MONTANA STATE UNIVERSITY

Center for Biofilm Engineering

Biofilm Regulatory Toolbox

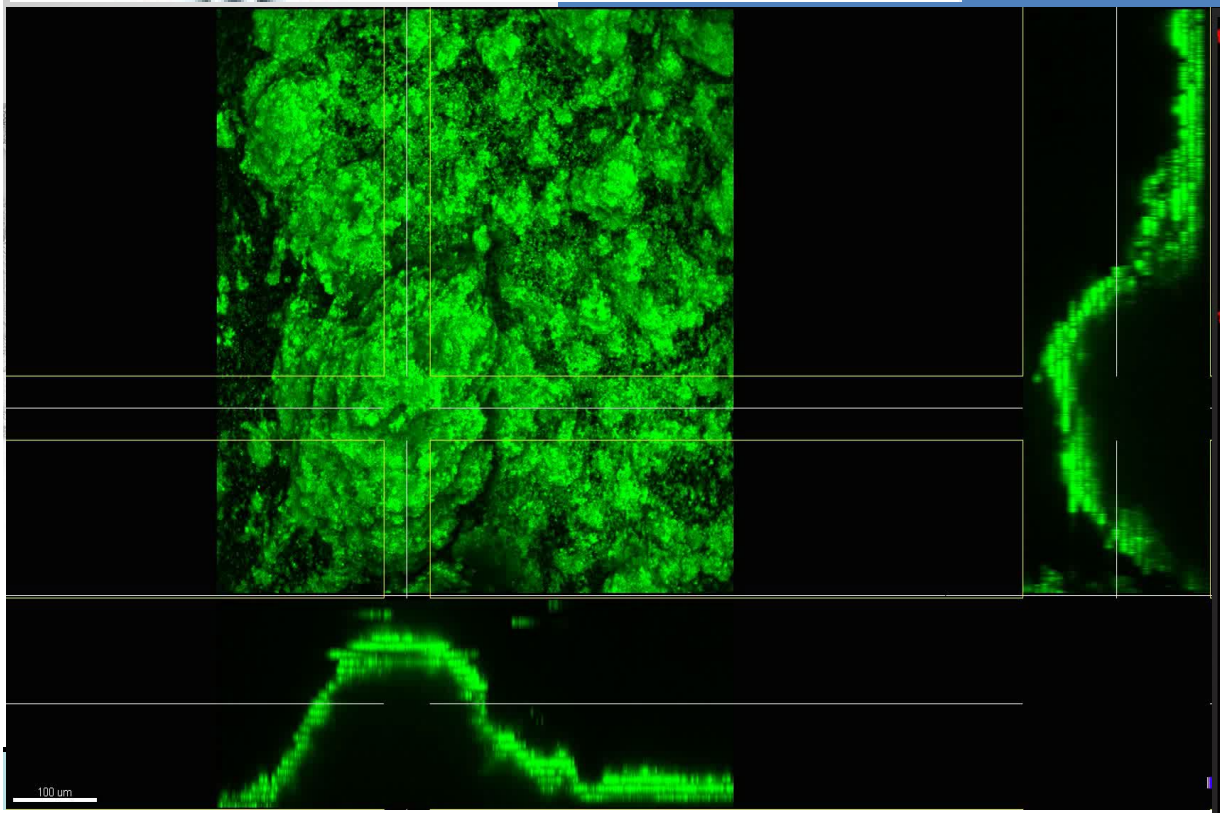
cost EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY

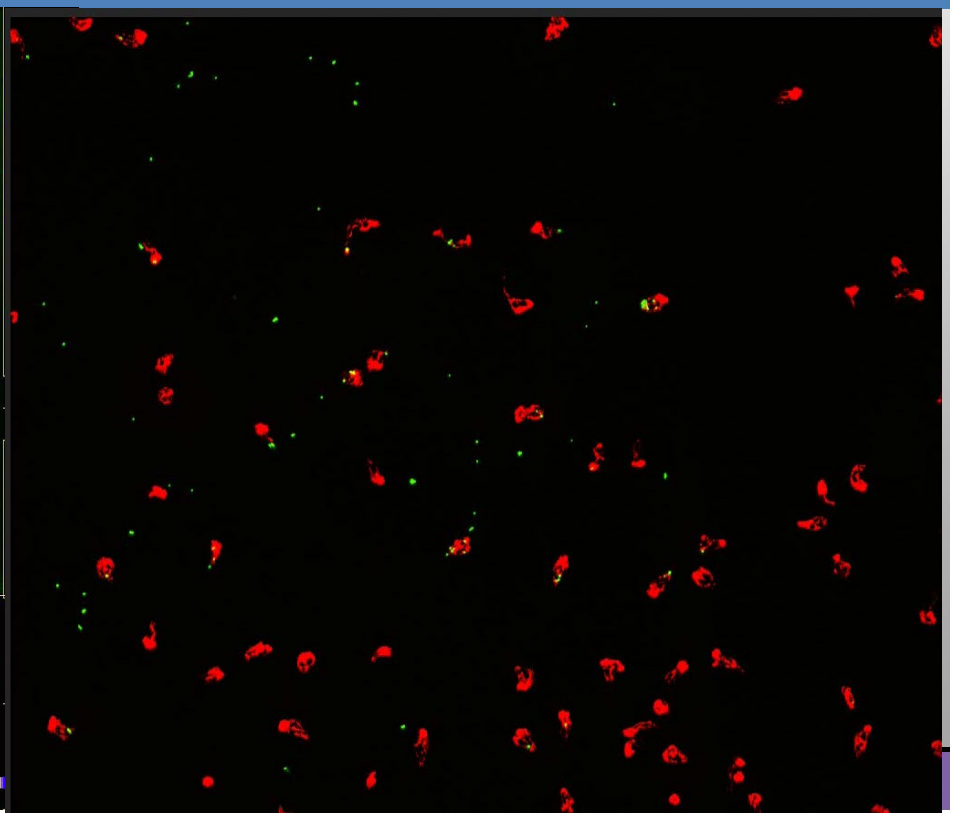
Funded by the European Union

Basic Stats, Repeatabilty & Responsiveness

8 Sept, 2026

Al Parker, PhD, Biostatistician, CBE





About me

Al Parker, PhD, Biostatistician
Center for Biofilm Engineering
Montana State University (MSU)
Bozeman, MT 59715
406-994-5145
albert.parker@montana.edu
www.math.montana.edu/parker

2019 - present	Research Professor, Department of Mathematical Sciences, MSU
2008 - present	Biostatistician, **Center for Biofilm Engineering**, MSU
2008 - present	Senior statistician for Office of Pesticide Programs, Biological and Economic Analysis Division, **US EPA**
2008 - present	Consulting statistician for Committee E35 on Pesticides, Antimicrobials, and Alternative Control Agents, **ASTM International**
2008 - 2014	Consulting statistician for Committee M on Antimicrobial Efficacy Testing, **AOAC International**
2007 - 2010	Post-doctoral Fellow, Bayesian Statistics, University of Auckland
2004 - 2006	Post-doctoral Fellow, Image Analysis, University of California, Santa Cruz

Attributes of an antimicrobial test method: Seven R's

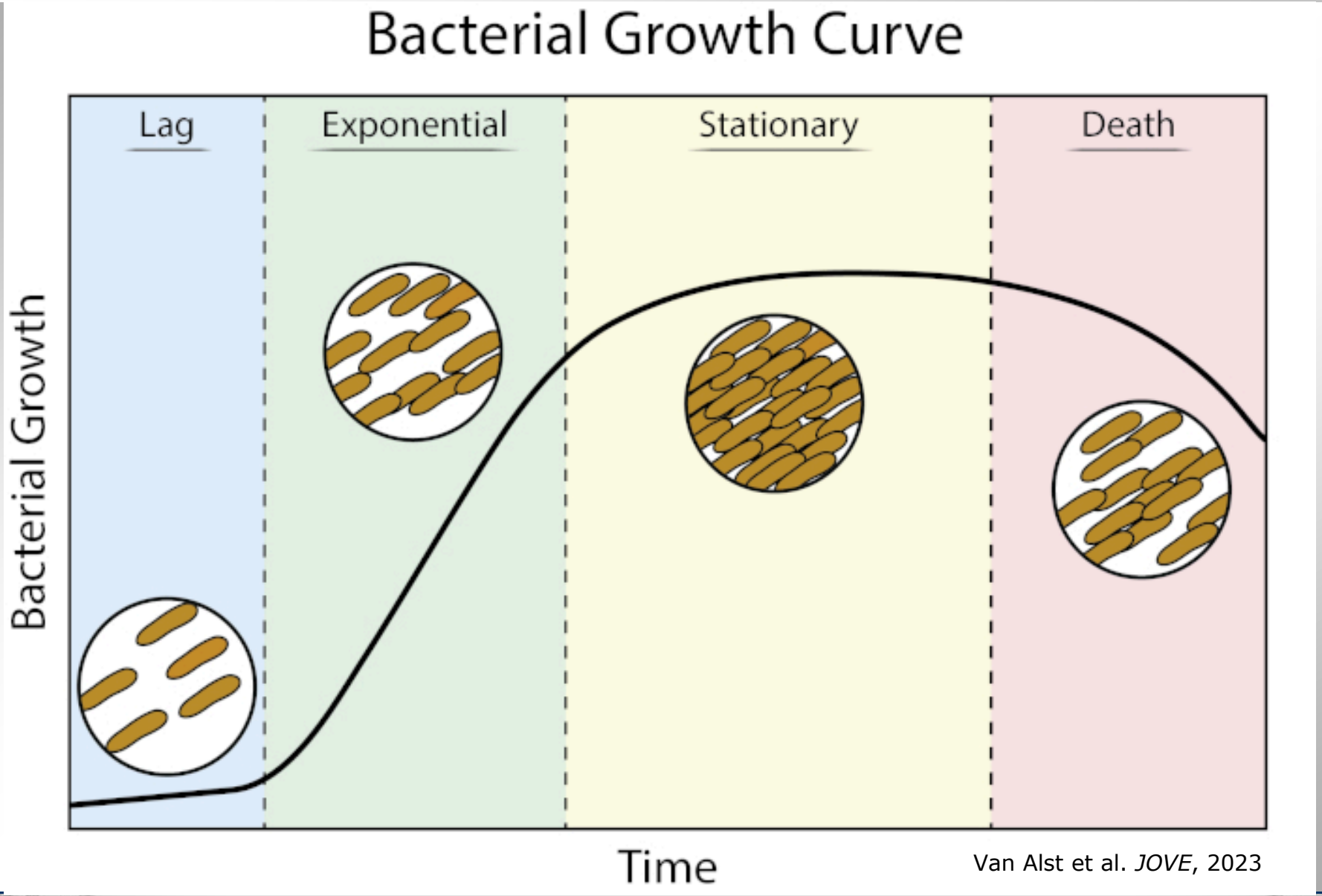
- Relevance
- Reasonableness
- Resemblance (positive controls)
- Repeatability (intra-lab)
- Responsiveness
- Ruggedness
- Reproducibility (inter-lab)

Attributes of an antimicrobial test method: 7 R's + 2

- Relevance
- Reasonableness
- Resemblance (positive controls)
- Repeatability (intra-lab) — ISO's "intermediate precision"
- Responsiveness
- Ruggedness
- Reproducibility (inter-lab)
- Unbiasedness — ISO's "trueness"
- Limit of Detection

Attributes of an antimicrobial test method: Seven R's

Which phase?



