LRs attained by each antimicrobial in the 6-lab study. Table 2 provides a quantitative assessment of the repeatability and reproducibility of the LRs for each antimicrobial and

efficacy level combination. The responsiveness of the STM was assessed by comparing the low and high efficacy levels for each antimicrobial. The STM was statistically significantly responsive when comparing the two levels of NaOCl (LR difference of 1.57, $p = 0.002$) and the quat-alcohol (LR difference of 1.87, $p = 0.013$), but not for the phenol (LR difference 0.45, $p = 0.170$).

Table 2. Repeatability and reproducibility of the LRs from the 6-lab collaborative study. LRs are unitless, SD denotes standard deviation.

Chemical	Conc	Mean	Within	Among	Repeatability	Reproducibility
		LR	lab %	lab %	SD	SD
NaOCl	HIGH	3.89	86%	14%	1.306	1.404
	LOW	2.32	71%	29%	0.751	0.891
Phenol	HIGH	3.78	18%	82%	0.580	1.349
	LOW	3.32	100%	0%	1.345	1.345
Quat-alcohol	HIGH	4.58	75%	25%	1.448	1.668
	LOW	2.71	63%	37%	0.731	0.922

Ruggedness to Sonicator Factors

To better understand what may have contributed to the high variability of the LRs for some of the treatments in the multi-laboratory study, a ruggedness test using a factorial design with center points was performed that studied variations in the following 4 settings: degassing the sonicator bath, sonication power, sonication duration, and placement of the tubes in the sonicator bath. The resulting LRs for a single NaOCl treatment, displayed in Figure 3, show a similar level of variability as seen in the 6-lab study. The ruggedness results are summarized by a regression equation in Table 3,

$$\text{LR} = 2.819 - 0.0065 (\text{Time} - 30) - 0.00037 (\text{Power} - 10) + 0.233 \text{ Degassed} \\ + 0.278 \text{ Placement.}$$

Time is in minutes; Power is in Watts; Degassed = 1 if the sonicator bath was degassed, otherwise Degassed = 0; and Placement = 1 if the tubes are suspended in a rack in the bath, otherwise Placement = 0. This equation allows one to predict the effect of the sonication parameters on the LR. The equation does not include any interactions among the explanatory variables because significance tests and graphical assessments failed to suggest any interaction effects. The magnitudes of the coefficients in the equation and the significance tests summarized in Table 3 suggest that the STM is rugged with respect to sonication duration (Time) and sonication power (Power), but not rugged with respect to either degassing or placement. Though not statistically significant, the large coefficients for Degassed and Placement suggest that the LR is predicted to be increased by as much as 0.5 logs when the sonicator bath is degassed and when the tubes are suspended in a rack in the sonicator bath (as specified by the method) as opposed to being placed on the bottom the bath. These findings are substantiated by a retrospective survey collected from the labs in the multi-lab study (ASTM 2013a).

Table 3. Regression output for LR as a linear function of sonication parameters based on results using sodium hypochlorite in a ruggedness test.

Term	Coef	SE Coef	95% CL	P-Value
Constant	2.819	0.540	(1.63,4.01)	0.000
Time-30	-6.5e-3	0.0152	(-0.0399,0.0269)	0.677
Power-10	-3.7e-3	5.06e-3	(-0.0115,0.0108)	0.943
Degassed	0.233	0.455	(-0.769,1.24)	0.619
Placement	0.278	0.445	(-0.769,1.28)	0.554

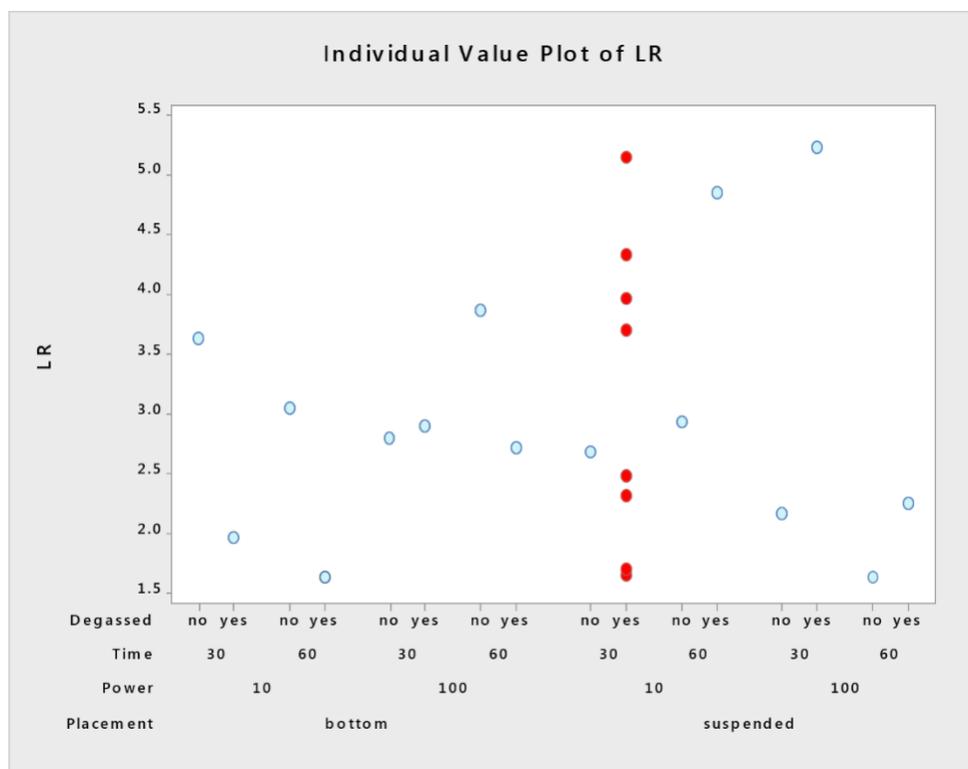


Figure 3. Each point in the figure is the LR for sodium hypochlorite in a single lab ruggedness study. The red points indicate LR values from running the standard STM protocol. The other points indicate LR values from the STM run with deviations to placement, degas, sonication power, and/or sonication time.

