MONTANA STATE UNIVERSITY

# Center for Biofilm Engineering

Biofilm Regulatory Toolbox

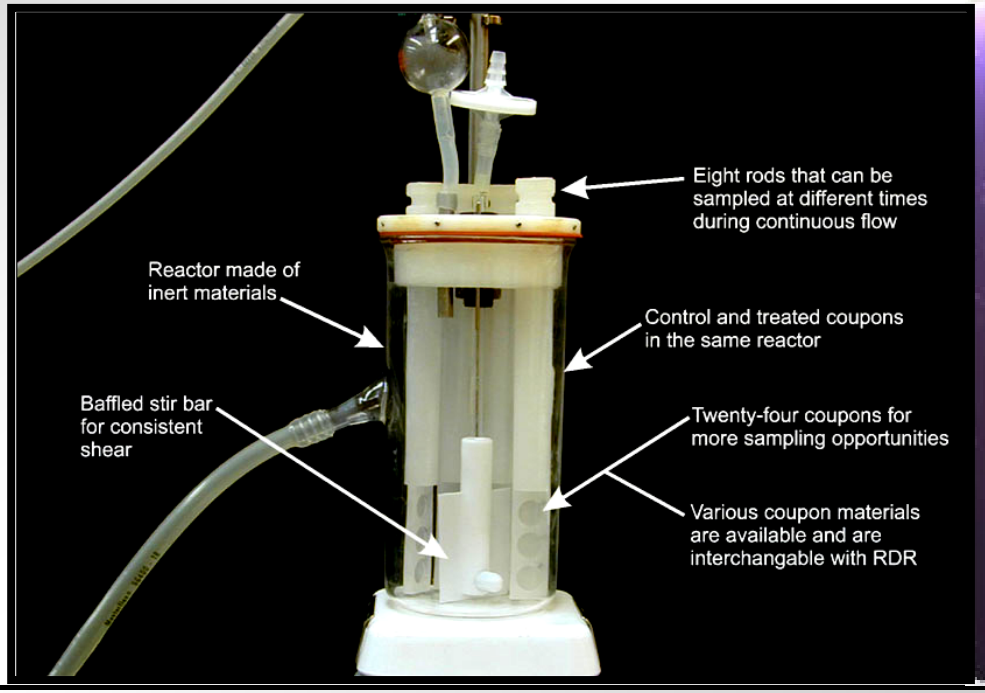
cost EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY