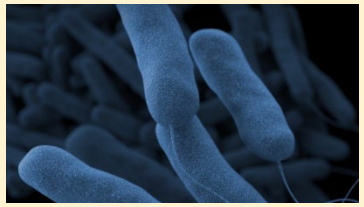
Van Alst et al. *JOVE*, 2023

Attributes of an antimicrobial test method: 7 R's + 2

- Relevance
- Reasonableness
- Resemblance (positive controls)
- Repeatability (intra-lab)
- Responsiveness
- Ruggedness
- Reproducibility (inter-lab)
- Unbiasedness
- Limit of Detection

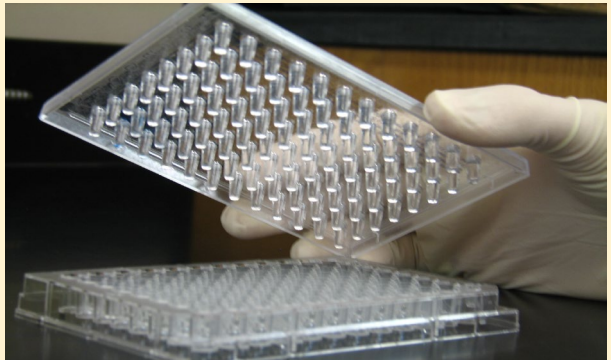
Three R Modules Based on This Talk

Resemblance (Control Repeatability)



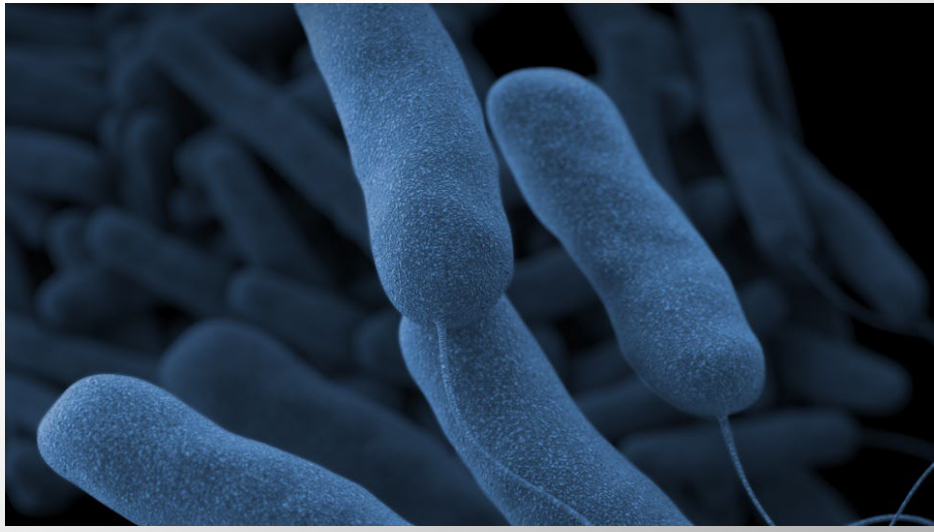
Repeatability

Responsiveness



Attributes of an antimicrobial test method: 7 R's + 2

- Relevance
- Reasonableness
- Resemblance (positive controls)
- Repeatability (intra-lab)
- Responsiveness
- Ruggedness
- Reproducibility (inter-lab)
- Unbiasedness
- Limit of Detection



Legionella biofilms

Bodle et al. *Biofilm*, 2026

Resemblance

The **bio-challenge (i.e., positive controls)** should be repeatable within a lab and reproducible across multiple labs, quantified by a small

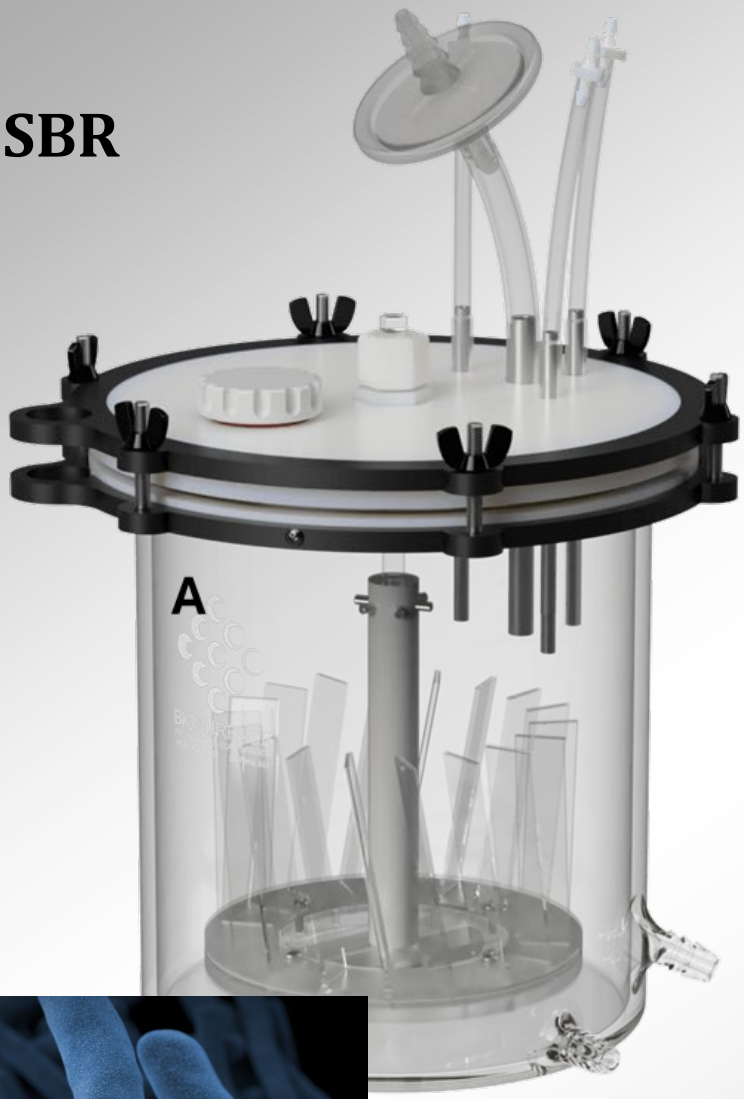
CS_r = repeatability SD

CS_R = reproducibility SD

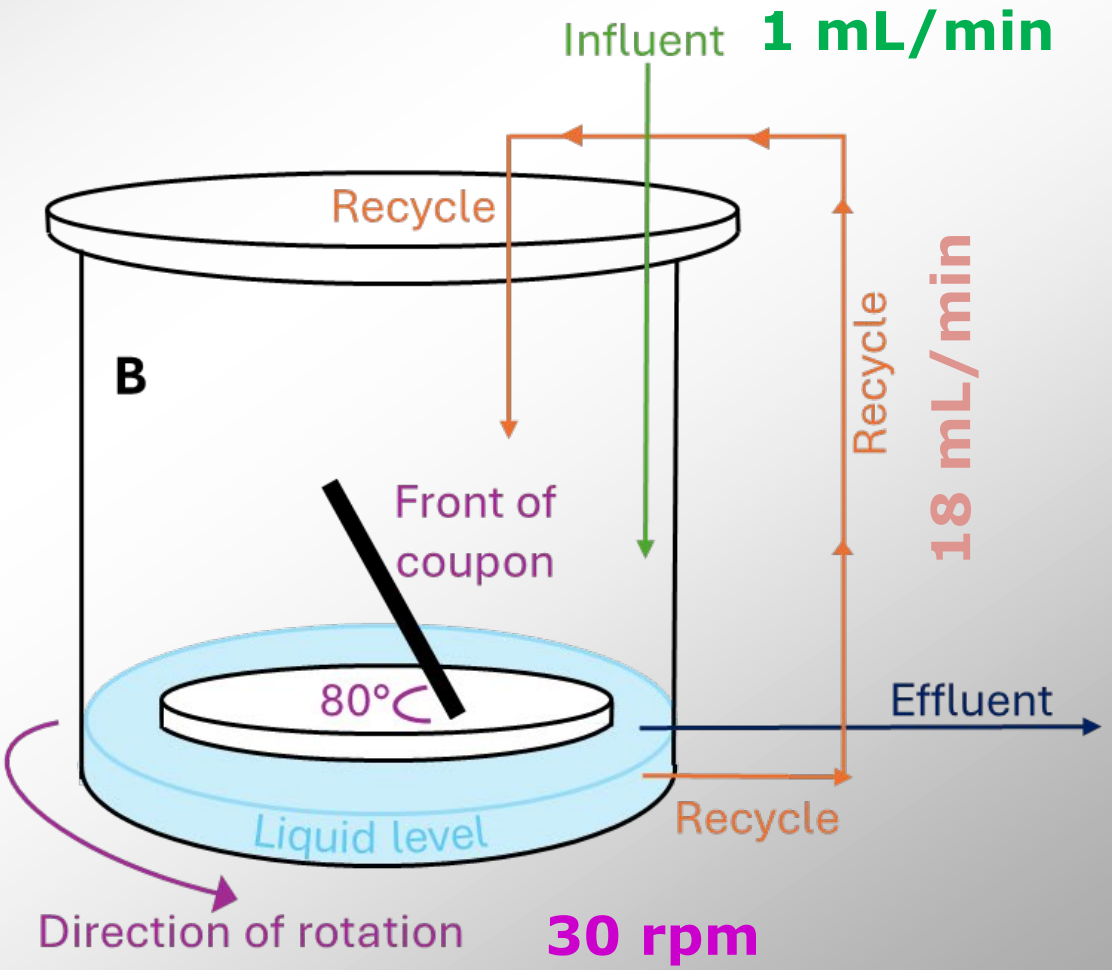
https://www.biofilm.montana.edu/documents-reports/knowledge-sharing-articles.html

Legionella, *Hartmonella* and bacteria found in drinking water formed a biofilm

ISBR



Cooling Tower Model



Legionella densities (D) and "log densities" (LD)

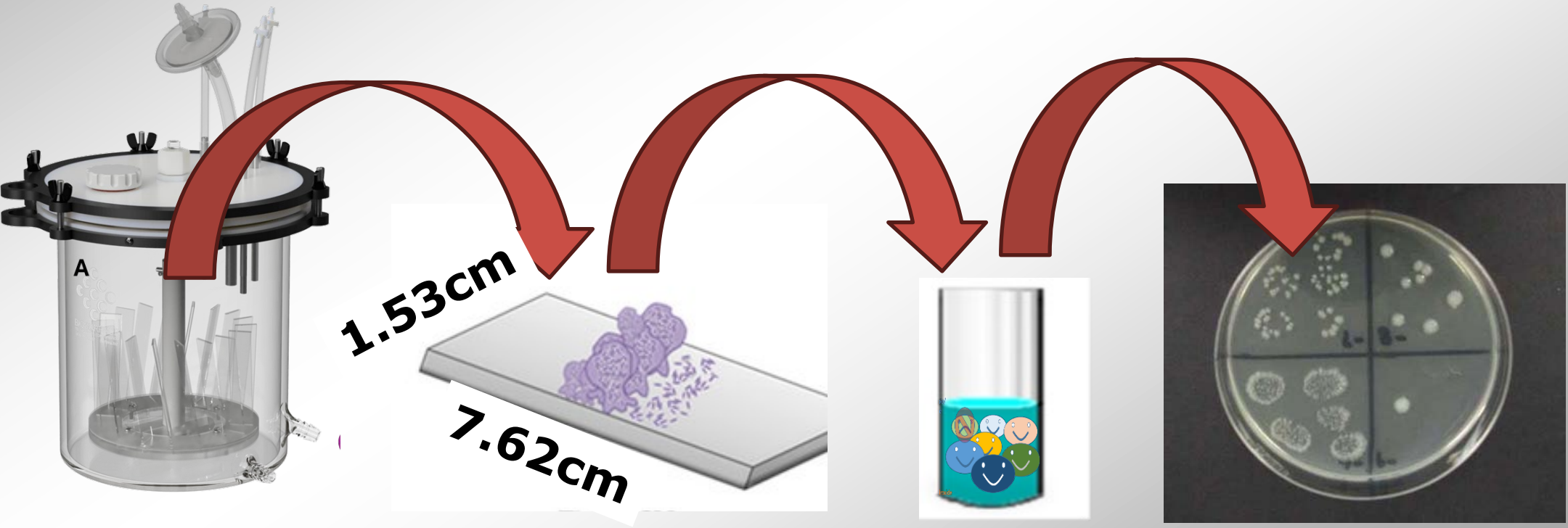
$$D = \frac{CFU \times V}{u \times 10^{k}} = \frac{8 \times 40\text{mL}}{0.1\text{mL} \times 10^{-2}} = 3.2 \times 10^{5} \frac{\text{CFU}}{\text{coupon}}$$

$$= 3.2 \times 10^{5} \frac{\text{CFU}}{1.52 \times 7.62} = 27,365 \frac{\text{CFU}}{\text{cm}^2}$$

$$LD = \log_{10}(27,365) = 4.437 \log_{10}\left(\frac{\text{CFU}}{\text{cm}^2}\right)$$

```
log10(8*40/(.1*10^-2)/
(1.53*7.62))
```

www.cdc.gov/legionella



From Fig 1, Floyd et al, PLOS Pathogens, 2015

V=40mL **u=10x10uL drops**

Attributes of an antimicrobial test method: 7 R's + 2

- Relevance
- Reasonableness
- Resemblance (positive controls)
- Repeatability (intra-lab)
- Responsiveness
- Ruggedness
- Reproducibility (inter-lab)
- Unbiasedness
- Limit of Detection