Splash Bias

Swabbing found nearly $6 \log_{10}(\text{CFU}/\text{cm}^2)$ viable bacteria on the inside wall of the tubes after dropping coupons into tubes using the STM. Although this is less than 1% of the typical $10^8 - 10^9$ biofilm bacteria grown on each control coupon, such a splashing effect could result in too many bacteria being attributed to the treated carriers which would bias the measured LR, especially for products that kill everything. To assess the potential for bias, three different methods of placing the coupons into the conical tube were investigated: dropping, placement with forceps, or dropping with a splashguard tube insert (Figure S3). In all 3 scenarios the coupons were then treated with 5000 mg/L, pH 7 NaOCl. Importantly, with 95% confidence, the control LDs for the 3 methods (not shown) were found to be statistically equivalent (*i.e.*, the same biochallenge was used for each coupon placement method). The LR values for this NaOCl treatment applied in 3 experiments are shown in Figure 4. The drop method gave LR values that were 0.69 less on

average than the LRs from the forceps method ($p = 0.038$) and 0.50 less on average than the LRs from the splashguard method ($p = 0.102$). This is consistent with the hypothesis that dropping the coupon causes a splash of bacteria onto the insides of the tube that are never subjected to the 4 mL volume of treatment. The LRs generated by NaOCl when using the forceps and splashguard methods were not statistically significantly different ($p = 0.568$) and in fact were statistically equivalent as long as mean LR differences as large as 0.70 can be considered negligible and not of practical importance. These results suggest that (1) dropping the coupons into the tube tend to bias downward the LR for the treatment being tested and (2) using either forceps or a splashguard can mitigate this negative bias on the LR.

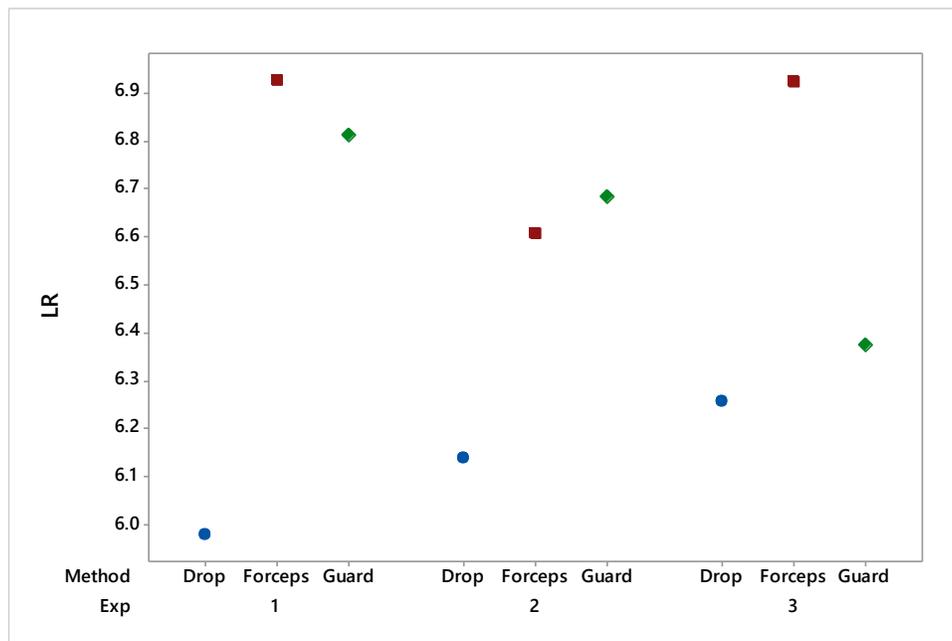


Figure 4. Each point in the figure is the LR for a treatment in a single lab study comparing three different methods of coupon placement.

Harvesting Bias

A different potential source of bias could result if the action of an antimicrobial either inhibits or enhances the proportion of viable biofilm cells harvested from the treated coupons compared to the controls (Hamilton, Buckingham-Meyer, and Goeres 2009). For antimicrobials that inhibit harvesting, the resulting LRs and percent reduction (PRs) would be biased upwards (*i.e.*, are too large). For

antimicrobials that enhance harvesting, the resulting LRs and PRs would be biased downwards (*i.e.*, are too small). The PR of the biofilm due to harvesting, calculated from optical densities of crystal violet staining (Figure S5 in Supplementary Material), was calculated for 3 different antimicrobials, each at 3 different concentrations. The PRs are shown in the top pane of Figure 5. When testing these 3 antimicrobials, 90% of the bacteria were harvested from the controls (*i.e.*, harvesting efficacy was 90%). However, at most 60% of the surviving bacteria were harvested from coupons treated by two of the antimicrobials tested (2 high concentrations of Divosan® (a hydrogen peroxide/acid blend) and all 3 concentrations of Diverflow® (sodium hydroxide)). Our data demonstrate that high concentrations of Divosan and Diverflow statistically significantly impeded the harvesting of biofilm bacteria on the treated coupons, which means that the calculated LRs for these treatments are biased upwards by between 0.2 and 0.3 logs as shown in the bottom pane of Figure 5.

