Interlaboratory Evaluations of a Standardized Quantitative Test Method for Determining the Bactericidal and Tuberculocidal Efficacy of Antimicrobial Substances on Hard Non-Porous Surfaces

Stephen F. Tomasino, Rebecca M. Pines, Darla M. Goeres, Albert E. Parker

Interlaboratory Evaluations of a Standardized Quantitative Test Method for Determining the Bactericidal and Tuberculocidal Efficacy of Antimicrobial Substances on Hard Non-Porous Surfaces

Tomasino, S., Pines, R., Goeres, D. and Parker, A.

Abstract

The development, validation, and use of new quantitative methodologies for testing the effectiveness of antimicrobial products are necessary to meet the regulatory challenges associated with an ever-changing marketplace, novel product claims, new infection control practices, and the emergence of new clinical pathogens. A series of four interlaboratory studies were conducted in a standardized manner on an interim quantitative method for testing liquid treatments against bacteria to assess its statistical performance. The Quantitative Method, a derivative of ASTM E2197, is designed to enumerate the number of viable microbes remaining on a test carrier following exposure to a liquid antimicrobial treatment; a $\log_{10}$ reduction (LR) in viable bacteria is calculated based on the difference between the mean $\log_{10}$ density values of the untreated control and treated carriers. The Quantitative Method uses 1 cm diameter disks (carriers) of brushed stainless steel as the material to represent a hard, non-porous surface. The LR value is used as the measure of product effectiveness, where higher LR values are indicative of greater microbial kill. The test microbes were *Staphylococcus aureus*, *Pseudomonas aeruginosa*, and *Mycobacterium terrae*. The liquid antimicrobial treatments used in these studies were highly relevant to those in the marketplace and provided a wide range of mean LR outcomes. The focus of the statistical assessment was on the repeatability of the LRs across experiments within a lab (S_r) and the reproducibility of the LRs across labs (S_R). Due to the

additional sources of variability, the S_R is expected to be higher than the variability within a laboratory (S_r); this was observed in the studies reported here. Across the studies, the S_r values for LR were small (i.e., less than 0.84), most notably for treatments generating high mean LRs (5 or above) where the S_r was as small as 0.12. Overall, the S_R values ranged from 0.227 to 1.217. Only three of the twenty-four treatment combinations over the study period resulted in S_R values above 1.0 – the associated LRs for the three treatments ranged from 2.22 to 3.26. Antimicrobial treatments with a LR of 4.5 or higher exhibited S_R of 0.561 or less. The statistical attributes reported here for the draft Quantitative Method when used to test *P. aeruginosa, S. aureus,* and *M. terrae* provide information for decision makers when considering the method as a candidate regulatory procedure. The data and statistical analyses contained in this report are historical in nature and provide useful baseline information for individuals conducting additional technical review of the method. Based on the data, the Quantitative Method displays a statistical profile consistent with other standard methods approved by standard-setting organizations where method performance data are available.

1. Introduction

Cleaning and disinfection of environmental surfaces are important practices for reducing the spread of public health pathogens in a wide range of clinical, household, institutional, veterinary, commercial, and food production settings (Abreu et al., 2013; AHRQ, 2015; Alfa et al., 2015; Boyce, 2016; Dancer, 2014; Donskey, 2013; Hota, 2004; Quinn et al., 2015; Rutala et al., 2013). To ensure the performance of antimicrobial products in the field prior to sale and distribution, regulatory authorities have established policies to document the effectiveness of antimicrobial

products according to their labeling. The availability of validated standardized laboratory methods is essential to the regulatory process. Throughout the world, a variety of different methods are recognized for the generation of efficacy data for antimicrobial products; however, to date, there is no single method or platform for assessing product efficacy that can be used and accepted by all countries. For example, in the United States (US), the regulatory oversight for antimicrobial products is the responsibility of the Environmental Protection Agency's (EPA) Office of Chemical Safety and Pollution Prevention (OCSPP), Office of Pesticide Programs (OPP). The statutory authority for the registration of antimicrobial products falls under the Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA). In the US, antimicrobial pesticides bearing claims for control of microorganisms on inanimate surfaces which are infectious to humans are considered directly related to human health – these are known as public health antimicrobial products. Under FIFRA, the registrant of an antimicrobial product with a public health claim is required to submit efficacy data to EPA in support of the product's registration. Efficacy test guidelines, Series 810.2000–Product Performance Test Guidelines: OCSPP 810.2200 Disinfectants for Use on Hard Surfaces-Efficacy Data Recommendations have been established by the OCSPP to inform the manufacturer which test methodology is appropriate to support a specific efficacy claim (US EPA OCSPP, 2018). The OPP Antimicrobials Division (AD) reviews the product efficacy data and if the data meet the efficacy standards and all other requirements for registration are met, the product is granted a registration and is issued an EPA registration number. An overview of EPA's regulatory process for antimicrobial pesticides has been published (Tomasino et al., 2021)

The development, validation and use of new quantitative methodologies for testing the effectiveness of antimicrobial products are necessary to meet the regulatory challenges associated

with an ever-changing marketplace, novel product claims, new infection control practices, and the emergence of new clinical pathogens. As part of this effort, the US and other OECD (Organization of Economic Co-operation and Development) member countries initiated an evaluation of a set antimicrobial test methods for the purpose of international harmonization of test methods for measuring the performance of antimicrobial products. The OECD is an intergovernmental organization in which representatives of 34 industrialized countries in North and South America, Europe and the Asia and Pacific region, as well as the European Commission, participate collaboratively in response to international problems (OECD, 2018). Most of the OECD's work is facilitated by more than 200 specialized committees and working groups composed of member country delegates. In 2002, an international workshop of the US and OECD member countries identified scientific and regulatory areas in need of harmonization for testing antimicrobial products. From the 2002 workshop, a primary goal of selecting and validating a quantitative test method to assess the activity of antimicrobial products for use on hard non-porous surfaces was proposed. The methodology down selected for consideration is derived from an existing standard method, ASTM E2197, the Standard Quantitative Disk Carrier Test Method (ASTM E2197, 2017). The technical attributes of ASTM E2197 have been previously summarized (Springthorpe, 2005) and investigated in two previous multi-laboratory studies (Parker et al. 2018). For the purpose of this paper, the method will be referred to as the Quantitative Method. The Quantitative Method retains the primary attributes of ASTM E2197; however, microbe-specific modifications and additional areas of optimization were applied during our studies.

A summary of EPA's approach to method development and standardization has been published (Tomasino, 2014) and the activities described in this paper are highly relevant to that process. For

background, the process for the development of OECD guidelines for testing chemicals is provided in an OECD Guidance Document (OECD, 2009). The OECD Test Guideline Program provides a structured consensus-based technical peer review process of a candidate method for an intended purpose, the data requirements, generation and review of method validation data, and a transparent approval process. In the early stages of the OECD-facilitated process, an *ad hoc* group of international experts under the auspices of the OECD Task Force on Biocides was tasked with overseeing the technical evaluation and direction of the project. Between 2008 and 2010, a series of international ring trials (interlaboratory studies) were conducted to validate the proposed method; the outcome of the ring trials was posted by the OECD (OECD, 2011). In 2011, the draft method and the validation report were reviewed by the OECD Working Group of National Coordinators (WNT) of the Test Guidelines Programme; the WNT is comprised of regulatory representatives from the various OECD member countries. Upon review, the WNT recommended several revisions to the method and additional data collection to further ascertain the performance of the method across laboratories. Main areas where the method was revised included the use of more stringent culture preparation practices, identification of control carrier count levels for a valid test, inclusion of a reference standard, and enhancements to the microbe recovery process. In response to the WNT's recommendations, four interlaboratory studies were performed in the US from 2011 through 2013 to address proposed revisions and to further assess the statistical performance of the revised method. The interlaboratory studies discussed in this report involve tests of *Staphylococcus aureus*, *Pseudomonas aeruginosa*, and *Mycobacterium terrae*.

2. Methods

115 *2.1 Quantitative Method Overview*

116 The Quantitative Method uses 1 cm diameter disks (carriers) of brushed stainless steel as the
117 material to represent a hard, non-porous surface. Each disk receives 10 µL of microbial inoculum
118 (with a three-part organic and inorganic soil load) deposited in the center of each carrier. The
119 inoculum is dried and exposed to 50 µL of the liquid antimicrobial treatment being tested; control
120 carriers receive an equivalent volume of an innocuous fluid (e.g., phosphate buffered saline). The
121 exposure time (e.g., 5 min) is allowed to elapse; 10 mL neutralizer is then added to the vial to halt
122 the antimicrobial action. Each vial with the carrier is vortexed, serially diluted, and then filtered or
123 direct plated to recover viable microorganisms. Based on the difference between the mean $\log_{10}$
124 density values of the untreated control and treated carriers, a $\log_{10}$ reduction (LR) in viable bacteria
125 is calculated. The LR value is used as the measure of product effectiveness, where higher LR
126 values are indicative of greater microbial kill.

127

128 *2.2 Frozen stock cultures*

129

130 Frozen stock cultures of *P. aeruginosa* (ATCC No. 15442) and *S. aureus* (ATCC No. 6538) were
131 prepared by rehydrating lyophilized cultures with approximately 1 mL tryptic soy broth (TSB) and
132 transferring the entire rehydrated mixture into 5 mL of TSB, mixing thoroughly and incubating at
133 36±1ºC for 24±2 h. The growth was streak isolated onto tryptic soy agar (TSA), incubated at
134 36±1ºC for 24±2 h, and several isolated colonies were collected and re-suspended in 1 mL of TSB.
135 Golden/yellow colonies were selected for *S. aureus*. Each of three colonial phenotypes was
136 selected for *P. aeruginosa*. The suspension (0.1 mL) was spread plated on 6-10 plates of TSA and
137 incubated for 18-24 h at 36±1°C. Following the incubation, 5 mL sterile cryoprotectant (TSB with

15% v/v glycerol) were added to the surface of each plate. The growth was re-suspended in cryoprotectant using a sterile spreader. The suspension from each plate was removed with a pipette, pooled in a sterile vessel, and thoroughly mixed. Small aliquots of the harvested suspension were transferred into cryovials and stored at -70ºC.

Frozen stock cultures of *M. terrae* (ATCC No. 15755) were prepared by adding 1.0 mL of Middlebrook 7H9 Broth with 10% (v/v) ADC enrichment (MADC) to a lyophilized culture and transferring the entire rehydrated pellet back into 5 mL of MADC. After mixing, the suspension (0.1 mL) was spread plated onto 6-10 plates of M7H9 or M7H11 agar and incubated for 20-22 days at 36±1°C. At the end of the incubation period, 5 mL of MADC was added to the surface of each plate and re-suspended using a sterile spreader to gently scrape the agar surface. The suspension was removed with a pipette, pooled, and mixed. While mixing continuously, small aliquots of the harvested suspension were placed into separate cryovials and stored at -70°C. The titer of the frozen stock was approximately 1×10^9 CFU/mL.

2.3 Test cultures

For *S. aureus*, TSB was used to grow the test cultures. For *P. aeruginosa*, synthetic broth (SB) was used for the generation of the test cultures in Study one; however, due to the inconsistent commercial availability of SB, TSB was used in Study two. Test cultures were prepared by defrosting a frozen stock culture and adding 0.1 mL of the culture to 10 mL of broth (SB or TSB per the study) vortex-mixed, and incubated for 18-24 h at 36±1°C.