Limit of Detection in Microbiology

$$\text{“LOD”} = \frac{1CFU \times 40\text{mL}}{0.1\text{mL} \times 10^0} = 400 \frac{\text{CFU}}{\text{coupon}}$$

$$= 400 \frac{\text{CFU}}{1.52 \times 7.62} = 34.3 \frac{\text{CFU}}{\text{cm}^2}$$

$$= 1.54 \log_{10}\left(\frac{\text{CFU}}{\text{cm}^2}\right)$$

www.cdc.gov/legionella

```
log10(1*40/(.1*10^0)/
(1.53*7.62))
```

If there are 400 CFU of *Legionella* on a glass slide in the ISBR then most of the time (51%-63%) they will be detected (CFU>0)

Decrease the LOD by

- **decreasing the total volume *V*=40mL**
- **increasing the plate volume *u*=100uL**

Commonly, CFU=LOD/2 or LOD is subbed in for CFU=0 – but consider better approaches if there are lots of zeros

Limit of Quantification in Microbiology

"LOD" $\neq \frac{14CFU \times 40mL}{0.1mL \times 10^0}$

but maybe ISO considers

"LOQ" $= \frac{14CFU \times 40mL}{0.1mL \times 10^0}$?

Jiang & Parker *European J Parenteral and Pharm Sci* 302 2025.

Data from 1 experiment

Data: viable plate counts of *Legionella* from 3 coupons

Exp	CFU/cm²	$\log_{10}$ (CFU/cm²)
1	27,365	4.437
1	27,365	4.437
1	15,490	4.190

Mean LD = 4.36

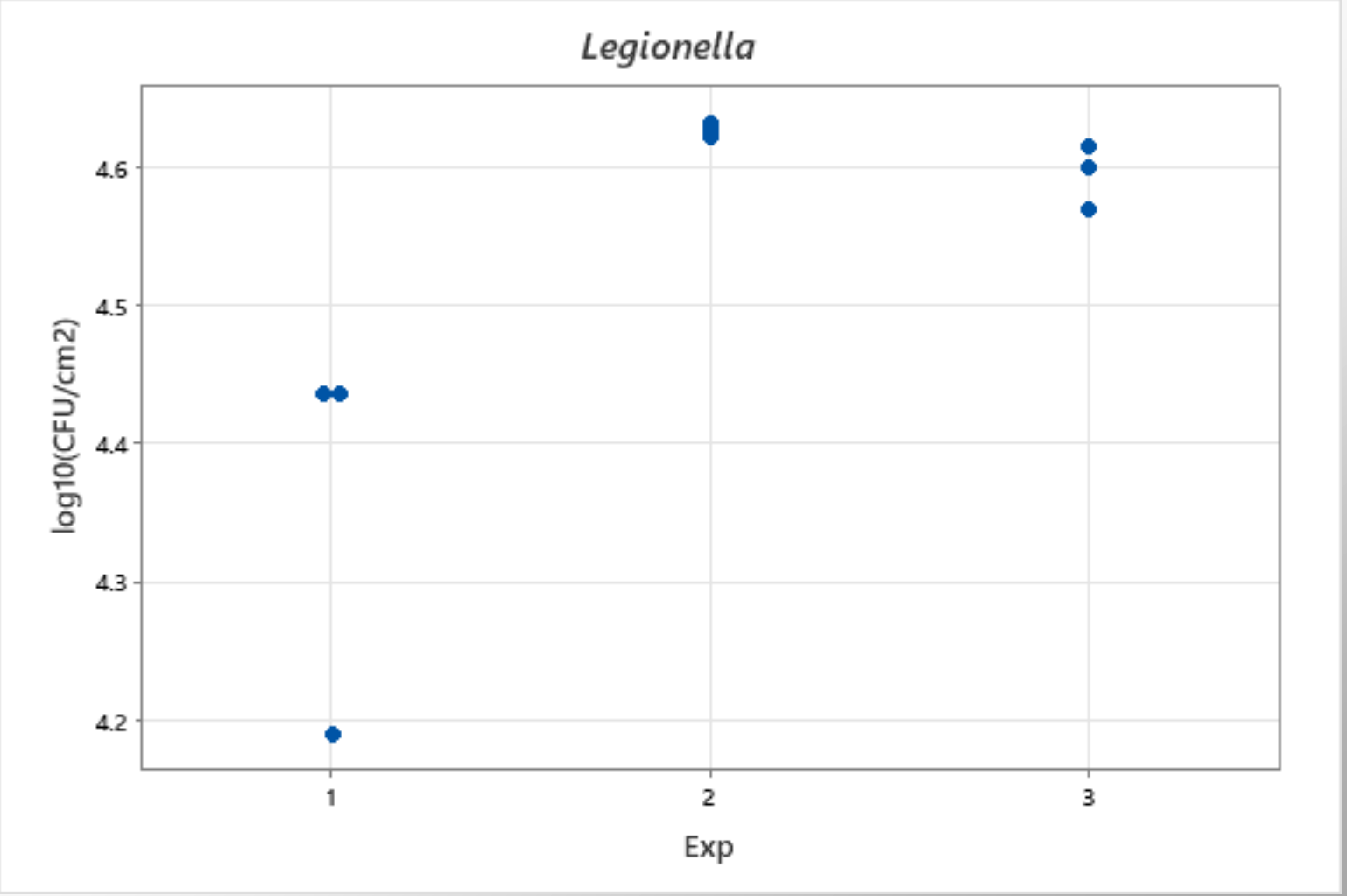
Coupon-to-Coupon SD = 0.143

Perform 3 biological replicates

Exp	log10 (CFU/cm2)
1	4.437
1	4.437
1	4.190
2	4.632
2	4.627
2	4.621
3	4.632
3	4.627
3	4.621

Bodle et al, in prep

First, always graph the data



Then Pool and Analyze the Data (in R)

Exp	log10 (CFU/cm2)	mean log10 (CFU/cm^2)	coupon-to-coupon SD (unit-less)
1	4.437		
1	4.437	4.355	0.143
1	4.190		
2	4.632		
2	4.627	4.627	0.005
2	4.621		
3	4.570		
3	4.616	4.595	0.023
3	4.599		

Mean	SE	CS_r
4.53	0.0859	0.1636
95% *likelihood profile*-CI		
(4.33, 4.72)		

74% Exps, 26% Coupons

0.74 is coupon correlation

Random Effect

```
lmer(log10(CFU) ~ (1|Exp), data = Legionella)
```

Pool and Analyze the Data (in Excel!)

Exp	log10 (CFU/cm2)	mean log10 (CFU/cm^2)	coupon-to-coupon SD (unit-less)
1	4.437		
1	4.437	4.355	0.143
1	4.190		
2	4.632		
2	4.627	4.627	0.005
2	4.621		
3	4.570		
3	4.616	4.595	0.023
3	4.599		

Mean	SE	CS_r
4.53	0.0859	0.1636
95% *t*-CI with DF=2 (WIDER THAN R!)		
(4.16, 4.90)		

Mixed Effects Model in EXCEL:

Overall Mean = AVERAGE(4.355, 4.627, 4.627) = 4.53

$S_r = [V_{exp} + V_{coupon}]^{0.5} = 0.1636$

V_{coupon} = AVERAGE(0.143^2, 0.005^2, 0.023^2)

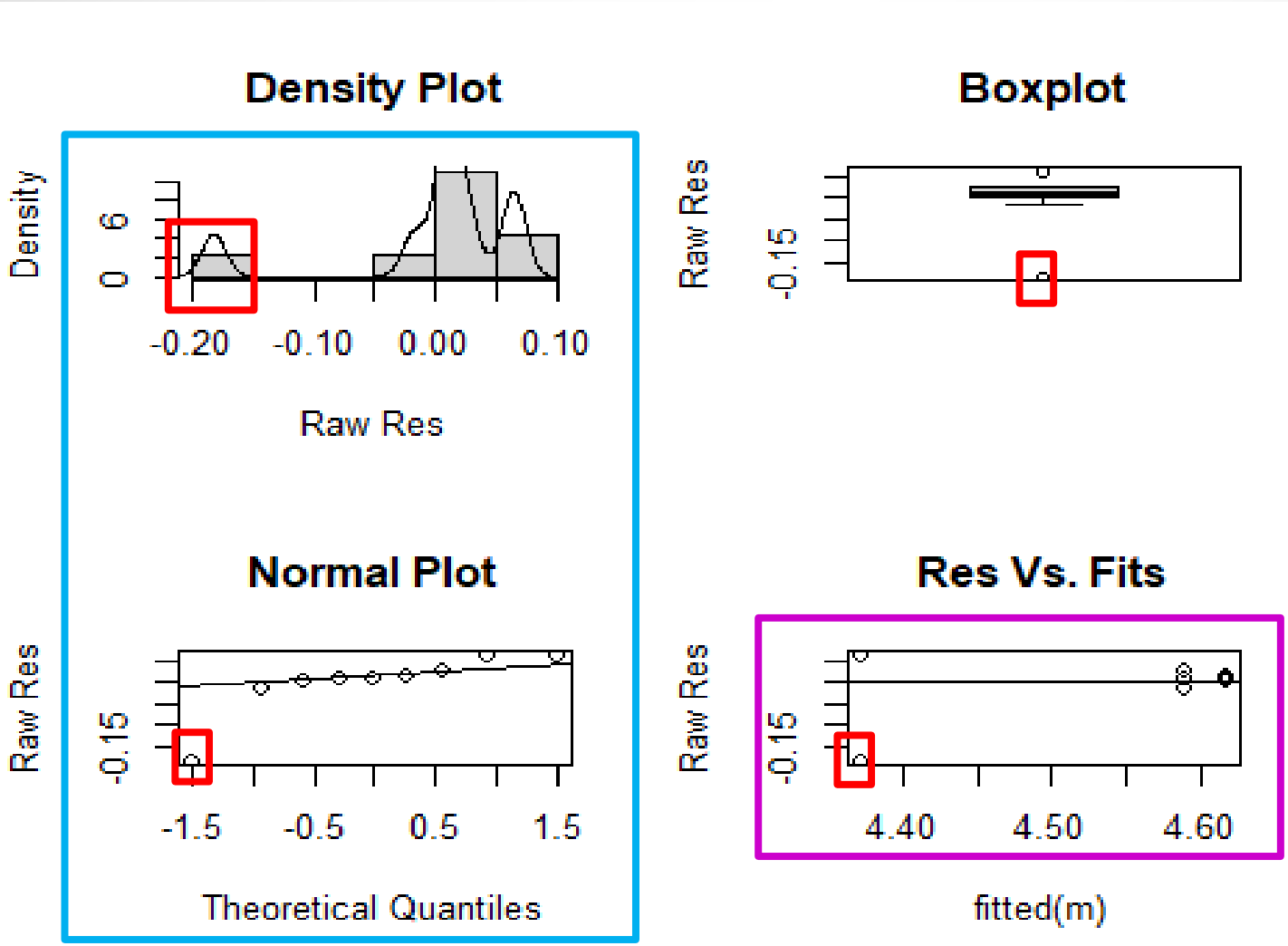
V_{exp} = VAR(4.355, 4.627, 4.627) – $V_{coupon}/3$

correlation of technical replicates = $V_{exp} / S_r^2 = 0.74$

Pooling Data with a Mixed Effects Model in R

```
m = lmer(log10(CFU) ~ (1|Exp), data = Legionella)
diagRES(m)
```

Fitting the model in R, we can check for constant variance, normality and outliers



Assess Outlier Influence by temporarily Removing and Refitting!

```
lmer(log10(CFU) ~ (1|Exp), data = Legionella[-3,])
```

Mean	SE	CS_r
4.53	0.0859	0.1636
95% *likelihood profile*-CI		
(4.33, 4.72)		

Mean	SE	CS_r
4.55	0.0586	0.1021
95% *likelihood profile*-CI		
(4.42, 4.69)		

My opinion (following ASTM guidance) is to leave outliers in a data set unless there is some deviation from protocol to justify removing them.