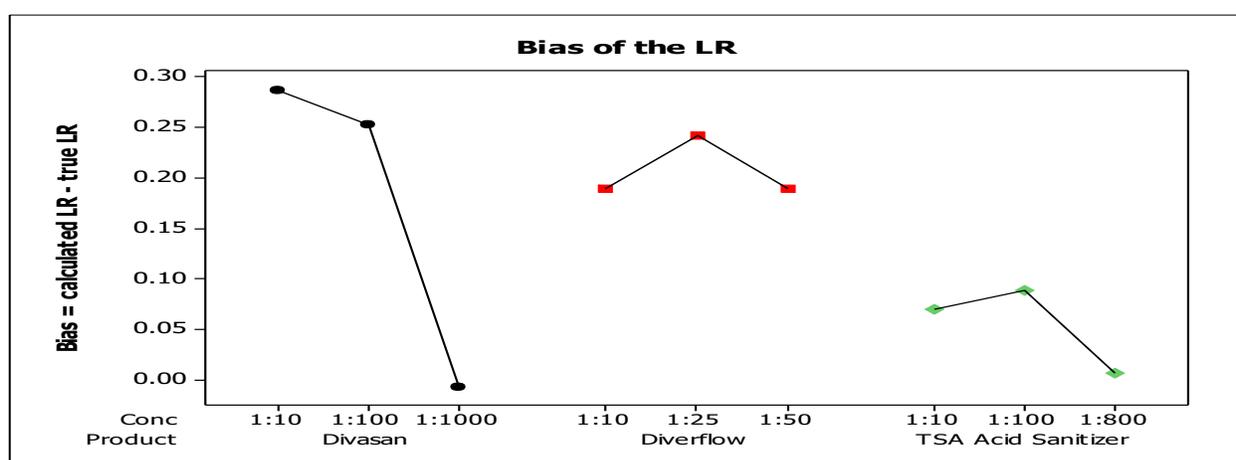
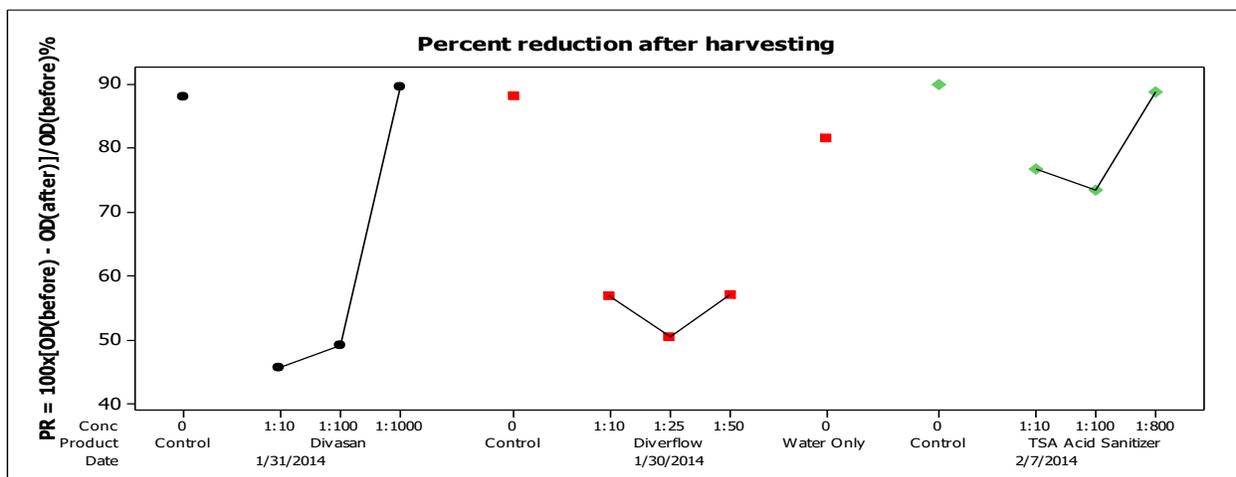


Figure 5. Top pane: Each point in the figure is the PR for a treatment in a single lab study of three different antimicrobials. Bottom pane: Each point in the figure is the harvesting bias for the LR in a single lab study of three different chemicals.

Discussion

It is desirable that efficacy testing in a laboratory be performed on a biofilm that adequately emulates “real use” conditions so as to predict how the antimicrobial will perform *in situ*. Although the intent is to develop a standard antimicrobial test method that represents the environment where an antimicrobial will be applied as closely as possible, operational factors are often either included or removed that simplify the method, because it is important that a standard method be conducted in most microbiology laboratories

stocked with typical equipment. Conversely, it is perfectly acceptable for a research method to require highly skilled researchers using specialized equipment. Although standard methods have key differences from methods used for research, it is desirable that a method being considered for standardization has a strong publication record over several years. For example, the Calgary device (also known as the minimum biofilm eradication concentration (MBEC) device), that grows biofilms under low shear in a batch system, was already a well-used research method (Ceri et al. 1999) before it went through the standardization process (Parker et al. 2014, ASTM 2017). This is an important example because it is the only other biofilm efficacy test that has been standardized and so it is informative to compare the STM results with those for the MBEC.

Both research and standard methods must possess good repeatability in a single laboratory. To assess repeatability, the method is studied by multiple independent experiments in a single lab until the response of interest, such as the LD or LR of biofilm bacteria, is achieved with an acceptable degree of precision. Unlike research methods, however, a standard method goes through an additional validation step to determine how reproducible the method is when the same antimicrobial is tested across multiple laboratories (Hamilton et al. 2013, ASTM 2016, AOAC 2016).