For *P. aeruginosa* cultures, visible pellicle was separated from the broth prior to harvest. The culture was discarded if the pellicle was disrupted and showed fragments in the broth. Using a

pipette, the remaining broth culture was withdrawn, avoiding any sediment on the bottom of the tube, and transferred into a 15 mL conical tube. For *S. aureus,* the cultures were briefly vortexed and transferred to a 15 mL conical tube. The culture was centrifuged at ~5,000 gN for 20±5 min and the supernatant was removed without disrupting the pellet. The pellet was re-suspended in a maximum of 10 mL PBS. Further dilution was allowed prior to the addition of the soil load to meet the prescribed control carrier count levels. This culture was used to prepare the final test suspension which included the soil load – see sections 2.4 and 2.5.

For *M. terrae*, the test culture was initially (Study three) grown statically (without agitation) by adding 1 mL thawed stock culture to a flask of 100 mL MADC and incubating at 36±1°C for 20-22 days. For Study four, 1 mL of thawed frozen stock culture was added to a flask of 100 mL MADC, sealed with foil and incubated at 36±1°C for 7-10 days under agitation at 150 rpm. Following incubation, 25 mL portions of the MADC broth culture were aliquoted into each of 2-50 mL conical tubes and centrifuged at ~10,000 gN for 20±5 min. The supernatant was removed, and the pellet was re-suspended in 25 mL sterile de-ionized (DI) water and centrifuged a second time. The pellets were re-suspended in a total of 5 mL sterile DI water, placed in a sterile glass tissue grinder, and homogenized. The culture was transferred to a small bottle (e.g., bijou bottle) with glass beads, capped, and vortexed for 5 min. The concentration of the final culture was adjusted with sterile DI water to achieve control carrier counts within the prescribed range for each study.

2.4 Three-part soil load

Prior to carrier inoculation, a soil load, with organic and inorganic components, was incorporated in the test suspension using the following three stocks: a) bovine serum albumin (BSA) prepared by adding 0.5 g BSA to 10 mL of PBS, mixed and filter sterilized, b) yeast extract prepared by adding 0.5 g yeast extract to 10 mL of PBS, mixed, and filter sterilized, and c) mucin prepared by adding 0.04 g mucin to 10 mL of PBS, mixed and sterilized by autoclaving (15 min at 121°C). Each stock was stored at -20±2°C.

2.5 Final test suspension

For each test day, a minimum of 500 µL of the final test suspension was prepared by combining the following components: 25 µL BSA stock, 35 µL yeast extract stock, 100 µL mucin stock, and 340 µL of the test suspension. The final test suspension (held at 22±2°C), was used to inoculate carriers within 30 min of preparation.

2.6 Treatment diluent – 375 ppm hard water

For the preparation of antimicrobial treatments requiring dilution, the test substance was diluted in 375 ppm hard water. Solution A was prepared by dissolving 19.84 g anhydrous $MgCl_2$ (or 42.36 g $MgCl \cdot 6H_2O$) and 46.24 g anhydrous $CaCl_2$ in DI water, and diluting to 1 L. Solution B was prepared by dissolving 35.02 g $NaHCO_3$ in water, and diluting to 1 L. Solutions A and B were filter sterilized and stored under refrigeration. Per 1 L of 375 ppm hard water, 6.0 mL of solution A and 8.0 mL of solution B were added to 600 mL DI water, mixed, and brought to a 1L volume with DI water. The target pH was 7.0±0.2. On the day of the test, the hardness of the water was verified using a water hardness test kit or other suitable titration method. The target hardness expressed as mg/L calcium carbonate ($CaCO_3$) was 375 mg/L +5/-10% (337-394 mg/L).

2.7 Carriers

One cm diameter disks made from 0.7 mm thick sheets of brushed and magnetized stainless steel (AISI #430) were utilized (e.g., Pegen Industries, part #430-107). Carriers were prepared (by the manufacturer) to achieve a number 4 finish (polished), soaked in a mild acid bath to remove any impurities and accumulated debris from the carrier surface, and tumbled with ceramic chips and a cleanser to remove the punching burrs from the edges of the disks. Prior to use, the brushed surface of the carriers was visually checked for abnormalities (e.g., rust, chipping, deep striations) and discarded if observed. Carriers were soaked in a detergent solution for 2-4 h to degrease them and then rinsed thoroughly in DI water. Approximately 20 clean dry carriers were placed on a piece of filter paper inside a glass Petri dish with the lid applied and sterilized. Sterile carriers were transferred topside up to sterile Petri dishes (plastic) without filter paper for inoculation. Carriers were single use only.

2.8 Inoculation and drying of carriers

Carriers were placed in Petri plates prior to inoculation. Immediately prior to carrier inoculation, the final test suspension was briefly vortexed and then re-vortexed periodically during the inoculation process. Depending on the study, 3-4 untreated control carriers and 3-4 treated carriers (per test substance treatment) were used. Ten microliters of the final test suspension were applied to the center of each carrier with a calibrated positive displacement pipette (with a 10 µL tip). Inoculated carriers where the final test suspension leaked over the edge were discarded. The Petri dish with the inoculated carriers was transferred to a desiccation unit, lid removed, unit sealed with vacuum applied and maintained at 0.068 to 0.085 MPa at 22±2°C for 60±10 min. Dried, inoculated carriers were held at room temperature and used for testing within one hour after drying.

2.9 Exposure of the inoculum to the test substance or PBS (control)

Using sterile forceps, each dried carrier was transferred (inoculated side up) into a flat-bottom vial and capped. In timed, staggered intervals, 50 µL of the test substance (equilibrated to 22±2°C) or control substance (PBS) was deposited with a micropipette over the dried inoculum on each test carrier, ensuring complete coverage. The pipette tip was not allowed to contact the carrier surface. The vials were left uncapped and test carriers held at 22±2°C for 5 min. Any carrier where the test substance leaked off was not used and was replaced with new carrier and vial.

2.10 Neutralization of test substance

At the end of the 5 min exposure period, 10 mL of neutralizer was added to each vial in the order treated, including controls. The neutralized solution with the carrier was deemed the 10^0 dilution for the quantification step. Each vial was briefly vortexed following the addition of the neutralizer. Following the addition of neutralizer across the entire set of carriers, each vial was vortexed again for 30 s at high speed to recover the surviving microorganisms. See section 2.12.6 for the neutralization confirmation protocol.

2.11 Dilution and recovery

Using the 10.05 mL neutralized suspension, serial 10-fold dilutions were initiated within 30 min of the addition of the neutralizer. For calculation purposes, a value of 10.00 mL was utilized for the neutralized suspension. The liquids were filtered using pre-wetted (with 10 mL PBS) 0.2 µm polyethersulfone (PES) membrane filters for *S. aureus* and *P. aeruginosa* and 0.45 µm PES membrane filters for *M. terrae*; filtration was initiated within 30 min of preparing the dilutions. Vacuum was applied to facilitate filtering. The entire contents of each treatment vial (with the

247 treated carrier) and dilution tube (e.g., 10^{-1}) were filtered. For the vial, a magnet was used to hold
248 the carrier in place while dispensing the liquid onto the membrane. The vial was rinsed with ~20
249 mL of PBS and briefly vortexed, while keeping magnet in place, poured into the same filter unit.
250 For dilution tubes, each tube was rinsed once with ~10 mL of PBS, vortex-mixed, and poured into
251 filter unit. With the vacuum on, the inside surface of each filtration unit was rinsed with an
252 additional ~40 mL PBS. The membrane filter was placed on the appropriate recovery medium;
253 TSA was used for *S. aureus* and *P. aeruginosa*, and M7H11 was used for *M. terrae*.

254 For the control carriers, contents of the vial were serially diluted and filtered (Study one only) or
255 direct plated (100 µL) in duplicate (Study two) on the appropriate recovery medium. For *S. aureus*
256 and *P. aeruginosa*, plates were incubated at 36±1°C for 48±4 h. For *M. terrae*, plates were
257 incubated at 36±1°C for 17-21 days.

258 The number of colonies was recorded at the end of the incubation period. Counts up to 200 and
259 300 were used in the calculations for filtered and direct plated aliquots, respectively. The colony
260 forming units (CFUs) per carrier were calculated by taking the weighted average across all
261 countable dilutions (Hamilton & Parker 2010) and then converted to log density by taking the $\log_{10}$
262 of the density (CFUs per carrier). The mean $\log_{10}$ densities across individual sets of treated and
263 control carriers were calculated. The $\log_{10}$ reduction (LR) for each test was calculated as the mean
264 $\log_{10}$ density for the control carriers in that test (*TestLD*) minus the mean $\log_{10}$ density for treated
265 carriers in that test.

266

267 *2.12 Interlaboratory evaluations*

268 *2.12.1 Basic study principles*

The overarching goal of the interlaboratory investigations was to determine the statistical profile of the draft Quantitative Method for testing liquid products on a hard non-porous surface, with targeted revisions, using a standardized study protocol. The conduct of these studies was consistent with EPA method validation guidelines (U.S. EPA, 2016), standard setting organizations such as AOAC International (AOAC International, 2016), ASTM International (ASTM standard E691, 2013), and recommendations provided by Hamilton et al., 2013. Each laboratory designated a technical team to conduct the study and was asked to practice the method prior to launching the study to gain proficiency. The technicians were asked to evaluate a set of liquid antimicrobial test substances in a blinded, randomized, and replicated fashion. Study documentation, including peer review, was consistent with Good Laboratory Practice Standards (US EPA, 1997).

The primary response variables were the $\log_{10}$-transformed carrier counts and the resulting LR outcomes as the measurement of treatment effectiveness. As an integral part of the test design, a set of test substances with the potential to provide a range of LRs was evaluated to simulate the challenges of the real-world uses. The treatments were evaluated in a blinded, randomized fashion. Three replications (separate test days) were conducted for each test substance.

The studies allowed for the assessment of four key method performance indicators: 1) consistent control carrier counts (resemblance), 2) the ability to produce repeatable LRs within the same lab (repeatability), 3) the ability to produce reproducible LRs among labs (reproducibility), and 4) the capacity of the method to differentiate the antimicrobial efficacy of two levels of a test substance with presumed high and low log reduction outcomes (responsiveness).

2.12.2 Standardized test conditions

Three test conditions were employed for the evaluation of each antimicrobial and reference

standard treatment: 1) incorporation of a three-part soil load into the final test suspension prior to carrier inoculation, 2) use of 375 ppm hard water as the treatment diluent (when necessary), and 3) an exposure time of 5 min.

2.12.3 Antimicrobial test substances – treatments

The test substances were derived from antimicrobial formulations with active ingredients relevant to the marketplace. The treatments were not evaluated to confirm the label claims of a product. The treatments were prescreened in advance by a single laboratory to presumptively provide a range of effectiveness (i.e., range of LRs). Each lab received a sub-sample of the same production batch of each test substance. Treatments were provided as "ready to use" or as dilutable concentrates. The neutralizer solutions were confirmed as effective in advance of testing per the methodology provided in section 2.12.6.