Good Repeatability of the Controls?

Does CS_r = 0.1636 indicate good repeatability of the untreated *Legionella* controls from the ISBR?

YES: $CS_r \leq 0.5$ indicates good repeatability of untreated biofilm controls

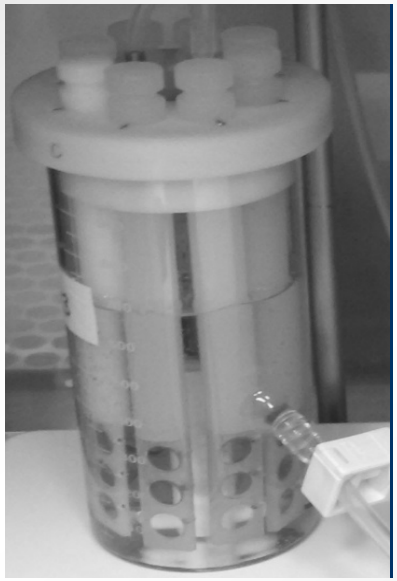
Tilton & Hamilton *JMM* 1999
Parker & Hamilton , CBE KSA-11, 2011

Let's compare to standardized biofilm methods

ISBR

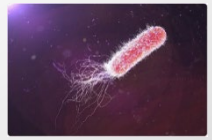
not standardized
biofilm on slides
low flow
low shear

Used by researchers

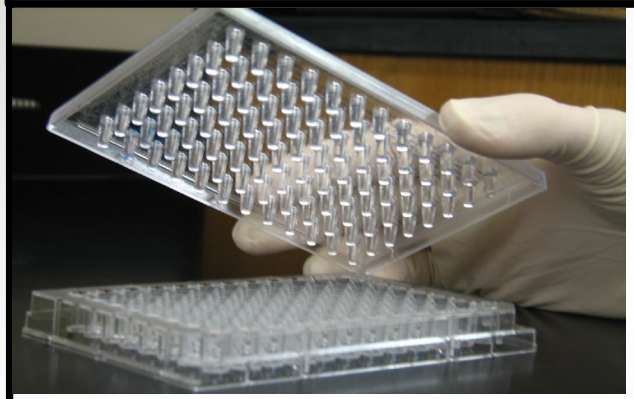


CDC/STM

ASTM E2562/E2871
biofilm on discs
continuous flow
high shear



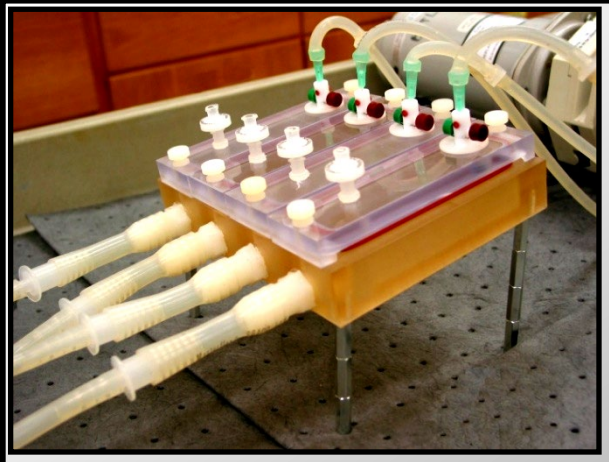
Used by US EPA



MBEC

ASTM E2799
biofilm on pegs
batch
low shear

Used by Health Canada

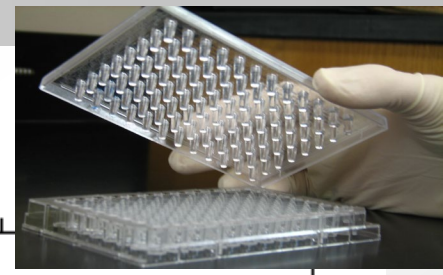


DFR

ASTM E2647
biofilm on slides
plug flow
low shear

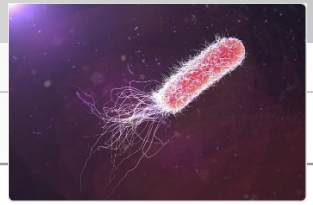
Used by researchers

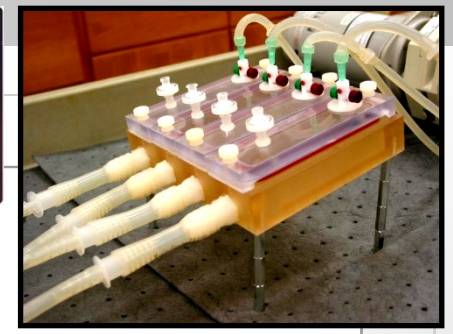
Multi-lab controls for the MBEC (2011)



Parker et al. *JMM*, 2014; ASTM RR:E35-1006, 2011

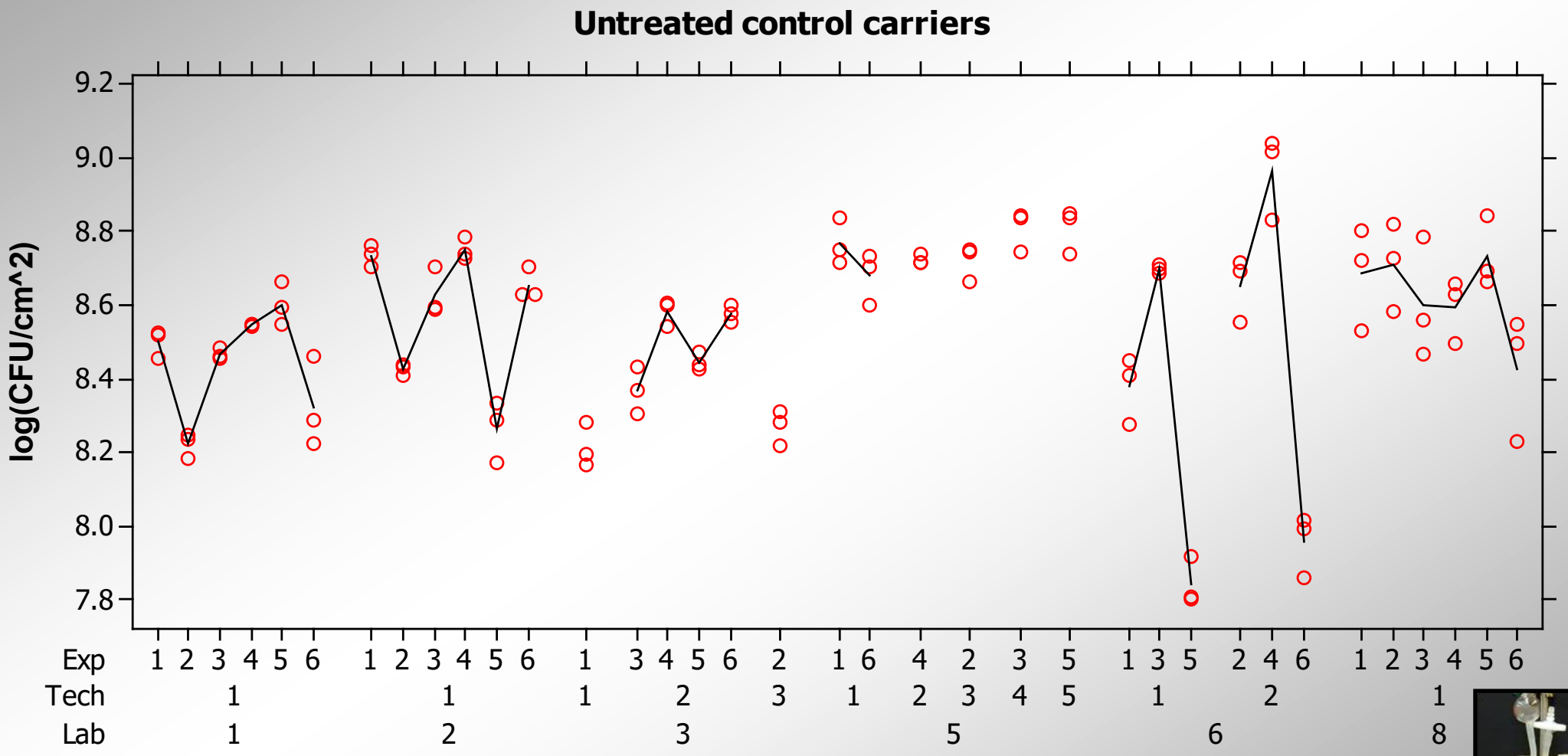
Multi-lab controls for the DFR (2019)

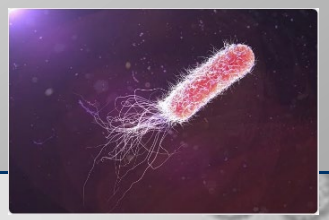




Goeres et al, JMM, 2020; ASTM RR:E35-1015, 2020

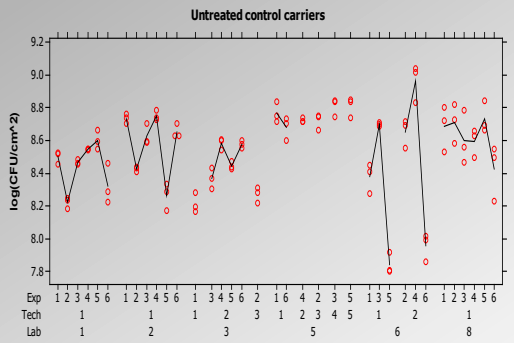
Multi-lab controls for the STM/CDC (2012)