After the pre-requisite ruggedness and/or multi-laboratory data are generated, the review process for a method going through standardization is different than the traditional peer review process that a research method undergoes. Standard setting organizations such as ASTM, AOAC, and ISO have committees that are organized by subject area with members of the committee originating from industry, academia and/or regulatory agencies. Committee members can review, comment and vote on the method before the method is accepted as a standard. After standardization, these standard setting organizations typically have a process for reviewing and updating the method based upon advances in science and technology. Upon review, some methods require no updates, others are updated and reapproved, some are archived as

no longer useful or relevant, and some are replaced by a new method that better represents the current state of the art.

In other words, standard methods are dynamic procedures that have a life cycle that reflects changes in the scientific, public health and regulatory landscape over time. Standard methods that are referenced in regulatory guidelines tend to have a long life cycle. Making changes to these methods can be challenging because companies have products with registrations based upon meeting performance standards using the standard method (Tomasino, Parker, and Hamilton 2014, Tomasino et al. 2012).

The multi-laboratory study of the STM indicated excellent resemblance of the control data generated by the CDC reactor across multiple laboratories and also across multiple experiments in a single lab. The remarkable reproducibility of the STM controls (reproducibility SD = 0.2442, Figure S4) is a consequence of the considerable testing and optimization of the CDC reactor growth protocol (Goeres et al. 2005, ASTM 2007).

Unfortunately, the variability of the LRs in the multi-lab study was much higher. Parker, Hamilton, and Goeres (2018) found that the LRs generated by the STM (Figure 2) had about the same level of reproducibility as did the MBEC, but that both these biofilm methods had less reproducibility compared to dried surface and sporicide tests. Because of the desire to use the STM as a regulatory tool, we decided to investigate the possible reasons for this variability and see if the precision and potential bias of the STM could be improved.

We hypothesized that differences in sonication across labs may have contributed to the large level of variability in the LRs that were observed in the multi-laboratory study (Figure 2 and Table 2). The labs were retrospectively surveyed about their sonicator settings, which suggested that sonication power, sonication time, sonicator bath placement and whether or not the sonicator bath was degassed may have

substantial effects on the LR outcome. We performed the ruggedness testing using NaOCl where these operational factors for sonication were purposely varied over a range of values that we thought accurately reflected how these factors might be varied from the SOP when performed in different laboratories. The purpose was to determine how small changes to the values of these factors specified in the protocol can change the LR. The results failed to provide strong evidence (Figure 3 and Table 3) that any of the sonication parameters were the culprit for the large STM variability in the LRs. Interestingly, a similar conclusion was found when ruggedness testing the MBEC (Parker et al. 2014).

For any method, the list of operational factors to ruggedness test is limitless. For general standard methods, including manufacturing and chemical assays, ASTM recommends testing sampling techniques, surface conditioning, temperature and humidity (ASTM 2018a). In the case of standard methods that test antimicrobial efficacy against biofilms, the operational factors are tied to the 4 discrete steps of Grow, Treat, Sample and Analyze. In addition to the sonication factors that we ruggedness tested for the STM, other factors that have been ruggedness tested and reported in the literature include: nutrient concentration, flow rate, shear stress, incubation time and temperature (*i.e.*, the time and temperature allowed for initial surface colonization), and treatment temperature (Goeres et al. 2005, Parker et al. 2014). Which factors to include in a future ruggedness test ultimately depends on previous data and experience gained throughout the method development process.

Demonstrable ruggedness and reproducibility explain why regulators use standard methods for testing the efficacy of antimicrobials. Although simplifications and concessions are made in the development of a standard method, the known ruggedness and reproducibility of the method enable regulatory agencies to make better decisions regarding which products are efficacious enough to enter the marketplace. Quantifiable ruggedness indicates how similar the product's performance will be when tested in more relevant field conditions (*e.g.*, when treatment temperature is varied). Quantifiable reproducibility predicts the variability in a product's efficacy when tested in different labs on different days. This gives decision

makers a high degree of statistical confidence that products that are not efficacious will fail tests using the standard method and a high degree of statistical power that products that are efficacious will pass the tests (Parker, Hamilton, and Tomasino 2014, Parker, Hamilton, and Goeres 2018).

Industry stakeholders identified the splashing of detached biofilm clumps onto the inside of the tube when the coupon was dropped as a potential source of method bias. This led to the hypothesis that splashed biofilm deposits might contribute to the variability in the STM. To test this hypothesis, a simple swabbing experiment was completed that found large numbers of viable bacteria on the inside wall as a result of dropping the coupon into the tube. Hence, a more rigorous experiment was conducted (Figure 4) that found that when dropping coupons, the LR can be biased downwards by as much as 0.69 when using antimicrobials that kill everything for a coupon up to the limit of detection.

A separate study was performed to evaluate the potential for harvesting bias of the STM. Harvesting bias occurs when the antimicrobial being tested either impedes or enhances harvesting of the microbes subjected to treatment compared to the harvesting of microbes not treated (*i.e.*, the controls). This type of study is not often performed, and less often reported in the literature, although it is a known issue for any antimicrobial test method (Hamilton, Buckingham-Meyer, and Goeres 2009). Our results found that LRs for treatments known to impede harvesting had LRs that were larger by up to 0.3 (Figure 5). Such bias could also occur when applying fixative-containing antimicrobials such as glutaraldehyde, not tested here, a protein crosslinker that may adhere biofilm more tightly on coupons and result in less efficient harvesting.