2.12.4 Sodium hypochlorite-based reference standard treatments

For the purpose of developing a reference standard and to measure method responsiveness, tests of certified laboratory grade sodium hypochlorite (NaOCl) were incorporated into the study design. Two concentrations of NaOCl were tested per microbe. Each laboratory was asked to prepare the test solutions and to verify the concentrations of available chlorine prior to conducting the tests using a Hach test kit, or comparable analysis. The two NaOCl treatments were tested side-by-side on the same test day with presumed high and low log reduction outcomes. The neutralizer solution was letheen broth with 0.1% (w/v) sodium thiosulfate ($Na_2S_2O_3$).

2.12.5 Control carrier counts

315 Quantification of the inoculum level on the untreated control carriers post-drying is an essential
316 component to the method. An acceptable range of mean log density of the control carriers (i.e.,
317 *TestLD*) was established for each study. The number of control carriers ranged from three to four
318 depending on the study. One set of control carriers was used to represent one or more treatments
319 on any given test day. Individual carrier counts or mean control carrier count values slightly
320 above or below the acceptable range were permissible upon discussion with the study director.

321 *2.12.6 Neutralization confirmation protocol*

322 The effectiveness of the neutralizer solution was determined in advance of each study using a
323 quantitative assay. In triplicate, the test substance (50 µL) and neutralizer (10 mL) were vortex-
324 mixed together for 10 s, 10 µL diluted final test suspension (20-200 CFU per vial) were added to
325 each mixed solution, and held for 10 min at room temperature. The entire contents of each vial
326 were filtered, and the test microbe enumerated. The titer of the final test suspension and the
327 toxicity of the neutralizer were determined for comparative purposes; 10 mL volumes of PBS
328 (for the titer) and the specified neutralizer were inoculated and enumerated per the procedure for
329 the neutralizer effectiveness component of the assay. To be deemed an acceptable neutralizer,
330 the recovered number of bacteria in the neutralizer effectiveness and neutralizer toxicity
331 treatments was at least 50% of the titer control (equivalent to a difference of 0.3 on the $\log_{10}$-
332 scale). Neutralizers utilized in these studies were letheen broth, letheen broth with 0.1% (w/v)
333 sodium thiosulfate, and PBS with 0.1% (v/v) polysorbate 80 and 0.1% (w/v) sodium thiosulfate.

334 *2.13 Interlaboratory studies involving S. aureus and P. aeruginosa*

335 *2.13.1 Study one*

Nine laboratories evaluated four antimicrobial treatments: phenolic (0.074%), hydrogen peroxide (0.5%), citric acid (0.6%), and quaternary ammonium compound (0.066%), and two NaOCl reference standard treatments. NaOCl concentrations of 2000 ppm and 500 ppm were tested against *S. aureus.* For *P. aeruginosa*, 2000 ppm and 250 ppm NaOCl were tested. Letheen broth was used as the neutralizer for the four antimicrobial treatments. Three replicate tests on independent days were conducted for each microbe and treatment combination. The target control carrier count level was 4.5 to 5.5 $\log_{10}$(CFU/carrier). Four untreated control carriers and three treated carriers (per antimicrobial test substance) were utilized. In this study, SB was used to generate the *P. aeruginosa* test cultures instead of TSB, and the control and treated carrier suspensions were filtered for recovery.

2.13.2 Study two

Following the completion of interlaboratory Study one, refinements to the method were introduced and a second study was conducted. The revised method was resubmitted for evaluation by seven laboratories. The main revisions included: 1) reducing the number of control carriers from four to three, 2) use of TSB instead of SB to generate the *P. aeruginosa* test suspensions, 3) increasing the target control counts to 5.0 to 6.0 $\log_{10}$(CFU/carrier), and 4) direct plating (in duplicate) of control carrier dilutions onto TSA. The treatments were the quaternary ammonium formulation tested in Study one and two NaOCl treatments. The low efficacy NaOCl treatment for *P. aeruginosa* was adjusted lower to 200 ppm. Three replicate tests on independent days were conducted for each microbe and treatment combination.

2.14 Interlaboratory studies with M. terrae

Two interlaboratory studies were conducted with *M. terrae* as the test microbe.

2.14.1 Study three

In the first study, seven laboratories evaluated four antimicrobial treatments, phenol (0.074%), hydrogen peroxide (4.4%) and peroxyacetic acid (0.23%), citric acid (0.60%), quaternary ammonium compound (0.21%), and a set of NaOCl treatments (2000 ppm and 500 ppm). The target control carrier count level was 4.5 to 5.5 $\log_{10}$(CFU/carrier). Three replicate tests on independent days were conducted for each treatment.

2.14.2 Study four

Participants in Study three indicated that the growth of the static culture was not homogeneous and often appeared pellicle-like in the flask. To resolve the issue, revisions were made to the preparation of the test suspension. For Study four, agitation of the MADC broth culture was used during incubation, followed by centrifugation and a homogenization step (see section 2.3 – *Test Cultures*). Also, the control carrier count range was adjusted to 5.0 to 6.0 logs CFU/carrier. To verify the changes, control carrier count data were generated over three replicate tests on independent days. No antimicrobial treatments were evaluated in this study.

2.15 Statistical methods

In all 4 studies, the mean of the log-transformed density of microbes on the untreated control carrier counts and the associated LRs were the primary responses.

2.15.1 Repeatability and Reproducibility

The main outputs of the statistical analysis are the *repeatability* of these responses from experiment to experiment in an individual laboratory, and the *reproducibility* of the responses across multiple laboratories. These attributes were quantified by a repeatability SD (S_r) and a reproducibility SD (S_R) respectively. Both these SDs are results from analyzing the *TestLD*s and LRs (for each agent) in each multi-laboratory study separately by a linear mixed effects model (LMM) with a random effect for laboratory (Hamilton et al. 2013, Pinheiro and Bates 2002). Each analysis provided the repeatability variance (S_r^2), and also the among-laboratory variance (S_{lab}^2). The reproducibility SD was then $S_R = [S_r^2 + S_{lab}^2]^{½}$. Residual plots were used to investigate potential outliers, confirm the homogeneous variance assumption, and to assess that the residuals approximately followed a normal distribution (Hamilton et al. 2013). These calculations, notation, and terminology are consistent with guidelines published by ASTM E691 and AOAC (2016). Calculations were performed using the software R (R Core Team 2017) package *nlme* (Pinheiro 2017). Explicit R code used for the analysis of a multi-laboratory data set is available on-line (Parker and Hamilton 2013).

2.15.2 Responsiveness

Responsiveness of the method to the differing efficacy levels of NaOCl was assessed by testing whether there was a difference in the mean LRs associated with the two efficacy levels. In each multi-laboratory study, because the two NaOCl treatments were always tested side-by-side on the same day, responsiveness was assessed by fitting an LMM to the difference in LRs observed on each test day.

3. Results

401 *3.1 Interlaboratory test data*

402 *3.1.1 Interlaboratory Study one – S. aureus* and *P. aeruginosa*

403 A total of 113 individual tests (tests involving two NaOCl solutions and four antimicrobial test
404 substances) across both microbes were conducted by 9 laboratories. A summary of the control
405 carrier count data (*TestLDs)* is provided in Table 1 (see also Figures S1-S2 in the Supplementary
406 Material). The specified control carrier count level was consistently achieved and ranged from 4.5
407 to 5.7. Seven of the 113 tests resulted in *TestLDs* values slightly above 5.5; however, the data were
408 deemed valid by the study director and were included in the analysis. The pooled *TestLD*s for the
409 NaOCl tests and the four antimicrobial treatment tests show mean *TestLDs* of 5.07 and 5.03 for *P.*
410 *aeruginosa* and *S. aureus*, respectively. The repeatability (S_r) and reproducibility (S_R) values were
411 less than 0.3 for both microbes – the lowest was the S_r for *S. aureus* at 0.172 and the highest was
412 the S_R for *P. aeruginosa* at 0.288.

413 The six antimicrobial treatments (two NaOCl treatments and four other antimicrobial test
414 substances per microbe) provided a broad range of mean LR values ranging from 3.17 to 5.40 for
415 *P. aeruginosa* tests and 1.51 to 5.28 for *S. aureus* tests (Table 2, Figures 1-2). Using the side-by-
416 side test data for assessing the presumed low and high efficacy solutions of NaOCl, the method
417 was responsive to the differing efficacy levels. In other words, statistically significant differences
418 in the mean LR (2.12 for *P. aeruginosa* and 3.83 for *S. aureus*) were exhibited for the two NaOCl
419 treatments per microbe (p-value < 0.001). The 2000 ppm NaOCl treatment often resulted in no
420 recovery of viable bacteria and the resulting S_r and S_R ranged from 0.162 to 0.276. Higher
421 variability occurred with the 250 ppm and 500 ppm NaOCl concentrations; however, the S_r and S_R
422 were less than 1.0.

The hydrogen peroxide and citric acid-based treatments consistently generated high LRs across both microbes with associated S_r and S_R ≤0.390. The phenolic treatment exhibited mean LRs of 5.33 and 4.50 when tested against *P. aeruginosa* and *S. aureus*, respectively. Similar S_r (e.g., 0.334 for *P. aeruginosa* and 0.429 for *S. aureus*) and S_R (e.g., 0.490 for *P. aeruginosa* and 0.561 for *S. aureus*) were observed for the tests of the phenolic. The quaternary ammonium treatment resulted in comparatively lower LR values for both microbes ranging from 3.26 for *P. aeruginosa* to 3.68 for *S. aureus*. The associated S_R for the quaternary ammonium treatment was higher (1.217 for *P. aeruginosa* and 0.852 for *S. aureus*) compared to the other treatments.

3.1.2 Study two – S. aureus and *P. aeruginosa*

In this method optimization study, a total of 38 individual tests were conducted by 7 labs. Two NaOCl solutions and one antimicrobial test substance per microbe were evaluated. Compared to Study one, the *TestLD*s for both microbes were approximately 0.5 logs higher (Table 1, see also Figure S3 in the Supplementary Material). The increase in the *TestLDs* was anticipated per the modification in the study protocol for a 0.5 log increase in the target range of *TestLDs*. Mean *TestLD*s of 5.45 and 5.67 were recorded for *P. aeruginosa* and *S. aureus*, respectively. S_r and S_R were similar to those observed in Study one; the highest level was the S_R for *P. aeruginosa* at 0.346. Thus, good resemblance of the control carrier counts was confirmed using the method modifications employed to optimize the test procedure.

The NaOCl treatments provided a 3 to 4 log difference in mean LR (Table 2; see also Figure 3). The 2000 ppm NaOCl treatment resulted in a high level of kill for both microbes, mainly resulting in no viable cells recovered. The 200 ppm concentration of NaOCl provided a mean LR of 2.69 for *P. aeruginosa*. *S. aureus* exhibited a mean LR of 1.84 when exposed to the 500 ppm NaOCl. The S_r values for each NaOCl treatment ranged from 0.115 (2000 ppm NaOCl treatment for *S.*

446 *aureus*) to 0.397 (200 ppm NaOCl treatment for *P. aeruginosa*). The S_R ranged from 1.141 for *P.*
447 *aeruginosa* when tested against the 200 ppm NaOCl to 0.234 for the 2000 ppm NaOCl treatment
448 tested against *S. aureus*. The mean difference in LR for the two NaOCl treatments was 3.0 for *P.*
449 *aeruginosa* tests (p-value < 0.001) and 4.1 for *S. aureus* tests (p-value < 0.001), indicating good
450 responsiveness to the differing concentrations of NaOCl for both microbes.