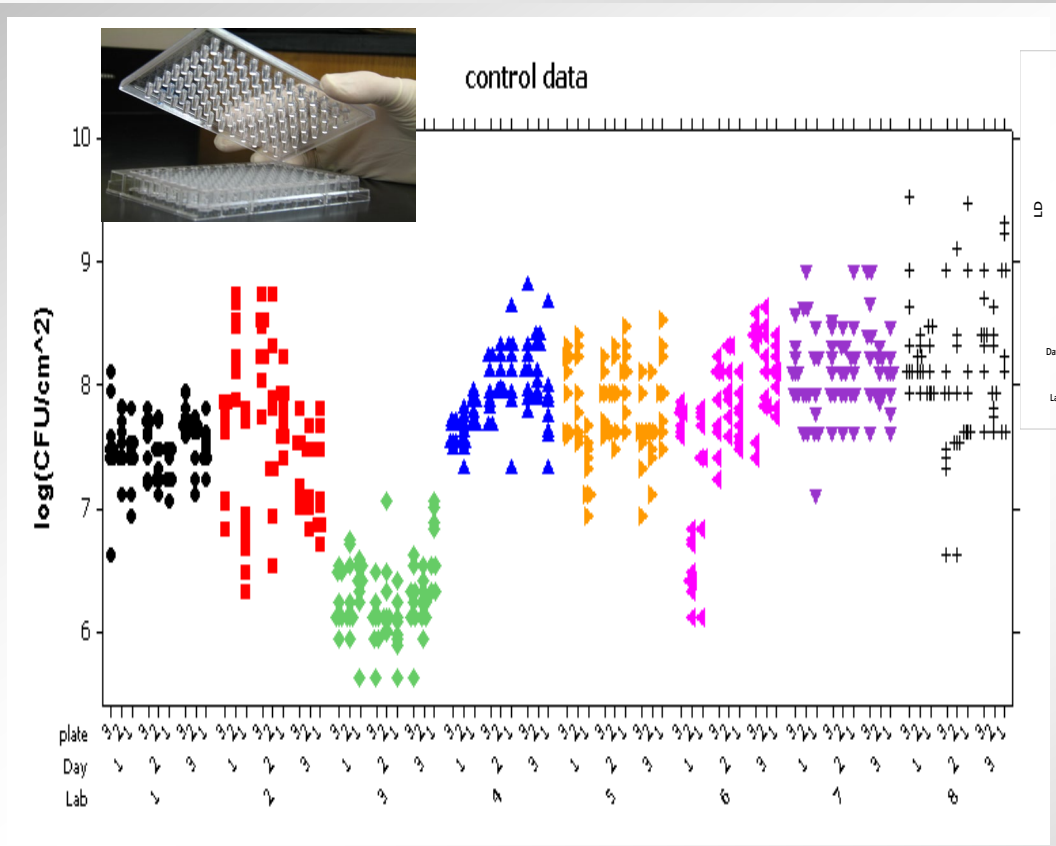


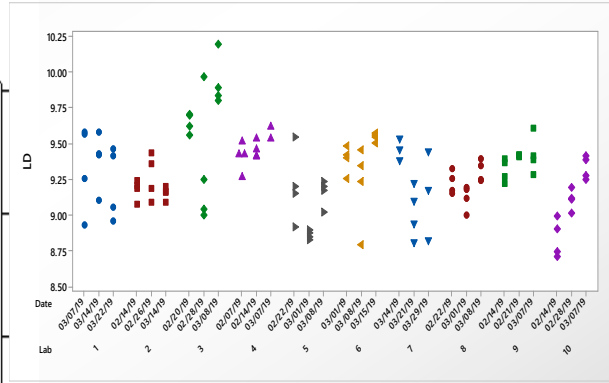


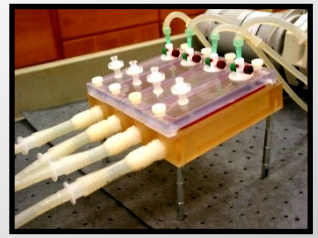
Goeres et al. *JMM*, 2019; ASTM RR:E35-1008, 2013

Comparing controls for the STM, MBEC & DFR









STM:

Repeatability SD: 0.221

Reproducibility SD: 0.244

MBEC

Repeatability SD: 0.325

Reproducibility SD: 0.667

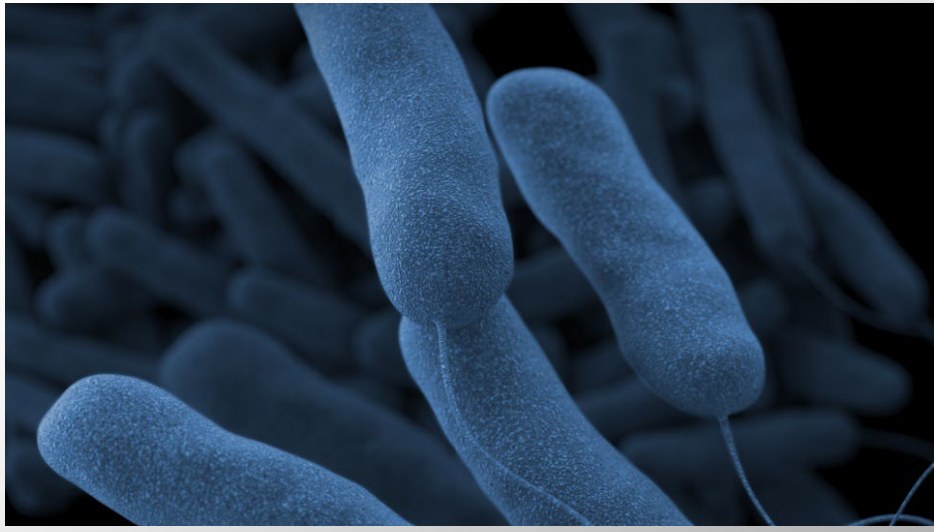
DFR

Repeatability SD: 0.228

Reproducibility SD: 0.268

Attributes of an antimicrobial test method: 7 R's + 2

- Relevance
- Reasonableness
- Resemblance (positive controls)
- Repeatability (intra-lab)
- Responsiveness
- Ruggedness
- Reproducibility (inter-lab)
- Unbiasedness
- Limit of Detection



Repeatability

Independent repeats of the same experiment in the same laboratory produce nearly the same response, as indicated by a small

repeatability standard deviation

S_r = STDEV(Response for each experiment)

http://www.biofilm.montana.edu/content/ksa-sm-10

Repeatability Example

Data: log reduction (LR)

LR = mean(control LDs) – mean(treated LDs)

Analyze LRs instead of the individual
control and treated log(CFU)'s because:

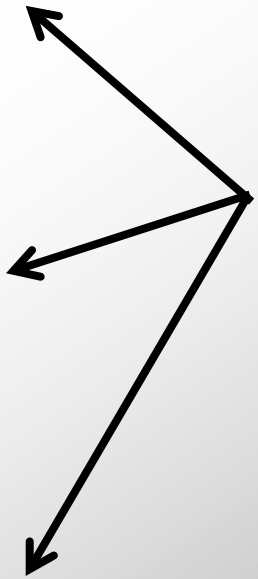
(1) this normalizes each treatment to the control in each experiment;
(2) *t*-tests and ANOVA can be applied to LRs;
(3) usually the control and treated log(CFU)'s exhibit different variability, which violates the homogeneity of variance assumption of an ANOVA model and the variance assumption of a Poisson model.

Here's the same untreated control data …

Exp	log10 (CFU/cm2)	mean log10 (CFU/cm^2)	coupon-to-coupon SD (unit-less)
1	4.437		
1	4.437	4.355	0.143
1	4.190		
2	4.632		
2	4.627	4.627	0.005
2	4.621		
3	4.570		
3	4.616	4.595	0.023
3	4.599		

Repeatability Example (synthetic treated)

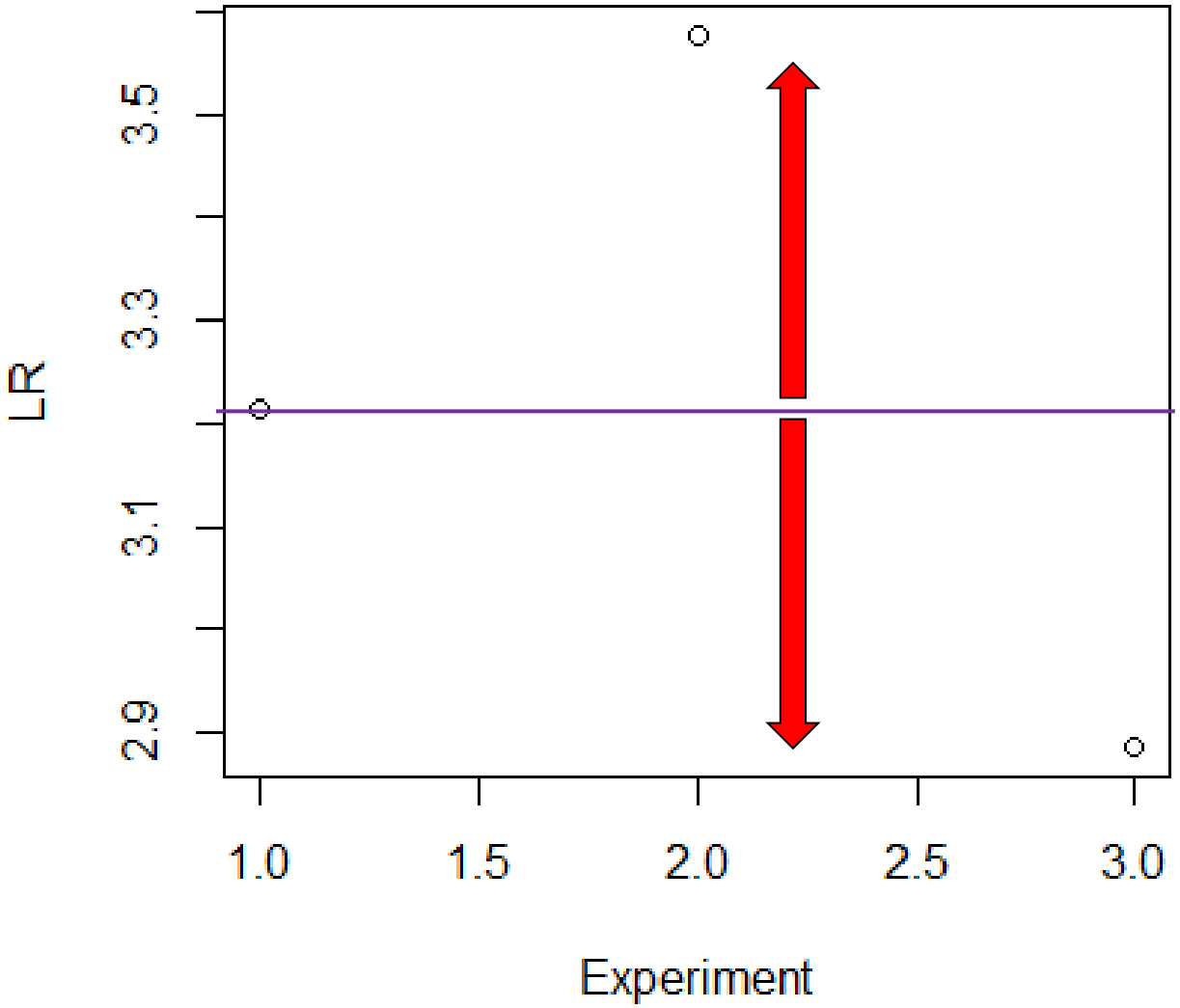
	coupon LD				
Exp	control	treated	Control LD	Treated LD	LR
1	4.437	1.079			
1	4.437	1.301	4.355	1.141	3.214
1	4.190	1.041			
2	4.632	0.903			
2	4.627	1.041	4.627	1.050	3.577
2	4.621	1.204			
3	4.570	1.740			
3	4.616	1.663	4.595	1.709	2.886
3	4.599	1.724			



Since there is no obvious pairing between the controls and treated coupons in each experiment, get 1 LR for each experiment

Mean LR = 3.23

Repeatability Example



Summary Statistics:

1. Mean LR = 3.23
the best guess for the
true mean LR

2. Repeatability SD:

S_r =STDEV(3.21, 3.58, 2.89)
= 0.3456

the typical distance
between
the LR for a
single experiment and the
true mean LR

Good Repeatability of the LRs in 1 Lab?

Does S_r = 0.346 indicate good repeatability of LRs of *Legionella* grown in the ISBR?

YES: $S_r \leq 1.0$ indicates good repeatability of biofilm LRs

Tilton & Hamilton *JMM* 1999
Parker et al *Scientific Reports* 2018

Estimating the True Mean LR in R

```
LR = c(3.214,3.577,2.886)

mean(LR)
## [1] 3.225667
sd(LR)
## [1] 0.3456477

t.test(LR)
## p-value = 0.003806
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  2.367030 4.084303

t.test(LR,alternative = "greater", mu=2)
## p-value = 0.01275
## alternative hypothesis: true mean is greater than 2

t.test(LR,alternative = "greater", mu=3)
## p-value = 0.1877
## alternative hypothesis: true mean is greater than 3
```

Attributes of an antimicrobial test method: 7 R's + 2

- Relevance
- Reasonableness
- Resemblance (positive controls)
- Repeatability (intra-lab)
- Responsiveness
- Ruggedness
- Reproducibility (inter-lab)
- Unbiasedness
- Limit of Detection

Responsiveness

An biofilm test method should be sensitive enough that it can detect important changes in in the biofilm (e.g., due to changes in antimicrobial concentration or contact time).