Conclusion

Although standardized biofilm growth protocols have been around for decades (ASTM 2002a, Goeres, Loetterle, and Hamilton 2007, ASTM 2007, 2008), standardized biofilm efficacy methods are relatively recent. The MBEC was standardized in 2011. Here we outline the early life cycle of the STM that led to its initial standardization in 2013 (ASTM 2013a). The subsequent ruggedness and bias testing reported

here led to two substantial modifications of the STM: the development of the splashguard and the use of microscopy to assess bias if differential harvesting is suspected between the control biofilms and those biofilms treated with an antimicrobial. The STM with these modifications (E2871-19 (ASTM 2019)) is the current regulatory tool for registering anti-biofilm products and substantiating anti-biofilm claims on hard non-porous surfaces in the US (EPA 2016b). Parker, Hamilton, and Goeres (2018) considered the variability of the regulatory version of the STM (ATMP-MB-20) and showed that it is sufficiently reproducible when testing highly efficacious antimicrobials. The current version of the STM enables public health officials, manufacturers of antimicrobials and regulatory agencies to make informed and statistically confident decisions about antimicrobial efficacy against biofilms.

REFERENCES

- AOAC. 2016. "Guidelines for Collaborative Study Procedures to Validate Characteristics of a Method of Analysis." In *Official Methods of Analysis of AOAC International*.
- ASTM. 2002a. "Standard Test Method for Quantification of *Pseudomonas aeruginosa* Biofilm Grown with Medium Shear and Continuous Flow Using Rotating Disk Reactor." *ASTM International* E2196.

- ASTM. 2002b. "Standard Test Methods for Evaluation of Inactivators of Antimicrobial Agents." *ASTM International* E1054.
- ASTM. 2007. "Standard Test Method for Quantification of *Pseudomonas aeruginosa* Biofilm Grown with High Shear and Continuous Flow using CDC Biofilm Reactor." *ASTM International* E2562.
- ASTM. 2008. "Standard Test Method for Quantification of a *Pseudomonas aeruginosa* Biofilm Grown Using a Drip Flow Biofilm Reactor with Low Shear and Continuous Flow." *ASTM International* E2647.
- ASTM. 2013a. "Interlaboratory Study to Establish Precision Statements for ASTM E2871: Standard Test Method for Evaluating Disinfectant Efficacy against *Pseudomonas aeruginosa* Biofilm Grown in CDC Biofilm Reactor using the Single Tube Method." *ASTM International* RR E35-1008.
- ASTM. 2013b. "Standard Test Method for Evaluating Disinfectant Efficacy Against *Pseudomonas aeruginosa* Biofilm Grown in CDC Biofilm Reactor Using Single Tube Method." *ASTM International* E2871.
- ASTM. 2016. "Standard practice for conducting an interlaboratory study to determine the precision of a test method." *ASTM International* E691.
- ASTM. 2017. "Standard Test Method for Testing Disinfectant Efficacy against *Pseudomonas aeruginosa* Biofilm using the MBEC Assay." *ASTM International* E2799.
- ASTM. 2018a. "Standard Practice for Conducting Ruggedness Tests." *ASTM International* E1169.
- ASTM. 2018b. "Standard Practice for Preparing a *Pseudomonas aeruginosa* or *Staphylococcus aureus* Biofilm using the CDC Biofilm Reactor." *ASTM International* E3161.
- ASTM. 2019. "Standard Test Method for Evaluating Disinfectant Efficacy Against *Pseudomonas aeruginosa* Biofilm Grown in CDC Biofilm Reactor Using Single Tube Method." *ASTM International* E2871.
- Behnke, S., A. E. Parker, D. Woodall, and A. K. Camper. 2011. "Comparing the chlorine disinfection of detached biofilm clusters with those of sessile biofilms and planktonic cells in single- and dual-species cultures." *Appl Environ Microbiol* 77 (20):7176-84. doi: 10.1128/AEM.05514-11.
- Buckingham-Meyer, K., D. M. Goeres, and M. A. Hamilton. 2007. "Comparative evaluation of biofilm disinfectant efficacy tests." *J Microbiol Methods* 70 (2):236-44. doi: 10.1016/j.mimet.2007.04.010.
- Ceri, H., M. E. Olson, C. Stremick, R. R. Read, D. Morck, and A. Buret. 1999. "The Calgary Biofilm Device: new technology for rapid determination of antibiotic susceptibilities of bacterial biofilms." *Journal of clinical microbiology* 37:1771–1776.
- Davies, D. 2003. "Understanding biofilm resistance to antibacterial agents." *Nat Rev Drug Discov* 2 (2):114-22. doi: 10.1038/nrd1008.
- EPA. 2016a. "Method and Proposed Guidance to Assess the Efficacy of Antimicrobial Pesticide Products Intended to Control Public Health Biofilms." <https://www.regulations.gov/docket?D=EPA-HQ-OPP-2016-0357>.
- EPA. 2016b. "Single Tube Method with Splashguard for Evaluating Disinfectant Activity against *Pseudomonas* Biofilm – 2015 Method Performance Study," accessed 2016/04/21. <https://www.regulations.gov/document?D=EPA-HQ-OPP-2016-0357-0002>.
- Fernandez-Rivero, M. E., J. L. Del Pozo, A. Valentin, A. M. de Diego, J. Peman, and E. Canton. 2017. "Activity of Amphotericin B and Anidulafungin Combined with Rifampicin,