451 The mean LR for the quaternary ammonium-based treatment (also tested in Study one) ranged
452 from 1.36 for *P. aeruginosa* to 2.98 for *S. aureus* (Table 2; see also Figure 3). Across both test
453 microbes, the variability in the data was similar for quaternary ammonium treatment (S_r of 0.537
454 to 0.589 and S_R of 0.718 to 0.798).

455 *3.1.3 Study three – M. terrae*

456 Data from 33 independent tests were generated by four labs. A summary of the *TestLD* for *M.*
457 *terrae* is presented in Table 1 (see also Figure S4 in the Supplementary Material). A mean *TestLD*
458 of 5.09 was exhibited with a S_r of 0.263 and S_R of 0.329 which are indicators of good resemblance
459 of the carrier count data.

460 Across the laboratories, the 500 and 2000 ppm NaOCl treatments provided LRs of 0.97 and 3.28,
461 respectively (Table 2, Figure 4). The S_r of the LR data was 0.247 (500 ppm) and 0.392 (2000 ppm).
462 The 2000 ppm NaOCl treatments exhibited S_R of 0.947 compared to 0.272 for the 500 ppm NaOCl
463 treatment (Table 2). A mean difference in LR between treatments of 2.3 (p-value <0.001) was
464 exhibited between the two NaOCl treatments. Hence, the method was responsive to the two
465 NaOCl treatments against *M. terrae*.

466 The LRs of *M. terrae* ranged from 1.53 (citric acid) to 5.21 (hydrogen peroxide) – see Table 2 and
467 Figure 4. The S_r was 0.257 for tests of citric acid to 0.452 for tests of phenol(Table 2). Tests of the

phenol and the quaternary ammonium treatments exhibiting S_R of 1.175 and 0.987, respectively (Table 2).

3.1.4 Interlaboratory study number four – optimization of the M. terrae test culture

Concerns were expressed by the participants in Study three that *M. terrae* grown in static liquid cultures was difficult to standardize and use of agitation during incubation may be more appropriate. Thus, another study was conducted to optimize the growth conditions of the test microbe. An increase in the *TestLD* to 5.38 from 5.09 was noted for Study four which incorporated the modified suspension methodology and called for a target control count level of 5.0 to 6.0 logs per carrier. Furthermore, S_r and S_R were smaller than in Study three thus verifying that the optimization of the protocol was successful (Table 1; see also Figure S5 in the Supplementary Material).

Table 1. Analysis of log densities of the untreated controls from Studies one through four.

Interlaboratory Study	Microbe	Num. Labs	Num. Tests	Mean *TestLD*	SEM[1] *TestLD*	S_r	S_R
Study one (2012)	*P. aeruginosa*	9	58	5.07	0.065	0.232	0.288
	S. aureus	9	55	5.03	0.075	0.172	0.274
Study two (2014)	*P. aeruginosa*	7	19	5.45	0.114	0.222	0.346
	S. aureus	7	19	5.67	0.079	0.163	0.244
Study three (2013)	*M. terrae*	6	33	5.09	0.092	0.263	0.329
Study four[2]	*M. terrae*			5.38		0.235	0.324

(2014)		4	11	0.132	

[1]SEM: standard error of the mean *TestLD* for the controls

[2]Study included control carrier count evaluation only.

Table 2. Mean log reduction values and associated repeatability and reproducibility standard deviations for the OECD Quantitative Method as evaluated in interlaboratory studies. The numbers of labs and experiments are the same as reported in Table 1 except for Study three where the NaOCl treatments were evaluated by 17 tests, and the other treatments by 16 tests, in 6 laboratories.

Study	Microbe	Treatment	Mean LR	SEM LR*	S_r	S_R
Study one (2012)	*P. aeruginosa*	250 ppm NaOCl	3.17	0.278	0.340	0.879
		2000 ppm NaOCl	5.29	0.076	0.192	0.276
		Phenolic	5.33	0.136	0.334	0.490
		Hydrogen peroxide	5.40	0.098	0.315	0.390
		Citric acid	5.39	0.099	0.248	0.360
		Quaternary ammonium	3.26	0.333	0.843	1.217
	S. aureus	500 ppm NaOCl	1.51	0.144	0.230	0.472
		2000 ppm NaOCl	5.34	0.062	0.162	0.227
		Phenol	4.50	0.146	0.429	0.561
		Hydrogen peroxide	5.28	0.102	0.157	0.331
		Citric acid	5.23	0.095	0.271	0.362
		Quaternary ammonium	3.68	0.261	0.407	0.852

Study two (2014)	*P. aeruginosa*	200 ppm NaOCl	2.69	0.416	0.397	1.141
		2000 ppm NaOCl	5.71	0.145	0.258	0.431
		Quaternary ammonium	1.36	0.208	0.589	0.718
	S. aureus	500 ppm NaOCl	1.84	0.243	0.175	0.655
		2000 ppm NaOCl	5.95	0.082	0.115	0.234
		Quaternary ammonium	2.98	0.257	0.537	0.798
Study three (2013)	*M. terrae*	500 ppm NaOCl	0.97	0.757	0.247	0.272
		2000 ppm NaOCl	3.28	0.365	0.392	0.947
		Phenol	2.22	0.457	0.452	1.175
		Hydrogen peroxide	5.21	0.113	0.400	0.419
		Citric acid	1.53	0.357	0.257	0.898
		Quaternary ammonium	2.30	0.370	0.494	0.987

488 * SEM: standard error of the mean log reduction

489

490 **4. Discussion**

491 Four interlaboratory studies were used to develop a statistical profile of an optimized version of
492 the draft Quantitative Method. The wide range of efficacy of the antimicrobial treatments
493 evaluated provided a wide range of mean LRs as well as an assessment of repeatability of the LRs
494 across experiments (S_r) and the reproducibility of the LRs across labs (S_R) – see Table 2. In studies
495 one, two and three, three distinct microbes were tested against relevant antimicrobial test
496 substances. In addition, the use of NaOCl solutions in the test design provided three functions, first
497 as a relevant active ingredient in the disinfectant marketplace; second, for use as a potential

reference standard material for the development of proficiency exercises; and third for verifying the responsiveness of the test method.

Here the authors have identified four main areas of method performance: 1) resemblance of untreated control carriers, 2) S_r of tests within a single laboratory, 3) S_R of tests between laboratories, and 4) responsiveness of the method to changes in presumed treatment efficacy. A quantitative method may be described as repeatable if several independent tests in the same laboratory of the same antimicrobial test substance produce comparable *TestLDs* and LRs. The method is reproducible if several independent tests in different laboratories of the same antimicrobial test substance produce comparable *TestLDs* and LRs. Thus, small S_r and S_R of the *TestLDs* and the LRs are highly desirable and indicative of acceptable method performance. Due to the additional sources of variability, the laboratory-to-laboratory variability (S_R) is expected to be higher than the variability within a single laboratory (S_r); this was observed across the studies reported here. Ultimately, acceptable standard methods will be used by multiple laboratories, thus the S_R level is the primary measure of variability of interest.

Good resemblance of control carrier counts is an essential property of a reliable quantitative microbiological assay for testing the efficacy of disinfectants. Although some variability will always exist in microbial assays, the use of the method's standardized procedures for the preparation of the microbial test cultures, carrier inoculation process, management of the inoculated carriers and microbe recovery ensure the generation of consistent *TestLD*s. Further, the acceptance criterion for meeting the target *TestLD*s provides assurance that a test is valid. The level of resemblance of the *TestLD*s accounts for the day-to-day variability associated with many factors: the final test cultures, environmental conditions affecting the rate of drying, loss of viability while drying, inoculum removal, and recovery. The small variability of *TestLD*s exhibited

by the Quantitative Method, as measured by S_r and S_R for the *TestLDs*, is a strong indicator of good resemblance. As the acceptable range for *TestLD* was increased from 4.5-5.5 (in Study one) to 5.0-6.0 $\log_{10}$(CFU/carrier) (in Study two), the *TestLD* for each test microbe exhibited comparable levels of variability. Among the test microbes evaluated, *S. aureus* control carrier count data exhibited the smallest S_r (ranging from 0.16 to 0.17) and S_R (ranging 0.24 to 0.27). The resemblance S_r and S_R for *P. aeruginosa* and *M. terrae* control carriers were slightly higher than *S. aureus*, but small with a maximum of 0.346 (the S_R for *P. aeruginosa* in Study two).

Across the studies, the S_r values for LR were small (i.e., less than 0.84), most notably for treatments generating high mean LRs (5 or above) where the S_r was as small as 0.12. This outcome is indicative of a repeatable method. Only three of the twenty-four treatment combinations over the study period resulted in S_R values above 1.0 – the associated LR for the three treatments ranged from 2.22 to 3.26. Antimicrobial treatments with a LR of 4.5 or higher exhibited S_R of 0.561 or less. The S_R values ranged from 0.227 (see Study one, 2000 ppm NaOCl treatment for *S. aureus*) to 1.217 (see study one, quaternary ammonium treatment for *P. aeruginosa*). Data from Study two which implemented targeted method optimizations resulted in S_R ranging from 0.234 for the 2000 ppm NaOCl treatment for *S. aureus* to 1.141 for the 200 ppm NaOCl treatment for *P. aeruginosa*. Also, in Study two, the S_R for the quaternary ammonium treatment for *S. aureus* and *P. aeruginosa* were 0.798 and 0.781, respectively.

Based on differences in the LR values generated from the side-by-side testing of the NaOCl treatments with weak and strong activity, the method was shown to be responsive to changes in the NaOCl concentration for each test microbe. The use of reagent grade (laboratory grade) NaOCl may serve as a standardized test chemical for training and proficiency testing and may serve as an

internal control to verify the quality of the test system on a day-to-day basis. In practice, other active ingredients with documented levels of LR may also be used as an internal reference standard; however, their efficacy profile should be verified by multiple laboratories in advance of use.