To assess Responsiveness to Antimicrobials, experiments should be designed so that at least two efficacy levels are included for each antimicrobial agent

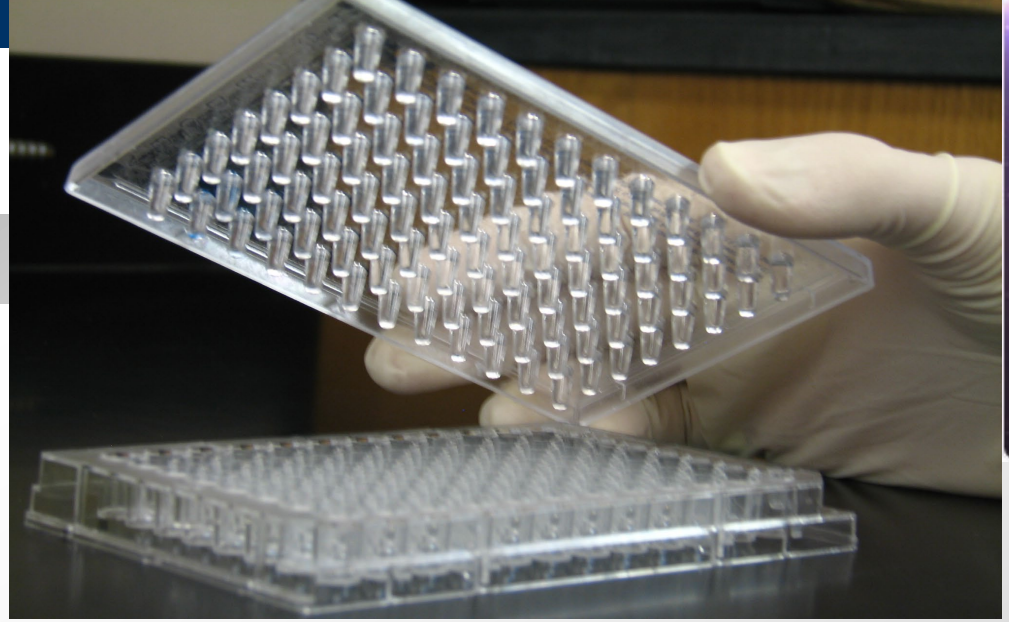
Responsiveness to different efficacy levels

Data: log reduction (LR)

LR = mean(control log(CFU)'s) – mean(disinfected log(CFU)'s)

Analyze LRs instead of the individual control and treated log(CFU)'s because:

(1) this normalizes each treatment to the control in each experiment;
(2) *t*-tests and ANOVA can be applied to LRs;
(3) usually the control and treated log(CFU)'s exhibit different variability, which violates the homogeneity of variance assumption of an ANOVA model and the variance assumption of a Poisson model.

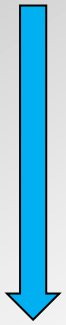


Pseudomonas biofilms

Parker et al. *JMM* 2014

MBEC 96 Well Testing Template

A: High Efficacy

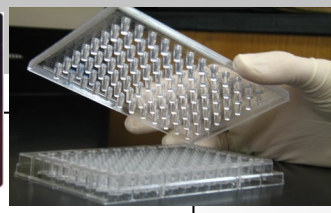


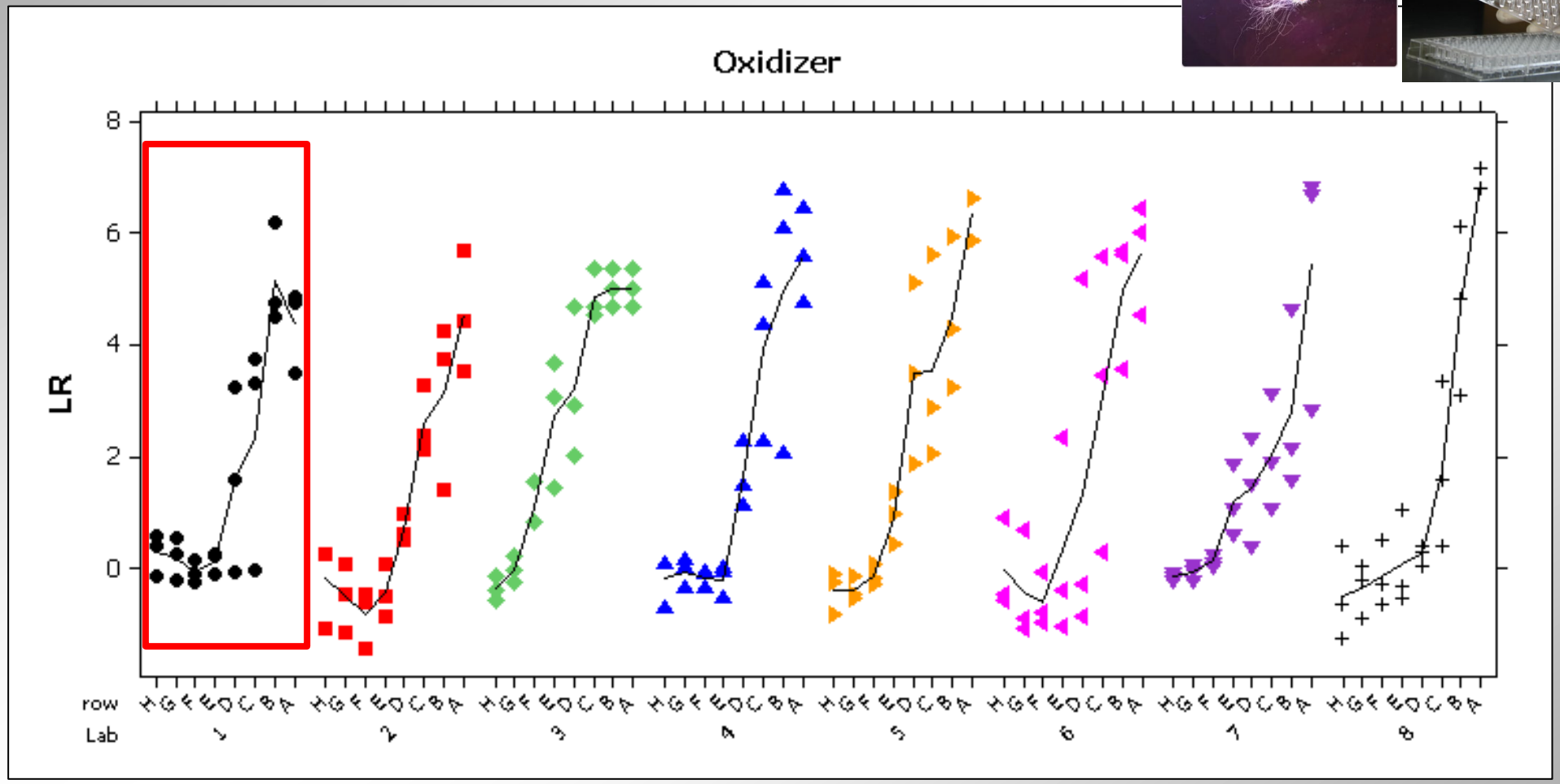
H: Low Efficacy

	1	2	3	4	5	6	7	8	9	10	11	12
A	100	100	100	100	100	50:N	N	GC				SC
B	50	50	50	50	50	50:N	N	GC				SC
C	25	25	25	25	25	50:N	N	GC				SC
D	12.5	12.5	12.5	12.5	12.5	50:N	N	GC				
E	6.25	6.25	6.25	6.25	6.25	50:N	N	GC				
F	3.125	3.125	3.125	3.125	3.125	50:N	N	GC				
G	1.563	1.563	1.563	1.563	1.563	50:N	N	GC				
H	0.781	0.781	0.781	0.781	0.781	50:N	N	GC				

disinfectant **neutralizer test** **control**

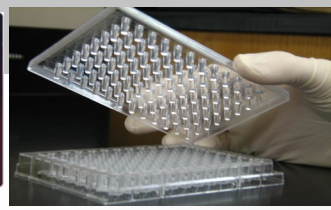
MBEC Responsiveness Across 6 Labs





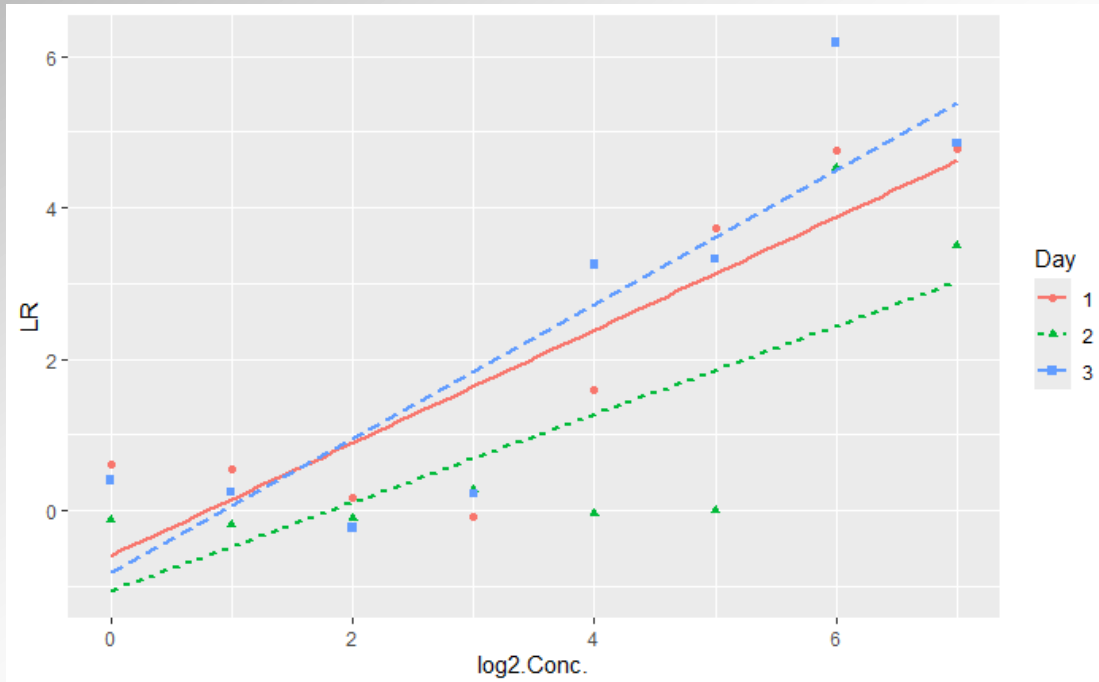
Parker et al. *JMM*, 2014; ASTM RR:E35-1006, 2011