- Clarithromycin, Ethylenediaminetetraacetic Acid, N-Acetylcysteine, and Farnesol against *Candida tropicalis* Biofilms." *J Fungi (Basel)* 3 (1). doi: 10.3390/jof3010016.
- 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 (Pt 3):757-62. doi: 10.1099/mic.0.27709-0.
- Goeres, Darla M, Martin A Hamilton, Nicholas A Beck, Kelli Buckingham-Meyer, Jackie D Hilyard, Linda R Loetterle, Lindsey A Lorenz, Diane K Walker, and Philip S Stewart. 2009. "A method for growing a biofilm under low shear at the air-liquid interface using the drip flow biofilm reactor." *Nature protocols* 4 (5):783.
- Goeres, Darla M, Linda R Loetterle, and Martin A Hamilton. 2007. "A laboratory hot tub model for disinfectant efficacy evaluation." *Journal of microbiological methods* 68 (1):184-192.
- Hamilton, Martin A., Kelli Buckingham-Meyer, and Darla M. Goeres. 2009. "Checking the validity of the harvesting and disaggregating steps in laboratory tests of surface disinfectants." *Journal of AOAC International* 92:1755–1762.
- Hamilton, Martin A., Gordon Cord Hamilton, Darla M. Goeres, and Albert E. Parker. 2013. "Guidelines for the Statistical Analysis of a Collaborative Study of a Laboratory Method for Testing Disinfectant Product Performance." *Journal of AOAC International* 96 (5):1138-1151. doi: 10.5740/jaoacint.12-217.
- Lewandowski, Zbigniew, and Haluk Beyenal. 2013. *Fundamentals of Biofilm Research, Second Edition*. 2 edition ed. Boca Raton: CRC Press.
- Lineback, C. B., C. A. Nkemngong, S. T. Wu, X. Li, P. J. Teska, and H. F. Oliver. 2018. "Hydrogen peroxide and sodium hypochlorite disinfectants are more effective against *Staphylococcus aureus* and *Pseudomonas aeruginosa* biofilms than quaternary ammonium compounds." *Antimicrob Resist Infect Control* 7:154. doi: 10.1186/s13756-018-0447-5.
- Mah, Thien-Fah C., and George A. O'Toole. 2001. "Mechanisms of biofilm resistance to antimicrobial agents." *Trends in Microbiology* 9 (1):34-39. doi: [https://doi.org/10.1016/S0966-842X\(00\)01913-2](https://doi.org/10.1016/S0966-842X(00)01913-2).
- Manner, S., D. M. Goeres, M. Skogman, P. Vuorela, and A. Fallarero. 2017. "Prevention of *Staphylococcus aureus* biofilm formation by antibiotics in 96-Microtiter Well Plates and Drip Flow Reactors: critical factors influencing outcomes." *Sci Rep* 7:43854. doi: 10.1038/srep43854.
- Parker, A. E., M. A. Hamilton, and S. Tomasino. 2014. "A Statistical Model for Assessing Performance Standards for Quantitative and Semi-quantitative Disinfectant Test Methods." *JAOAC International* 97 (1):58-67.
- Parker, A. E., D. K. Walker, D. M. Goeres, N. Allan, M. E. Olson, and A. Omar. 2014. "Ruggedness and reproducibility of the MBEC biofilm disinfectant efficacy test." *J Microbiol Methods* 102:55-64. doi: 10.1016/j.mimet.2014.04.013.
- Parker, Albert E., and Martin A. Hamilton. 2011. "KSA-SM-10: Assessing resemblance, repeatability, and reproducibility for quantitative methods." Center for Biofilm Engineering at Montana State University, Bozeman, MT. <http://www.biofilm.montana.edu/documents-reports/sharing-articles.html>.
- Parker, Albert E., Martin A. Hamilton, and Darla M. Goeres. 2018. "Reproducibility of antimicrobial test methods." *Scientific Reports* 8 (1):12531. doi: 10.1038/s41598-018-30282-3.

- Sauer, Karin, Anne K. Camper, Garth D. Ehrlich, J. William Costerton, and David G. Davies. 2002. "Pseudomonas aeruginosa Displays Multiple Phenotypes during Development as a Biofilm." *Journal of Bacteriology* 184 (4):1140.
- Stewart, P. S., and A. E. Parker. 2019. "Measuring Antimicrobial Efficacy against Biofilms: A Meta-Analysis." *Antimicrob Agents Chemother*. doi: 10.1128/AAC.00020-19.
- Stewart, Philip S. 2015. "Antimicrobial tolerance in biofilms." In *Microbial Biofilms*, 269-285. Washington, DC., USA: ASM Press.
- Tilt, Nicola, and Martin A. Hamilton. 1999. "Repeatability and reproducibility of germicide tests: a literature review." *Journal of AOAC International* 82:384-389.
- Tomasino, S. F. 2013. "Development and assessment of disinfectant efficacy test methods for regulatory purposes." *Am J Infect Control* 41 (5 Suppl):S72-6. doi: 10.1016/j.ajic.2012.11.007.
- Tomasino, Stephen F, Albert E Parker, Martin A Hamilton, and Gordon C Hamilton. 2012. "Performance of the AOAC use-dilution method with targeted modifications: collaborative study." *Journal of AOAC International* 95 (6):1618-1628.
- Tomasino, Stephen F., Albert E. Parker, and Martin A. Hamilton. 2014. "Use of Statistical Modeling to Reassess the Performance Standard for the AOAC Use-Dilution Methods (955.15 and 964.02)." *Journal of AOAC International* 97 (1):68-77. doi: 10.5740/jaoacint.13-216.
- Yoon, H. Y., and S. Y. Lee. 2017. "Establishing a laboratory model of dental unit waterlines bacterial biofilms using a CDC biofilm reactor." *Biofouling* 33 (10):917-926. doi: 10.1080/08927014.2017.1391950.