Although standard setting organizations such as AOAC International and ASTM International have not established formal guidance on acceptable levels of variability for quantitative methods, the literature and several examples of approved standard methods may be used for comparative purposes. Based on a review of the historical literature of suspension and dried surface tests, Tilt and Hamilton provide an assessment that a S_r less than 1.0 and a S_R less than 1.5 should be considered acceptable in terms of method performance (Tilt et al., 1999). Here we provide three explicit examples. First, data to support the validation of the AOAC Three Step Method (AOAC method 2008.05, 2008), a quantitative procedure for testing sporicidal chemicals against spores of *Bacillus subtilis*, were generated from an eight-laboratory study (Tomasino et al., 2008). In the study, the TSM control carriers exhibited S_r and S_R of 0.15 and 0.27, respectively. The resulting LR for the TSM's treated carriers exhibited S_r ranging 0.17 to 0.72 and S_R 0.34 to 1.43. The TSM is currently recognized by the EPA as an acceptable test method to support sporicidal product claims against spores of *B. anthracis*. In the second example, data from an eight-laboratory study were used to support the MBEC™ Physiology & Genetics Assay as an ASTM standard (ASTM E2799-17, 2017; Parker et al., 2014) for measuring disinfectant efficacy against *P. aeruginosa* biofilm. The S_r and S_R for the untreated controls for the MBEC™ method were 0.33 and 0.67, respectively. The LR associated with the disinfectant treatments tested with the MBEC™ method exhibited S_r levels ranging from 0.22 to 1.61, and S_R levels of 0.27 to 1.70. Our third and last comparative example is the qualitative AOAC Use-dilution Method (AOAC Method 955.15, 2012;

AOAC Method 964.02, 2012) that was assessed in a five-laboratory collaborative study. The LRs exhibited S_r of 0.15 to 0.33 and S_R of 0.32 to 0.55 in tests involving *S. aureus* and *P. aeruginosa* (Tomasino et al., 2012). In the same study, the control carriers exhibited small S_r (range of 0.16 to 0.25) and S_R (range of 0.30 to 0.38).

5. Conclusions

In 2013, the data provided in this report were used to support the approval of an OECD Guidance Document on Quantitative Methods for Evaluating the Activity of Microbicides used on Hard Non-Porous Surfaces by the WNT (OECD, 2013); however, additional peer review is ongoing in the US to further standardize the method. Furthermore, the data and statistical analyses contained in this report are historical in nature and useful baseline information for individuals conducting additional technical review of the method.

The statistical attributes reported here for the draft Quantitative Method when used to test *P. aeruginosa, S. aureus,* and *M. terrae* provide information for decision makers when considering the method as a candidate regulatory procedure. Based on the data, the Quantitative Method displays a statistical profile consistent with other standard methods approved by standard-setting organizations where method performance data are available. Currently, the Quantitative Method is recommended by the EPA for use in supporting product claims designed to treat hard non-porous surfaces contaminated with spores of *Clostridium difficile* (US EPA, 2018) and for cells of *Candida auris* (US EPA, 2017), an emerging yeast pathogen. Additionally, the technical quality and flexibility of ASTM E2197 for testing the effectiveness of antimicrobial substances against a wide variety of microorganisms has been widely demonstrated in the published domain (Alhmid et al., 2016; Cadnum et al., 2017; Cook et al., 2016; Kanamori et al., 2018; Majcher et al., 2008;

590 Sabbah et al., 2018; Sattar et al., 2003; Springthorpe et al., 2005, Springthorpe et al., 2007;
591 Uwamahoro et al., 2018).
592
593 Use of the Quantitative Method could potentially provide measurable benefits to regulators and
594 the stakeholder community due to its ease of use, clear interpretation of the quantitative results,
595 reduced laboratory biosafety hazards (e.g., reduced amount of inoculum required), likely reduction
596 in repeat testing, and the method's flexibility to accommodate multiple active ingredients and
597 microbes. The method also may be used to test spray product formulations (as liquids), thus
598 reducing the need for a separate spray test. The results presented here provide examples of how
599 the method may be successfully utilized for testing a variety of microbes. The current use of the
600 method by the EPA to support *C. difficile* and *C. auris* claims provide examples of its utility as a
601 regulatory procedure for testing high-consequence public health pathogens. The microorganisms
602 tested in our experiments are considered to be representatives of clinical pathogens; however,
603 individual regulatory authorities may require additional or alternative microorganisms to be tested
604 to support specific label claims. Furthermore, the LR data generated from this method may be used
605 to develop microbial kill curves for product research and development and may be useful in risk-
606 based analyses (Haas et al., 2014, US EPA 2012). In theory, for manufacturers seeking to market
607 products internationally, successful testing with an alternative harmonized method may allow for
608 product registration in other OECD member countries; thus minimizing pre-market costs and
609 providing regulators with simpler data sets to review and analyze. Harmonization of test
610 procedures could also benefit the end-users by the development of more consistent, streamlined
611 product use directions.
612

As demonstrated in this report, the availability of quantitative data allows for rigorous statistical analyses to inform the end-user of the amount of potential variability associated with the untreated control and treated populations of microbes on a hard, non-porous surface. Based on the observations over the three multi-laboratory studies involving the measurement of LRs, it appears that highly effective (e.g., LR $\geq$ 5.0) antimicrobial treatments resulting in few to no surviving microbes and antimicrobial treatments with limited effectiveness (e.g., LR $\leq$1.0) exhibit greater reproducibility. In contrast, antimicrobial treatments resulting in intermediate LR (e.g., 2.0 to 4.0) show higher variability in the results, thus lower reproducibility. These general observations are supported by a review conducted by Parker et al., 2018 which quantifies the relationship between variability (i.e., S_R) and the associated mean LR for several antimicrobial test methods (including ASTM E2197). A similar analysis of the trends in the current data set will be necessary to more completely ascertain the relationship between the mean LR and the repeatability and reproducibility of the Quantitative Method. Further, by understanding the level of variability associated with the Quantitative Method, stakeholders and regulators may use the data to propose appropriate LR-based product performance standards and testing criteria to ensure high confidence in the outcome of the test. An example of this modeling exercise is provided in the reassessment of the AOAC Use dilution Method's performance standards where a multi-laboratory collaborative study was used to generate method performance data and variability values (Tomasino et al., 2014; Parker et al., 2014). The statistical methodology utilizes error rates (i.e., Type I pass-error and Type II fail-error) to formulate a set of testing criteria (e.g., number of independent tests and the required LRs) necessary to achieve high confidence in the outcome of the tests. The approach can accommodate criteria that require tests against different microbes, even potentially with different

levels of variability as we demonstrate here. The application of this statistical model on the current data set will be covered under a separate report.

References

Abreu, A., Tavares, R., Borges, A., Mergulhao, F., Simoes, M. 2013. Current and emergent strategies for disinfection of hospital environments. Journal of Antimicrobial Chemotherapy. 68:2718-2732.

Agency for Healthcare Research and Quality (AHRQ), Technical Brief. 2015. Environmental cleaning for the prevention of healthcare-associated infections. AHRQ publication No. 15-EHC020-EF.

Alfa, M., Lo, E., Olson, N., MacRae, M., Buelow-Smith, L. 2015. Use of daily disinfectants cleaner instead of a daily cleaner reduced hospital-acquired infection rates. American Journal of Infection Control, 43: 141-146.

Alhmid, H., Koganti, S., Cadnum, J., Rai, H., Jencson, A., Donskey, C. 2016. Evaluation of a Novel Alcohol-based Surface Disinfectant for Disinfection of hard and Soft Surfaces in Healthcare Facilities. Open Forum Infectious Diseases. Brief Report. 1-3.

AOAC, *Guidelines for Collaborative Study Procedures to Validate Characteristics of a Method of Analysis*, in *Official Methods of Analysis of AOAC International*. 2016.

655 AOAC Official Methods of Analysis. Testing disinfectants against *Staphylococcus*
656 *aureus*. Use-dilution method. 2012; 18th Ed., AOAC INTERNATIONAL, Gaithersburg,
657 MD, Method 955.15.

658 AOAC Official Methods of Analysis. Testing disinfectants against *Pseudomonas*
659 *aeruginosa*. Use-dilution method. 2012; 18th Ed., AOAC INTERNATIONAL,
660 Gaithersburg, MD, Method 964.02.

661 AOAC Official Methods of Analysis. Three Step Method. 18th Ed., AOAC
662 INTERNATIONAL, Gaithersburg, MD, Method 2008.05.

663 ASTM Standard E691, 2013. Standard Practice for Conducting an Interlaboratory Study
664 to Determine the Precision of a Test Method. ASTM International, West Conshohocken,
665 PA.

666 ASTM Standard E2197-17. 2017. Standard quantitative disk carrier test method for
667 determining the bactericidal, virucidal, fungicidal, mycobacterial, and sporicidal activities
668 of liquid chemical germicides.

669 ASTM Standard E2799-17. 2017. Standard Method for Testing Disinfectant Efficacy
670 against *Pseudomonas aeruginosa* Biofilm using the MBEC Assay. ASTM International,
671 West Conshohocken, PA.

672 Boyce, J., 2016. Modern technologies for improving cleaning and disinfection of
673 environmental surfaces in hospitals. Antimicrobial Resistance and Infection Control.
674 5:10.

675 Cadnum, J., Shaikh, A., Piedrahita, C., Sankar, T., Jencson, A., Larkin, E., Ghannoum,
676 M., Donskey, C. 2017. Effectiveness of Disinfectants against *Candida auris* and other

677 Candida Species. Infection Control and Hospital Epidemiology. Vol. 38, No. 10. 1240-
678 1243.

679 Cook, B., Cutts, T., Nikiforuk, A., Kobasa, D., Theriault, S. 2016. The Disinfection
680 Characteristics of Ebola Virus Outbreak Variants. Scientific Reports. 6:38293

681 Dancer, S., 2014. Controlling hospital-acquired infection: focus on the role of the
682 environment and new technologies for decontamination. 27 (4) 665-690.

683 Donskey, C., 2013. Does improving surface cleaning and disinfection reduce health care-
684 associated infections? American Journal of Infection Control. 41: s12-s19.

685 Haas, Rose, Gerba. QMRA. John Wiley, 2014

686 Hamilton, M., Hamilton, G., Goeres, D., Parker, A., 2013. Guidelines for the statistical
687 analysis of a collaborative study of a laboratory method for testing disinfectant product
688 performance. J. AOAC Int. Vol. 96 (5), 1138-1151. https://doi.org/10.5740/jaoacint.12-
689 217

690 Hamilton, M. 2011. Montana State University Center for Biofilm Engineering.
691 Knowledge Sharing Articles. KSA-SM-011: How to decide whether reproducibility
692 standard deviation is small enough.

693 Hamilton MA, Parker AE "KSA-SM-06—Enumerating viable cells by pooling counts for
694 several dilutions," Knowledge Sharing Articles, Center for Biofilm Engineering at
695 Montana State University, Bozeman, MT. September 22, 2010

696 Hota, B., 2004. Contamination, disinfection, and cross-colonization: are hospital surfaces
697 reservoirs for nosocomial infection? Clinical Infectious Diseases. 39: 1189-1189.

698 Kanamori, H., Rutala, W., Gergen, M., Sickbert-Bennett, E., Weber, D. 2018. Germicidal
699 Activity against Carbapenem/Colistin-Resistant *Enterobacteriaceae* Using a Quantitative
700 Carrier Test Method. Antimicrobial Agents and Chemotherapy. 62:7.

701 Majcher, M., Bernard, K., Sattar, S. 2008. Identification by quantitative carrier test of
702 surrogate spore-forming bacteria to assess sporicidal chemicals for use against *Bacillus*
703 *anthracis*. Applied and Environmental Microbiology. 74: 676-681.

704 Organisation for Economic Co-operation and Development (OECD). 2009. Guidance
705 Document for the Development of OECD Guidelines for the Testing of Chemicals.
706 OECD Environmental Health and Safety Publications. Series on Testing and Assessment.
707 Number 1.