MBEC Responsiveness at Lab 1

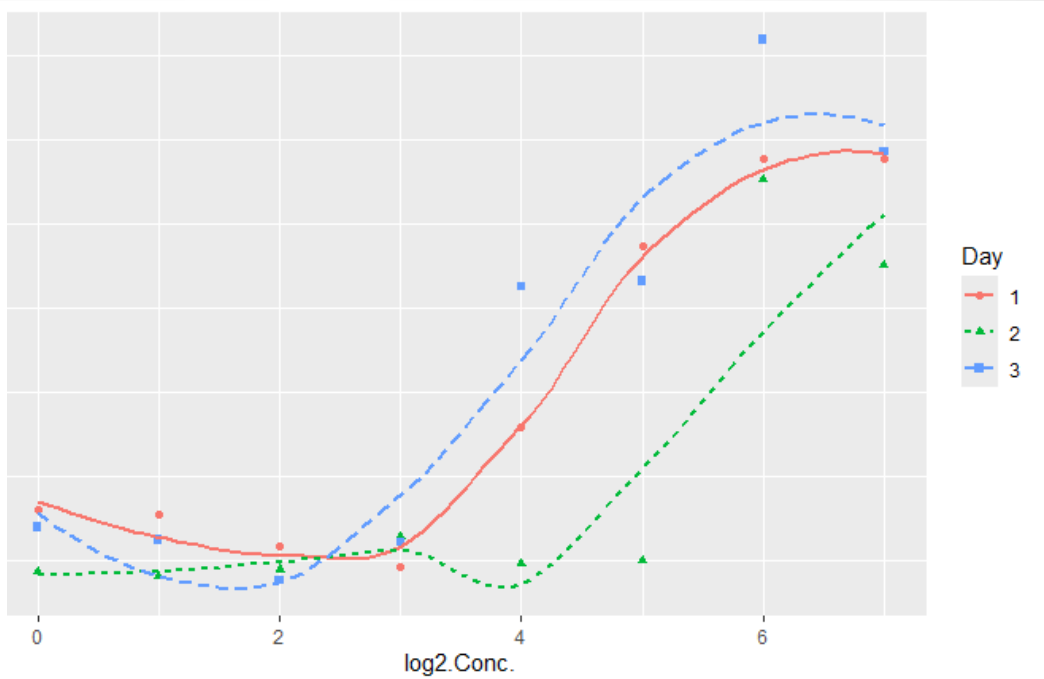


Modeling Increasing Concentrations of an Oxidizer

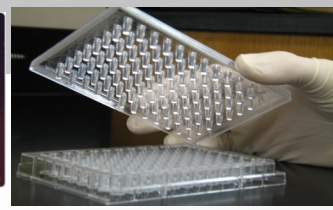
1. Linear Fit



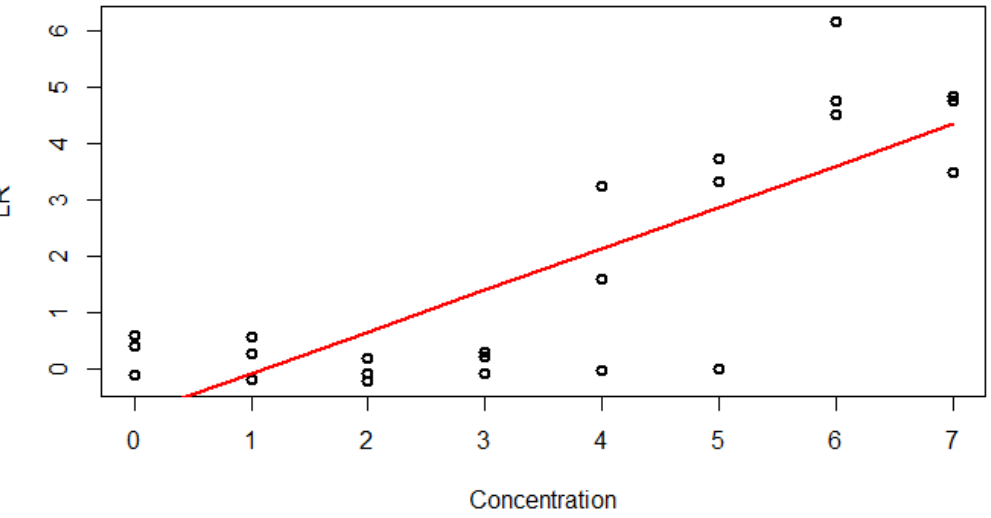
2. Smoother Fit (GAM)



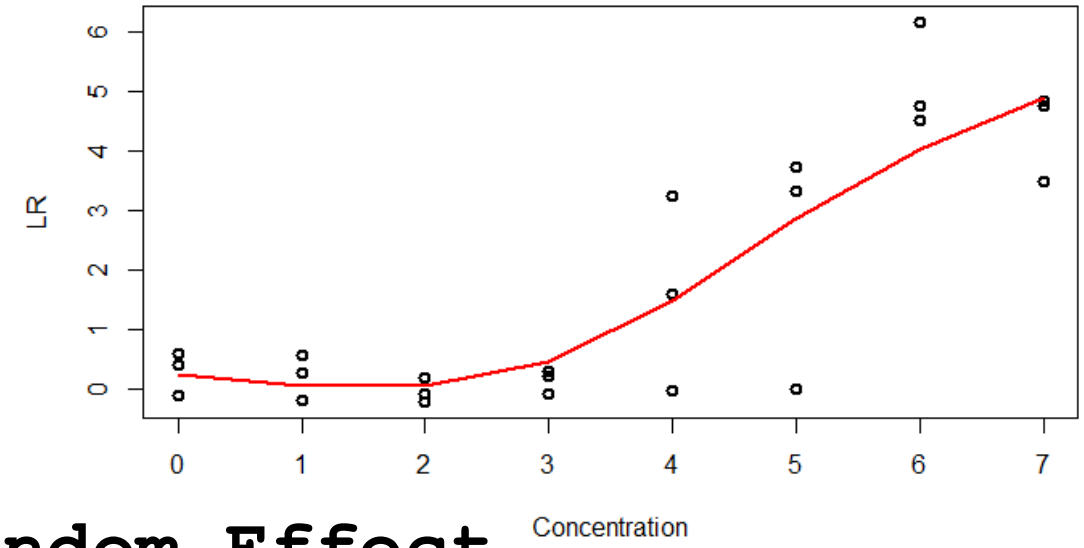
Modeling Responsiveness in R



1. Linear Fit



2. Smoother Fit (GAM)

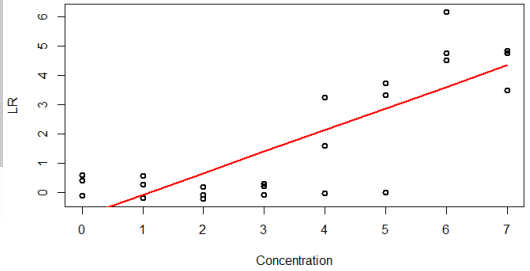


Random Effect

```
lmer(LR~log2.Conc.+(1|Day),
     data = Pseudomonas))
```

```
gamm(LR~ti(log2.Conc.),
     random=list(Day=~1),
     data = Pseudomonas)
```

1. Linear Fit of Responsiveness at Lab 1

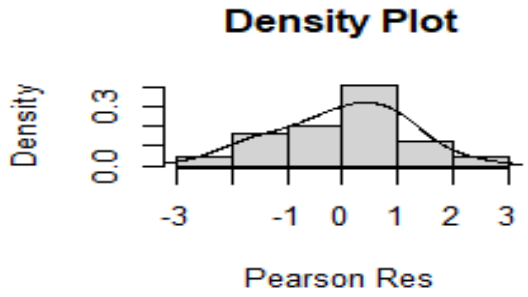


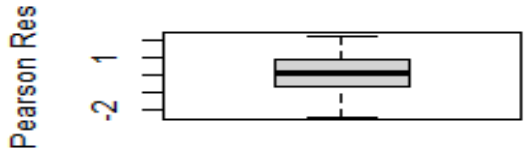

```
mean(dt$log2.Conc.)
## [1] 3.5
m = lmer(LR ~ I(log2.Conc.-mean(log2.Conc.)) + (1|Day),
       data=dt)summary(m)
```

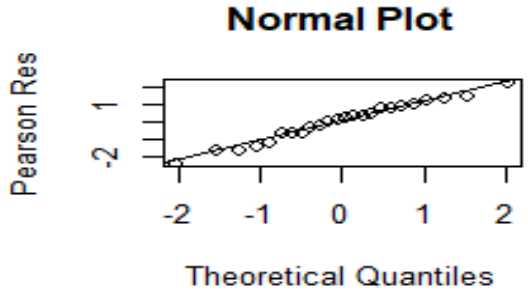
	Estimate	Std. Error	df	tvalue	Pr(>|t|)
(Intercept)	1.7575	0.3963	2.0000	4.434	0.0473 *
I(log2.Conc. - mean(log2.Conc.))	0.7383	0.1068	20.0000	6.915	1.03e-06 ***

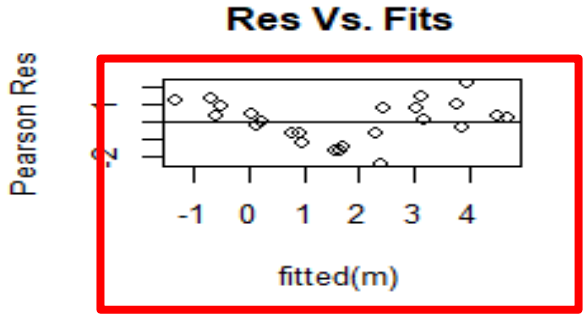
```
diagRES(m)
```

Checking for inappropriate fit

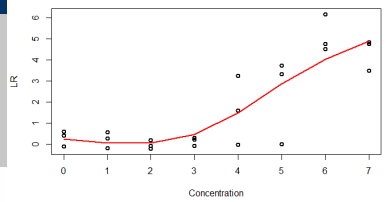








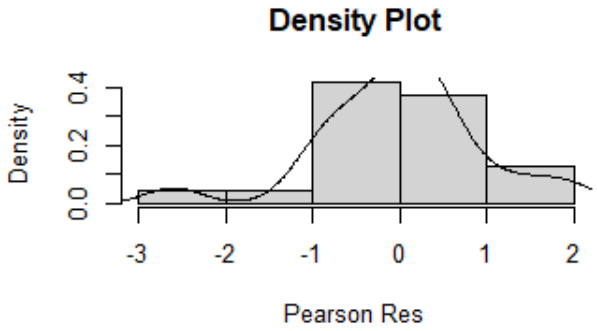
2. Smoother Fit of Responsiveness at Lab 1

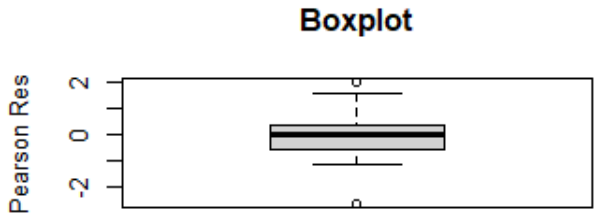


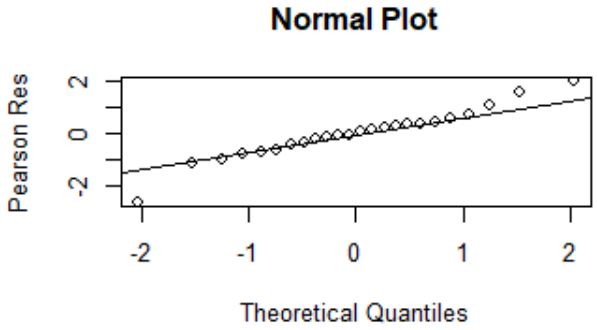

```
g = gamm(LR ~ ti(log2.Conc.),random=list(Day=~1),data=dt)
summary(g$gam)
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)        1.7575     0.3306   5.316 3.28e-05 ***
## Approximate significance of smooth terms:
##                      edf Ref.df     F p-value
## ti(log2.Conc.) 2.893  2.893 32.82  <2e-16 ***
diagRES(m)
```

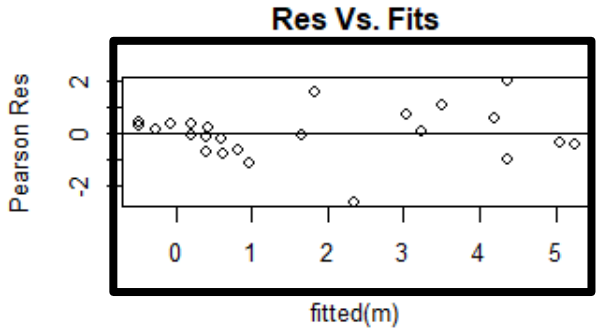
No quantification of responsiveness!

Is the fit better?