SUPPLEMENTARY MATERIAL

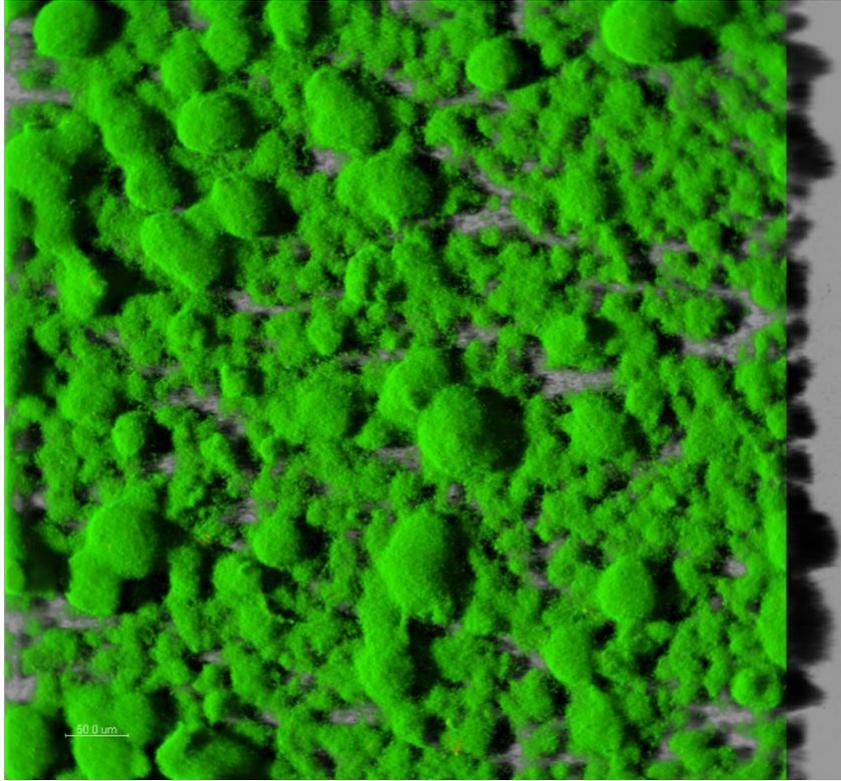


Figure S1. A 620 μm x 620 μm confocal microscope image of a *Pseudomonas aeruginosa* biofilm grown using the STM. The biofilm was stained with LIVE/DEAD BacLight Bacterial Viability Kit stain (Invitrogen #L7012, Carlsbad, CA) for 20 minutes, rinsed, and then imaged in a fully-hydrated state using extra-long working distance water immersion objectives.

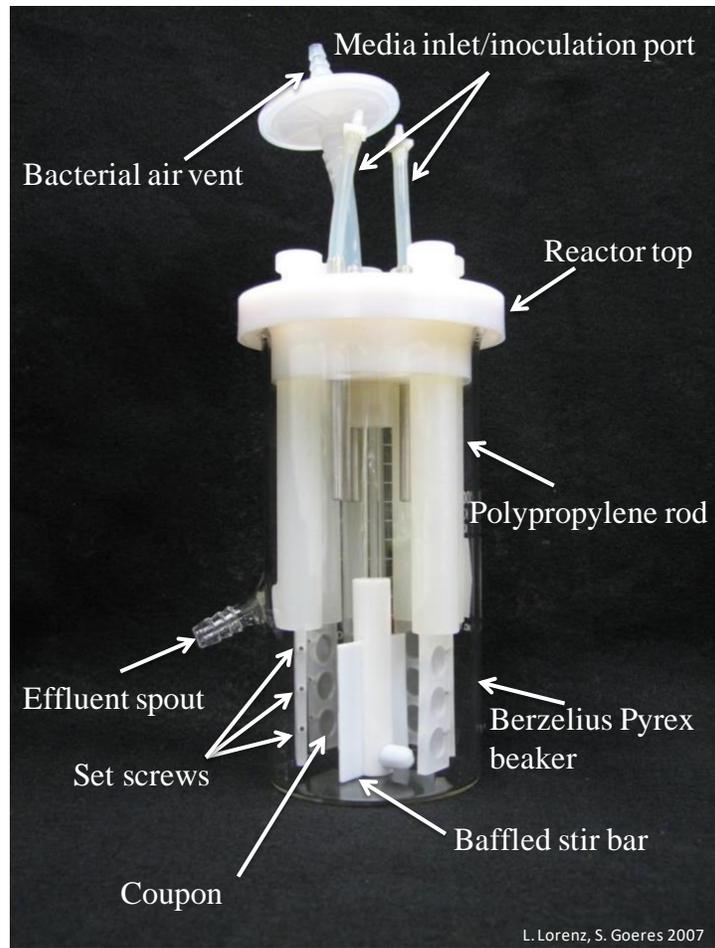
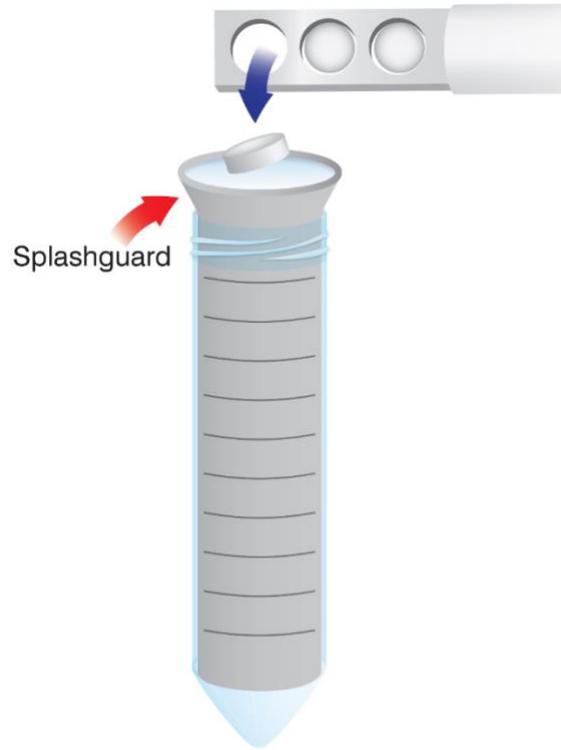


Figure S2. Photograph of an assembled CDC Biofilm Reactor with associated description of parts, including rods housing coupons of known surface area for biofilm growth.