708 Organisation for Economic Co-operation and Development (OECD). 2011. Series on
709 Testing and Assessment: Other (e.g., Pesticide residue chemistry, biocides efficacy,
710 miscellaneous). No. 154. Validation report: validation of efficacy methods for
711 antimicrobials used on hard surfaces, Part 1 and Part 2.

712 Organisation for Economic Co-operation and Development. 2018. About the OECD.
713 http://www.oecd.org/about/whodoeswhat/.

714 Parker, A., Hamilton, M., 2011. Assessing resemblance, repeatability and reproducibility
715 for quantitative methods. Center for Biofilm, Engineering, KSA-SM-10
716 (http://www.biofilm.montana.edu/content/ksa-sm-10).

717 Parker, A., Hamilton, M., 2013. Using R to assess resemblance, repeatability, and
718 reproducibility for quantitative and semi-quantitative disinfectant methods. Center for

719 Biofilm, Engineering, KSA-SM-13 (http://www.biofilm.montana.edu/content/ksa-sm-
720 13).

721 Parker, A., Tomasino, S., Hamilton, M.A., 2014. Statistical model for assessing
722 performance standards for quantitative and semi-quantitative disinfectant test methods. J
723 AOAC Int. Vol. 97 (1), 58–67. https://doi.org/10.5740/jaoacint.13-105

724 Parker, A., Walker, D., Goeres, D., Olson, M., Omar, A. 2014. Ruggedness and
725 reproducibility of the MBEC biofilm disinfectant efficacy test. Journal of
726 Microbiological Methods. 102 55-64.

727 Parker, A., Hamilton, M., Goeres, D. 2018. Reproducibility of antimicrobial test
728 methods. Scientific Reports. 8:12531.

729 Pinheiro, J.C. and D.M. Bates, *Mixed-Effects Models in S and S-PLUS*. 1st ed. 2000.
730 Corr. 3rd printing 2002 edition ed. 2002, New York: Springer. 528.

731 Pinheiro, J.C., et al., *nlme: Linear and Nonlinear Mixed Effects Models*. 2017.

732 Quinn, M., Henneberger, P., Braun, B., Delclos, G., Fagan, K., Huang, V., Knaack, J.,
733 Kusek, L., Lee, S., Le Moual, N., Maher, K., McCrone, S., Mitchell, A., Pechter, E.,
734 Rosenman, K., Sehulster, L., Stephens, A., Wilburn, S., Zock, J. 2015. Cleaning and
735 disinfecting environmental surfaces in health care: toward an integrated framework for
736 infection and occupational illness prevention. 43: 424-434.

737 R Core Team, *R: A language and environment for statistical computing*. 2017, R
738 Foundation for Statistical Computing: Vienna, Austria.

739 Rutala, W., Weber, D., Disinfectants used for environmental disinfection and new room
740 technology. 2013. American Journal of Infection Control. 41: S36-S41.

.

741 Sabbah, S., Springthorpe, S., Sattar, S. 2010. Use of a Mixture of Surrogates for
742 Infectious Bioagents in a Standard Approach to Assessing Disinfection of Environmental
743 Surfaces. 2010. Applied and Environmental Microbiology. Vol. 76. No. 17. Pp. 6020-
744 6022.

745 Sattar, S., Springthorpe, S., Abegbunrin, A., Zafer, A., Busa, M. 2003. A disc-based
746 quantitative carrier test method to assess the virucidal activity of chemical germicides.
747 Journal of Virological Methods. 112: 3-12.

748 Springthorpe, S., Sattar, S., 2005. Carrier tests to assess microbicidal activities of
749 chemical disinfectants for use on medical devices and environmental surfaces. J AOAC
750 Int. Vol. 88 (1), 182-201.

751 Springthorpe, S., Sattar, S., 2007. Application of a quantitative carrier test to evaluate
752 microbicides against mycobacteria. J. AOAC Int. Vol. 9 (3), 817-824.

753

754 Tilt, N., Hamilton, M. 1999. Repeatability and reproducibility of germicide tests: a
755 literature review. J AOAC Int. 82 (2), 384-389.

756 Tomasino, S., Blackburn, T., Hardy, J., 2021. Regulation of Antimicrobial Pesticides by
757 the US Environmental Protection Agency. Chapter 72. Block's Disinfection, Sterilization
758 and Preservation. Sixth edition.

759 Tomasino, S., Hamilton, M., 2006. Modification to the AOAC sporicidal activity of
760 disinfectants test (method 966.04): collaborative study. J. AOAC Int. Vol. 89, (5). 1373-
761 1397.

.

762 Tomasino, S., Parker, A., Hamilton, M., Hamilton, G., 2012. Performance of the AOAC
763 use-dilution method with targeted modifications: collaborative study. J. AOAC Int. 95
764 (6), 1618-1628. https://doi.org/10.5740/jaoacint.12-170

765

766 Tomasino, S., Parker, A., Hamilton, M., 2014. Use of statistical modeling to reassess the
767 performance standard for the AOAC use-dilution methods (955.15 and 964.02). J. AOAC
768 Int. Vol. 97 (1), 68-77. https://doi.org/10.5740/jaoacint.13-216

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

772

773

774 Tomasino SF, Pines RM, Hamilton MA, 2009. Improving the AOAC use-dilution
775 method by establishing a minimum log density value for test microbes on inoculated
776 carriers. J. AOAC Int., 92 (5), 1531-1540.

777 Tomasino SF, Pines RM, Hamilton GC, 2012. Procedural revision to the use-dilution
778 method: establishment of maximum log density value for test microbes on inoculated
779 carriers. J. AOAC Int. Vol. 95 (4), 1059-1063. https://doi.org/10.5740/jaoacint.12-171

780 United States Environmental Protection Agency. CFR, Title 40—Protection of
781 Environment chapter 1-Part 160—Good Laboratory Practice Standards).

782 United States Environmental Protection Agency, 1997. EPA Good Laboratory Practice
783 Standards (Code of Federal Regulations, Title 40, Volume 14, Parts 150 to 189)

.

United States Environmental Protection Agency, 2012. Microbial Risk Assessment Guideline: Pathogenic Microorganisms with Focus on Food and Water.https://www.epa.gov/risk/microbial-risk-assessment-guideline-pathogenic-microorganisms-focus-food-and-water

United States Environmental Protection Agency, 2012. Office of Chemical Safety and Pollution Prevention Harmonized Test Guidelines. Antimicrobial Test Guidelines. Series 810.2000–Product Performance Test Guidelines for Public Health Uses of Antimicrobial Agents. http://www.epa.gov/ocspp/pubs/frs/publications/Test_Guidelines/series810.htm

United States Environmental Protection Agency, 2016. Method Validation of U.S. Environmental Protection Agency Microbiological Methods of Analysis Prepared for: The EPA Forum on Environmental Measurements (FEM) Prepared by: The FEM Microbiology Action Team FEM Document Number 2009-01.

United States Environmental Protection Agency, 2017. Methods and Guidance for Testing the Efficacy of Antimicrobial Products Against Biofilms on Hard, Non-Porous Surfaces. https://www.epa.gov/pesticide-analytical-methods/methods-and-guidance-testing-efficacy-antimicrobial-products-against

United States Environmental Protection Agency, 2018. Methods and Guidance for Testing the Efficacy of Antimicrobial Products Against Spores of *Clostridium difficile* on Hard Non-Porous Surfaces. https://www.epa.gov/pesticide-registration/methods-and-guidance-testing-efficacy-antimicrobial-products-against-spores.

Uwamahoro, M., Massicotte, R., Hurtubise, Y., Gagne-Bourque, F., Mafu, A., Yahia, L., Evaluating the Sporicidal Activity of Disinfectants against *Clostridium difficle* and *Bacillus amyloliquefaciens* Spores by Using the Improved Methods Based on ASTM E2197-11. 2018. Frontiers in Public Health. Vol. 6. Article 18:1-6.

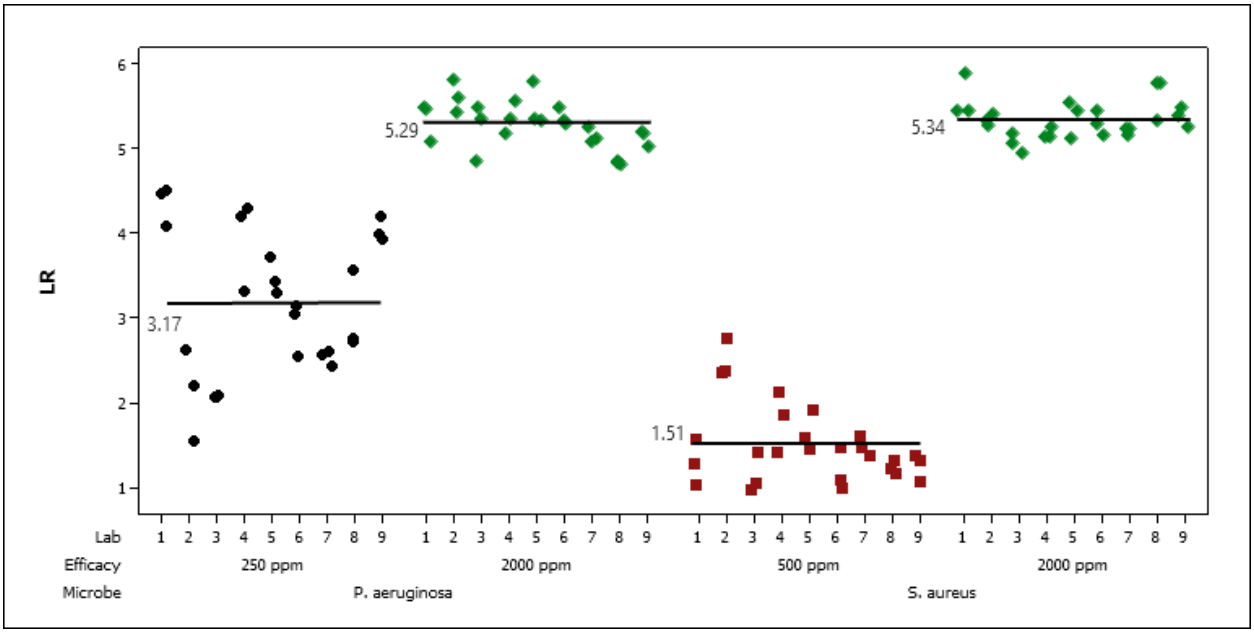


Figure 1. The LRs for *P. aeruginosa* and *S. aureus* for the NaOCl treatments in Study one (2012). Each point is a LR for a single test. Horizontal black lines indicate the mean LR for each treatment. Different colors and symbols are used to differentiate the treatments. Some horizontal jitter has been added to the data to help with visualization.