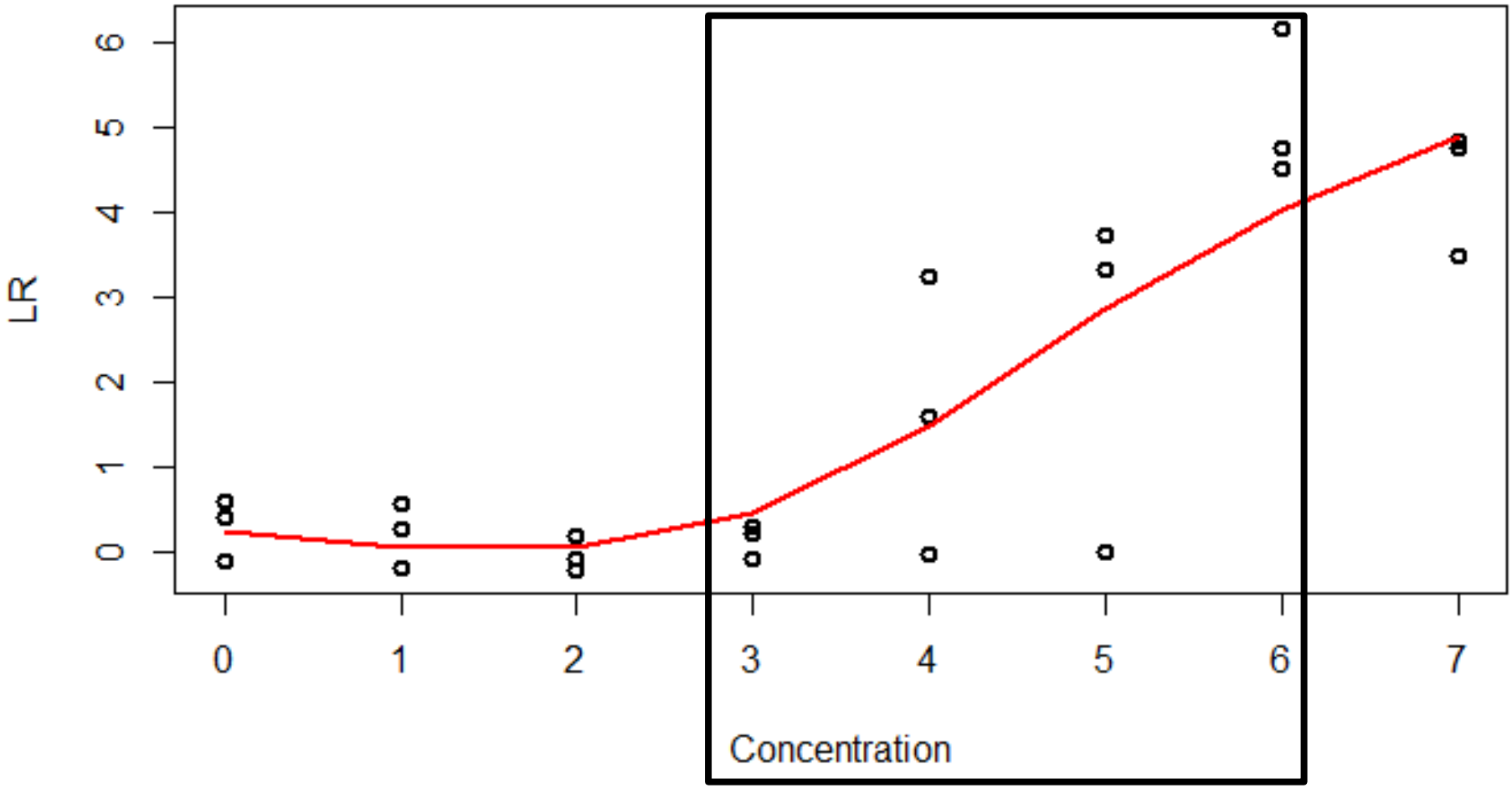






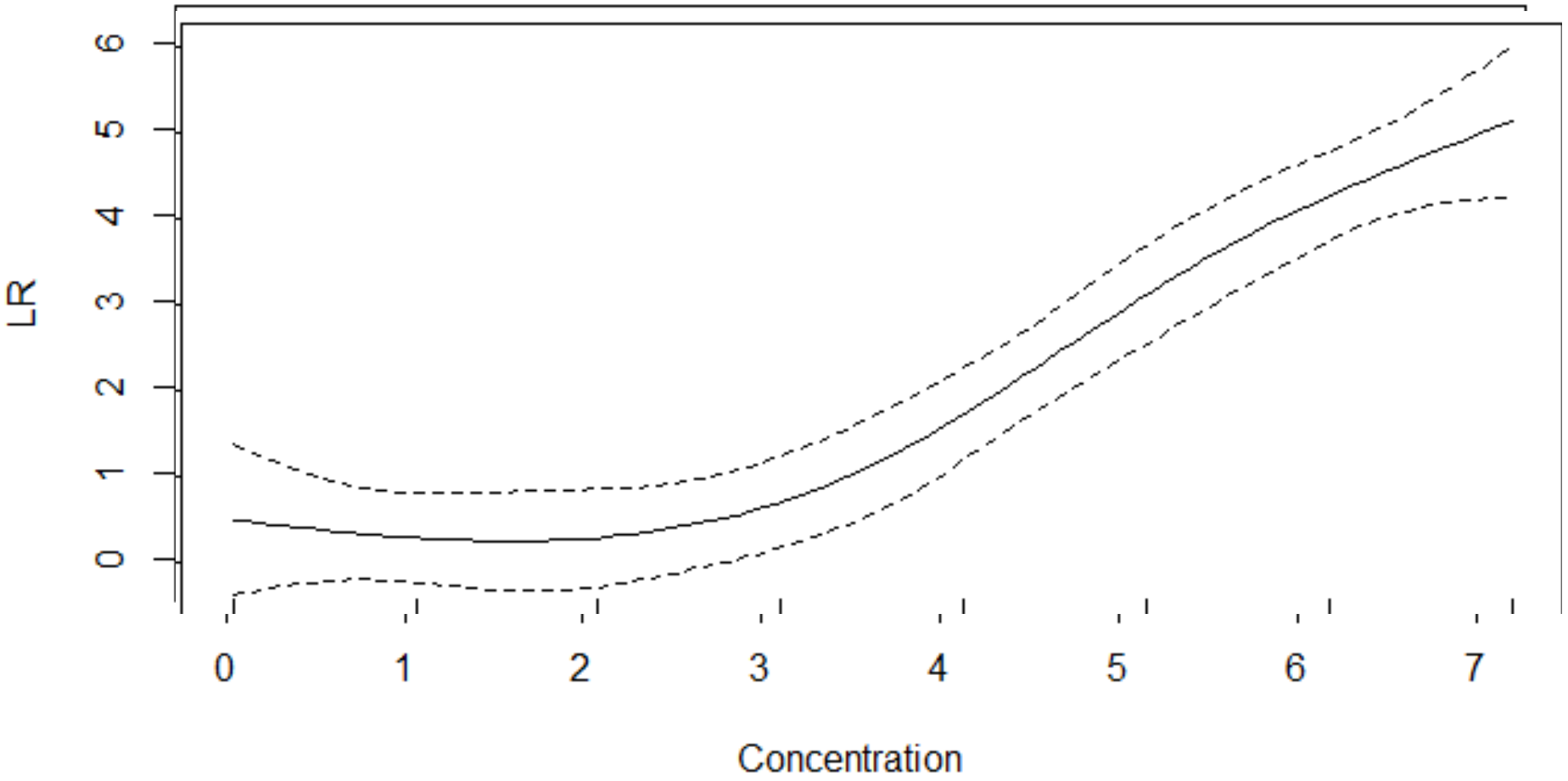
How to quantify responsiveness from a GAM or any ML algorithm?

ANSWER: In Microbiology, look at the Slope of the GAM in the Linear Region – subjective!

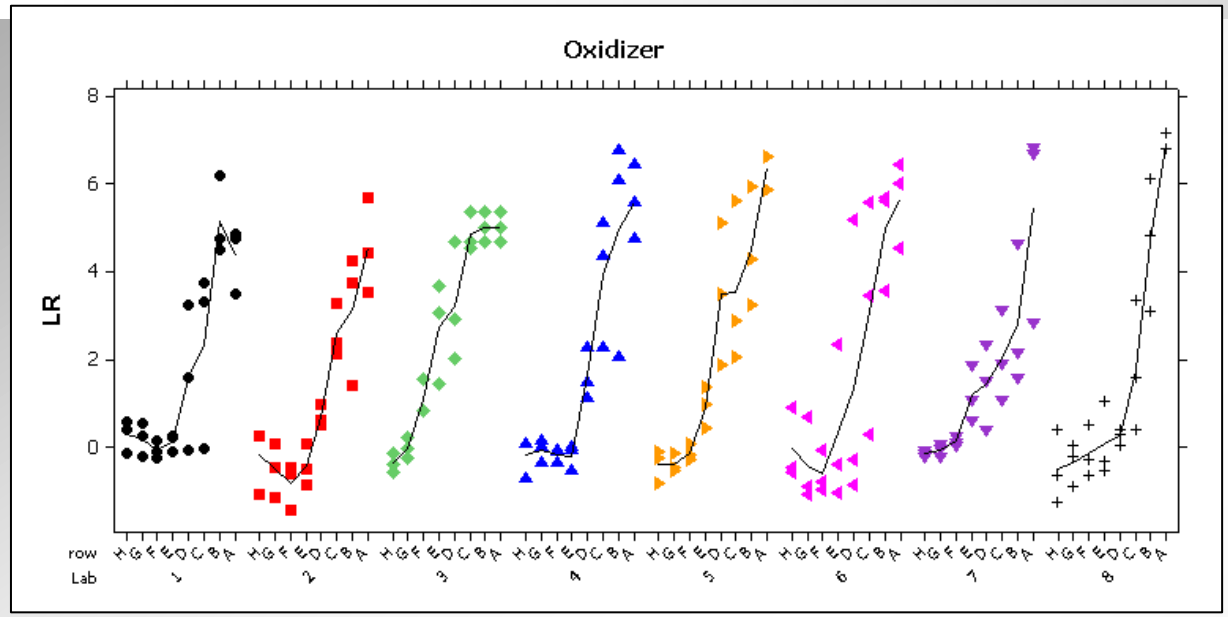


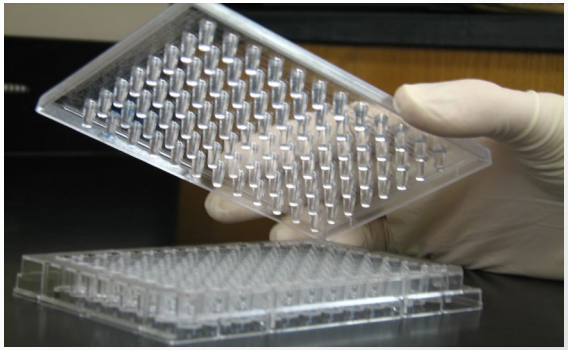
How to quantify uncertainty from a GAM or any ML algorithm?

ANSWER: By putting Error Bars on Predictions



MBEC Responsiveness Across All 6 Labs to 3 Antimicrobials



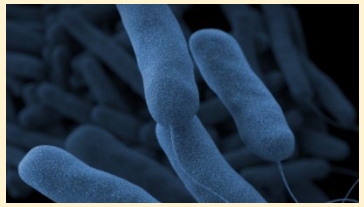


Disinfectant	Trend	SE	*p*-value
Non-chlorine oxidizer	0.87	0.0379	<0.00005
Phenol	0.87	0.0376	<0.00005
Quat	0.50	0.0275	<0.00005

Parker et al. *JMM*, 2014; ASTM RR:E35-1006, 2011

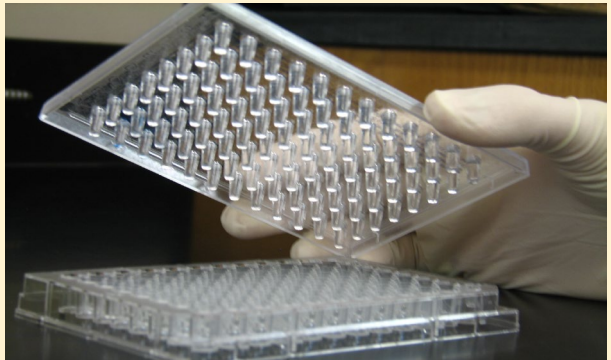
Three R Modules Based on This Talk

Resemblance (Control Repeatability)



Repeatability

Responsiveness



Any questions?