© 2016, MSU Center for Biofilm Engineering, J Parchen.

Figure S3. Diagram of splashguard insert in 50 mL conical tube that alleviates splash bias.

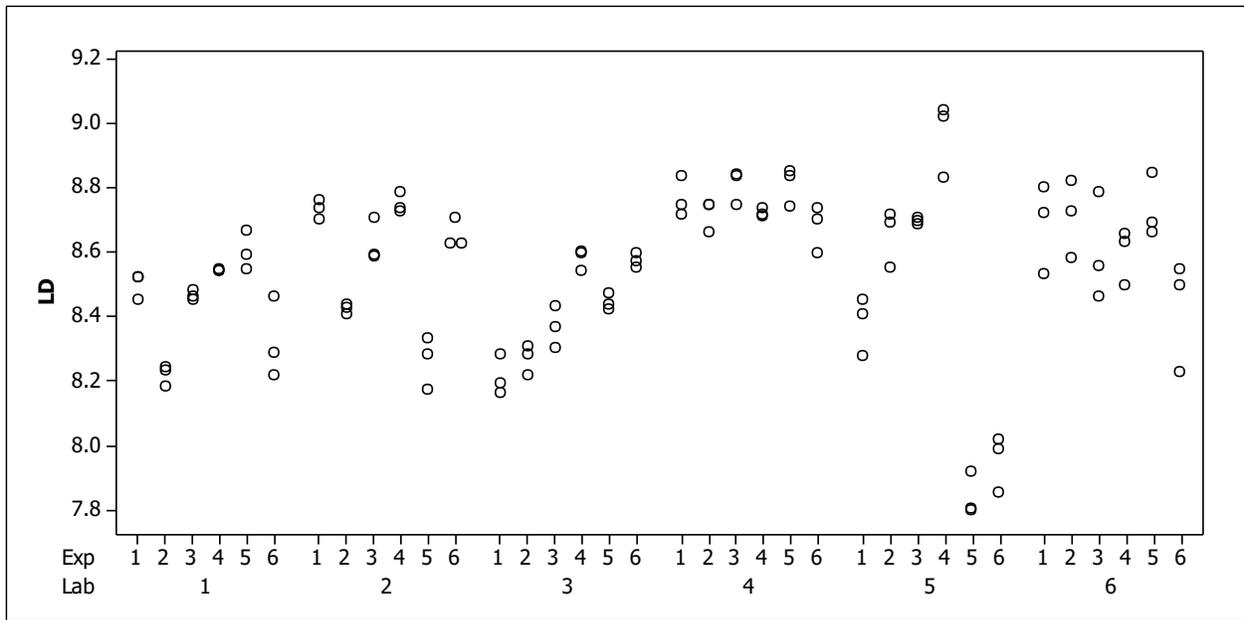


Figure S4. Untreated control data from the 6-lab study. Each point in the graph is the LD = $\log(\text{CFU}/\text{cm}^2)$ of biofilm bacteria grown on a single CDC coupon then subjected to buffered water for the specified contact time and then the neutralizer (2x D/E broth). Along the horizontal axis are listed numerical identifiers for the six labs (1-6), and also numerical identifiers for the 6 experimental days at each lab.

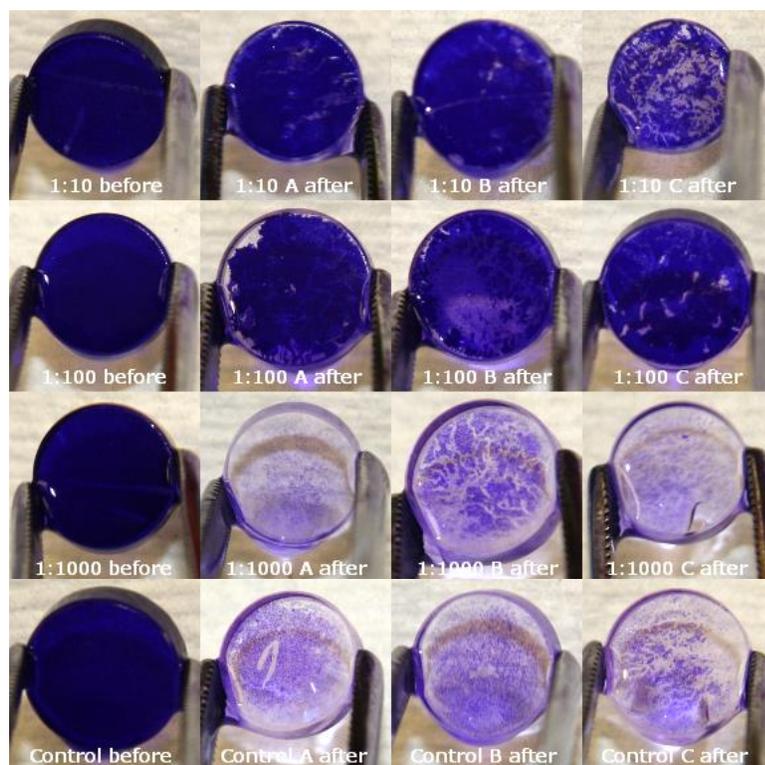


Figure S5. Assessing the potential for bias that occurs when the percent of bacteria harvested from the control coupons is different than the percent of bacteria harvested from the treated coupons. Shown here are the results of a crystal violet assay of controls and a Divosan MH treatment at 3 concentrations (1:10, 1:100 and 1:1000) before sonication/vortexing; after sonication/vortexing (3 replicates: A, B, C).