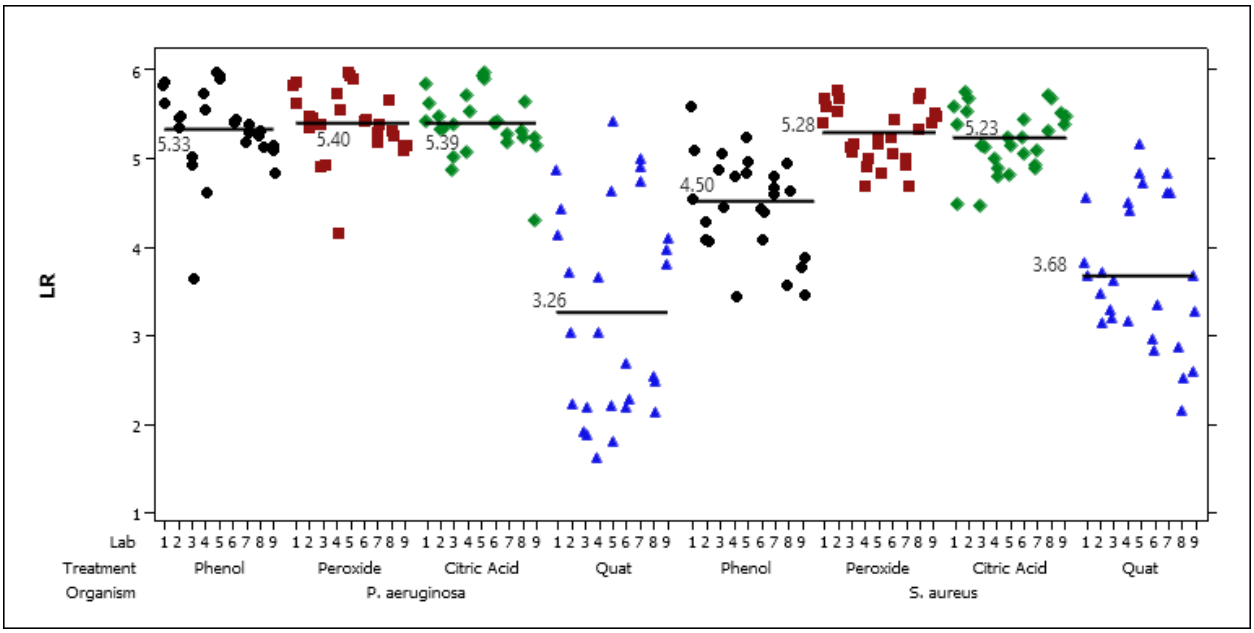
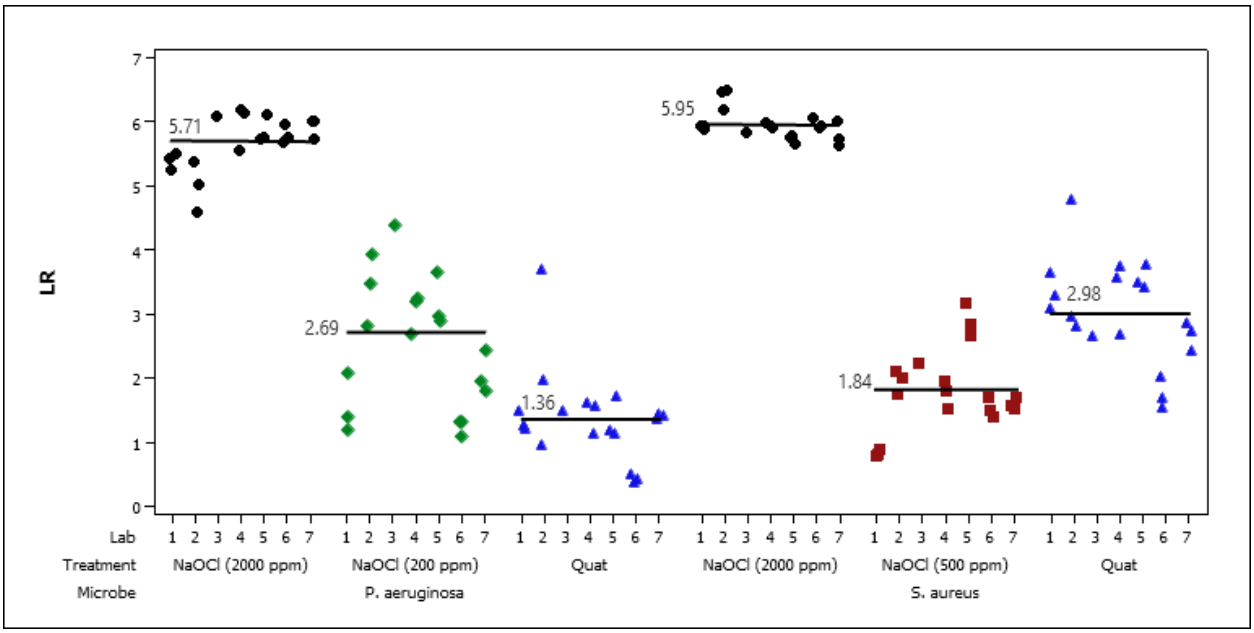


Figure 2. The LRs for *P. aeruginosa* and *S. aureus* for the phenolic, hydrogen peroxide, citric acid, and quaternary ammonium treatments in Study one (2012). Each point is a LR for a single test. Horizontal black lines indicate the mean LR for each treatment. Different colors and symbols are used to differentiate the treatments. Some horizontal jitter has been added to the data to help with visualization.



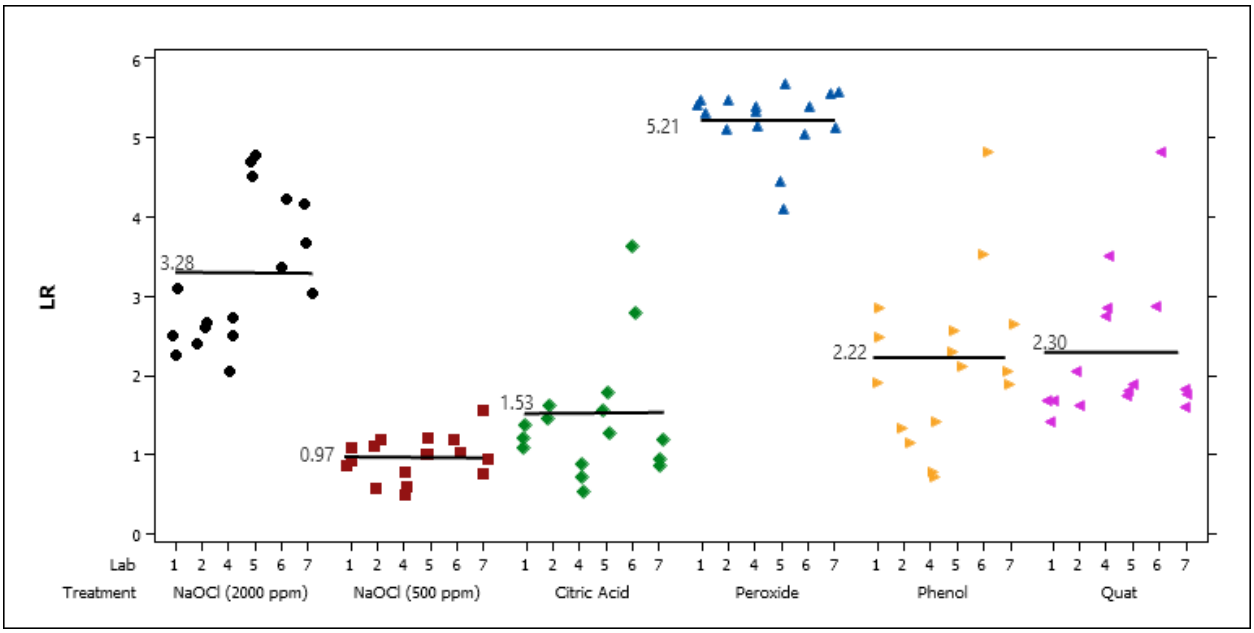
826

827 **Figure 3.** LRs for *P. aeruginosa* and *S. aureus* from Study two (2014). Each point is a log
828 reduction for a single test. Different colors and symbols are used to differentiate the treatments.
829 Horizontal black lines indicate the mean LR for each treatment. Some horizontal jitter has been
830 added to the data to help with visualization.

831

832

833



834

835 **Figure 4.** LRs for *M. terrae* for the NaOCl, phenolic, hydrogen peroxide, citric acid and
836 quaternary ammonium treatments in Study three (2013). Each point is a log reduction for a
837 single test. Different colors and symbols are used to differentiate the treatments. Horizontal
838 black lines indicate the mean LR for each treatment. Some horizontal jitter has been added to
839 the data to help with visualization.

840

841

842

843

844

845

846

847

848

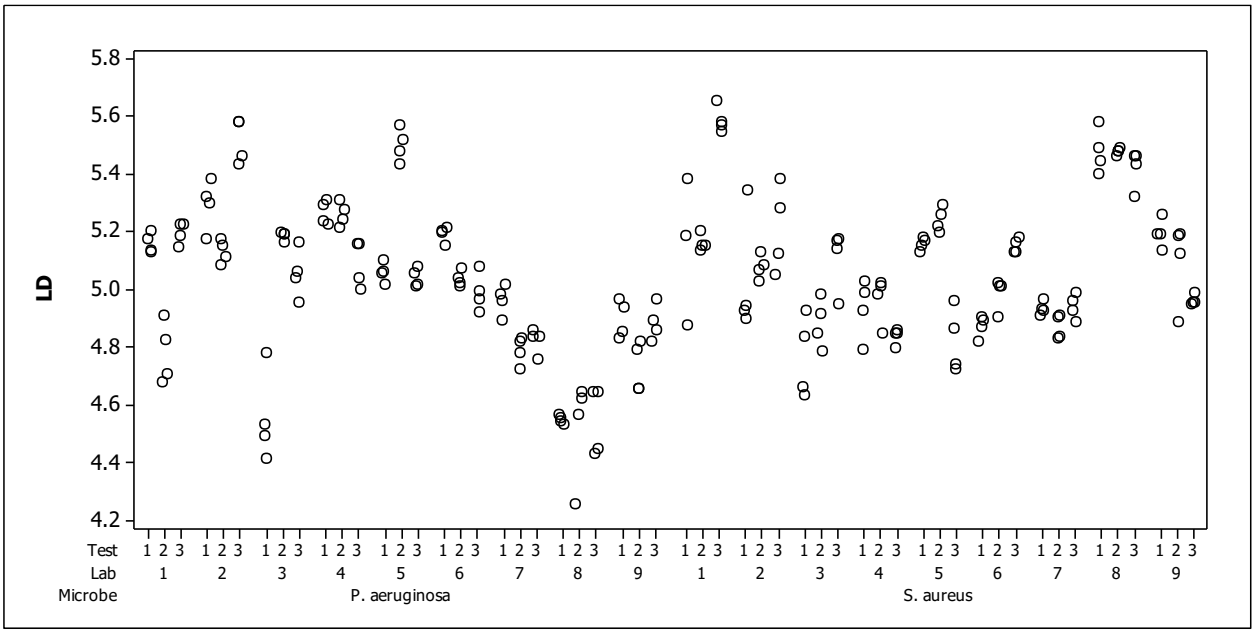
849

850 **Supplementary Material**

851 Plots of the control count data are provided for each of the 4 studies in the following 4 sections.

852 **S.1 Study one**

853



854

855 **Figure S1.** The untreated controls for *P. aeruginosa* and *S. aureus* for the NaOCl treatments in

856 Study one (2012). Each point is the log (CFU/carrier) for a single carrier. Some horizontal jitter

857 has been added to the data to help with visualization.

858

859

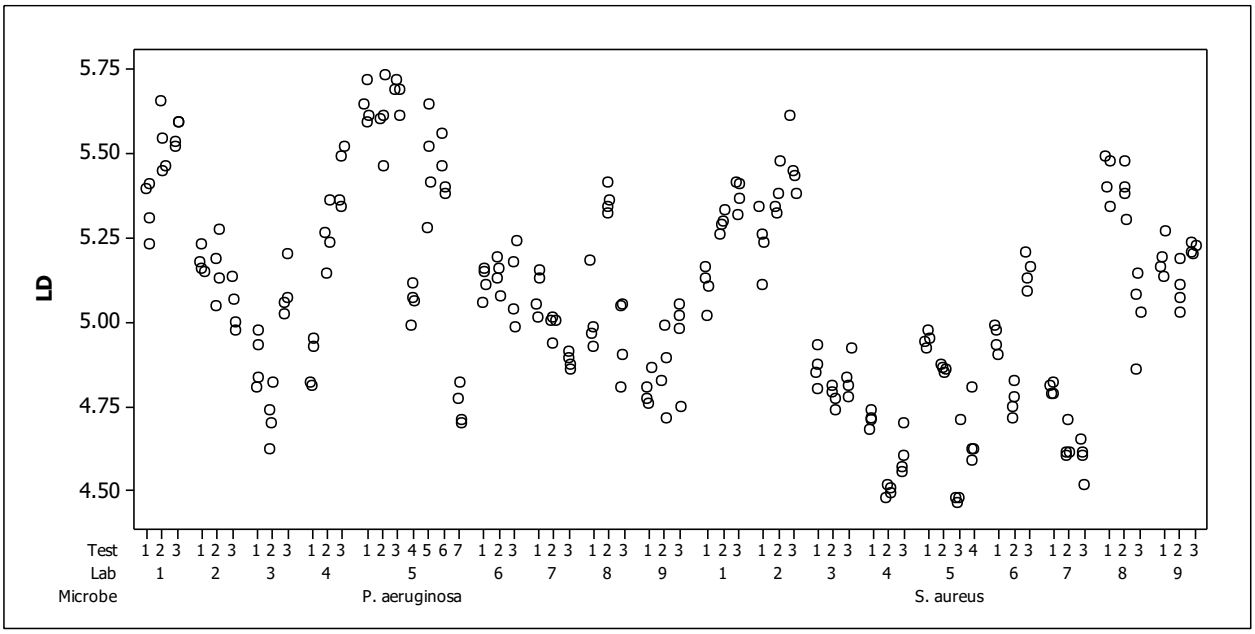


Figure S2. The untreated controls for *P. aeruginosa* and *S. aureus* for the phenolic, hydrogen peroxide, citric acid, and quaternary ammonium treatments in Study one (2012). Each point is the log (CFU/carrier) for a single carrier. Some horizontal jitter has been added to the data to help with visualization.

S.2 Study two

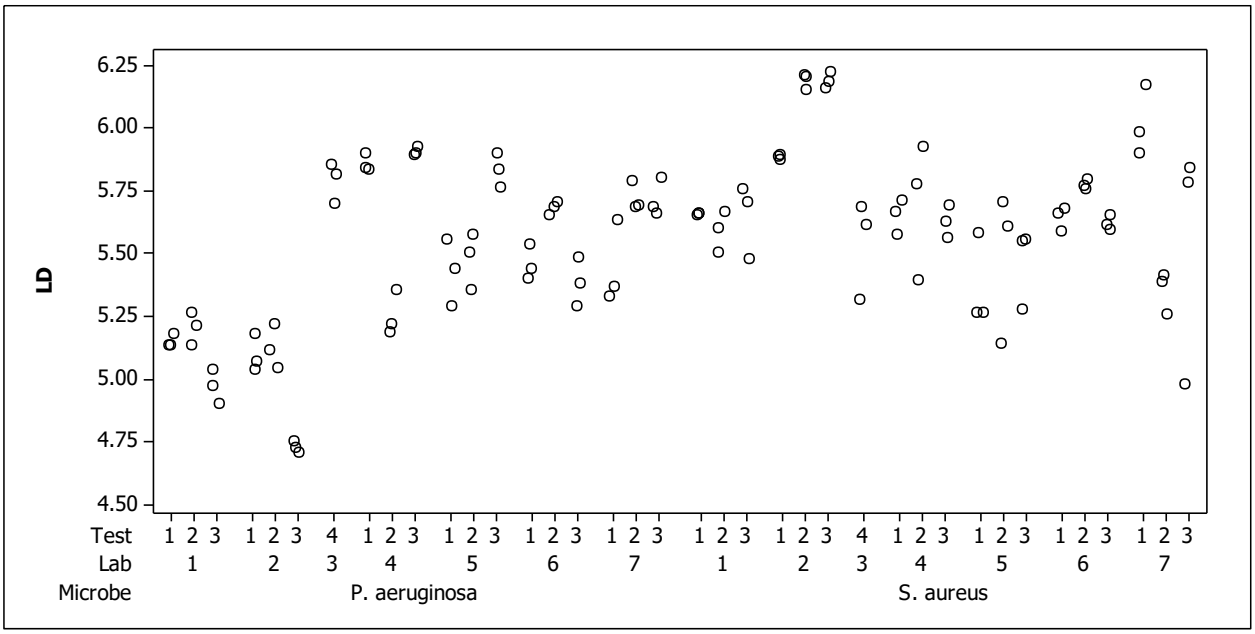


Figure S3. Untreated controls for *P. aeruginosa* and *S. aureus* from Study 2 (2014). Each point is the $\log_{10}$ (CFU/carrier) for a single control carrier in a test. Some horizontal jitter has been added to the data to help with visualization.

S.3 Study three

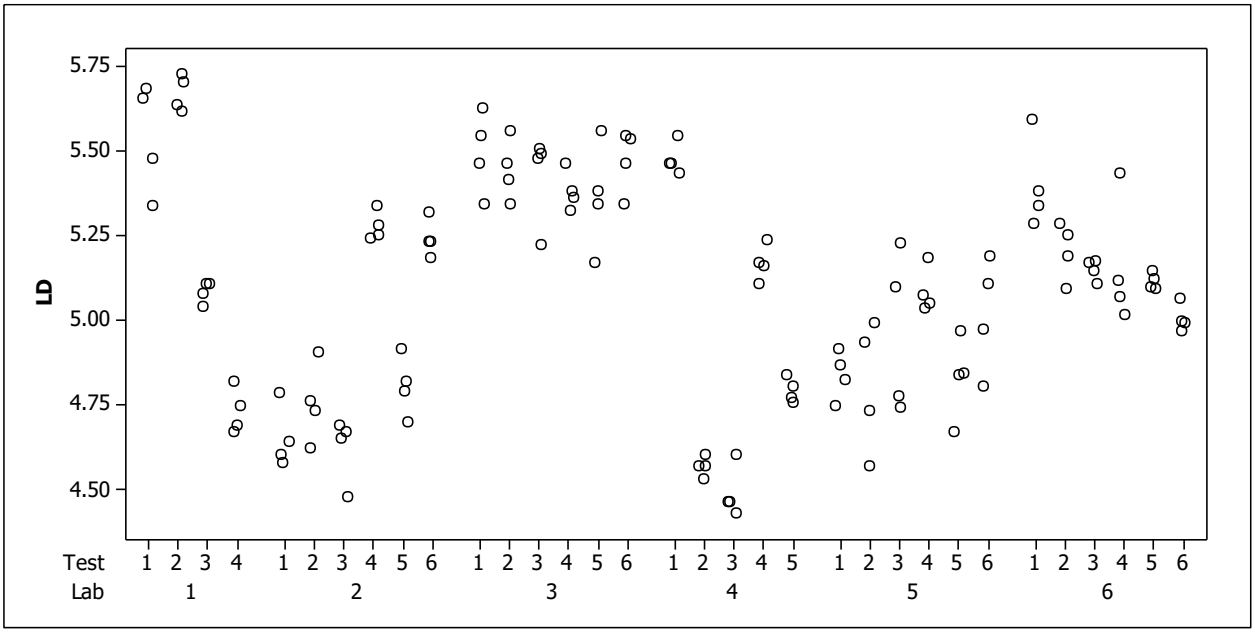
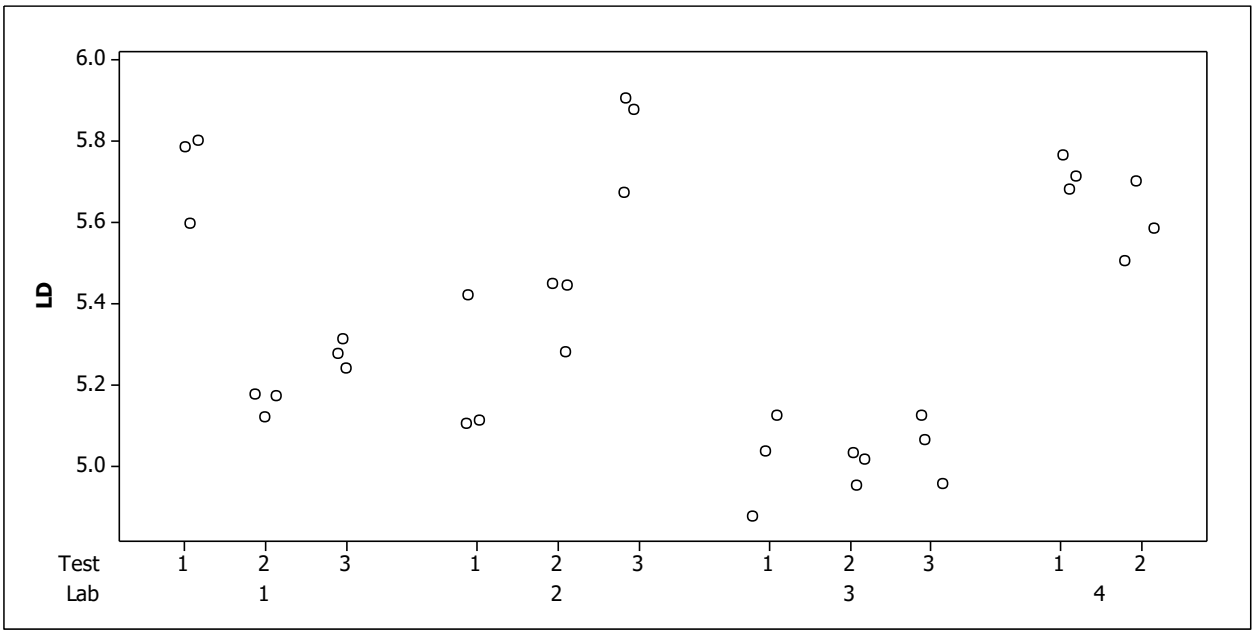


Figure S4. Untreated controls for *M. terrae* from Study three (2013). Each point is the $\log_{10}$ (CFU/carrier) for a single control carrier in a test. Some horizontal jitter has been added to the data to help with visualization.

S.4 Study four



911

912 **Figure S5.** Untreated controls for *M. terrae* from Study four (2014). Each point is the $\log_{10}$

913 (CFU/carrier) for a single control carrier in a test. Some horizontal jitter has been added to the

914 data to help with visualization.

915