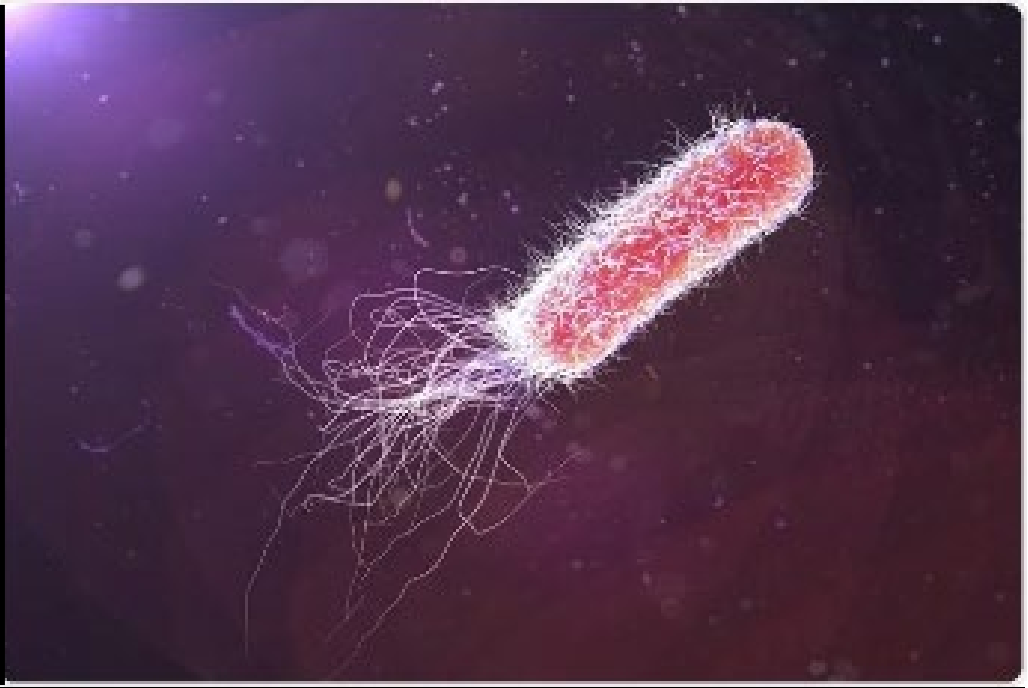
Funded by the European Union

## Ruggedness & Reproducibility

**8 Sept, 2026**

**Al Parker, PhD, Biostatistician, CBE**



# Attributes of an antimicrobial test method: Seven R's

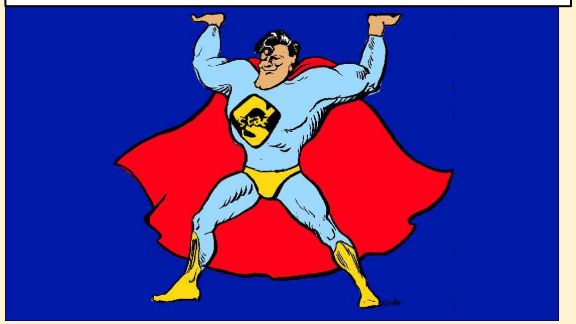
- Relevance
- Reasonableness
- Resemblance (positive controls)
- Repeatability (intra-lab)
- Responsiveness
- Ruggedness
- Reproducibility (inter-lab)

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

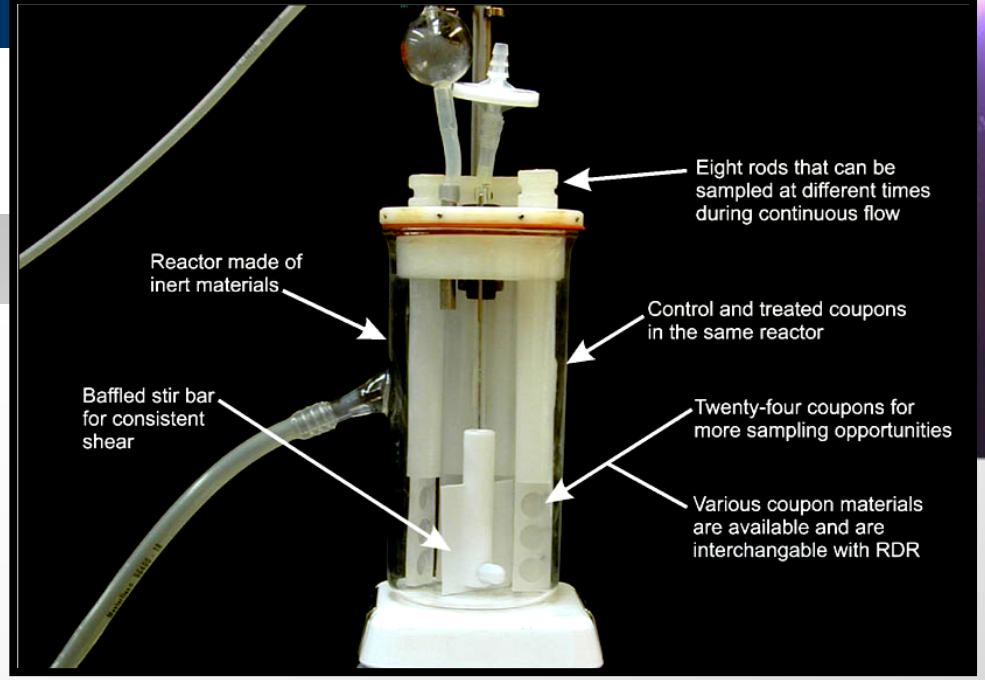
- Relevance
- Reasonableness
- Resemblance (positive controls)
- Repeatability (intra-lab)
- Responsiveness
- Ruggedness
- Reproducibility (inter-lab)
- Unbiasedness
- Limit of Detection

# Two R Modules Based on This Talk

## Ruggedness



## Reproducibility




# *Pseudomonas* biofilms

Goeres et al. *JMM* 2019

# 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

# Ruggedness

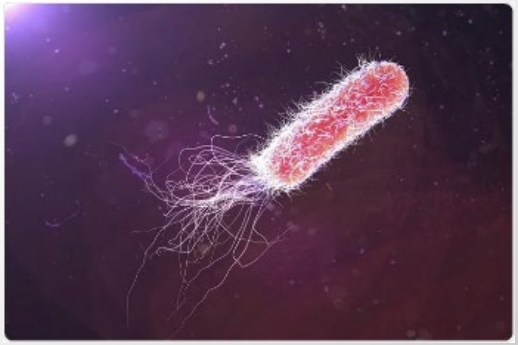
A standard laboratory method is said to be rugged if the outcome is unaffected by slight departures from the protocol.

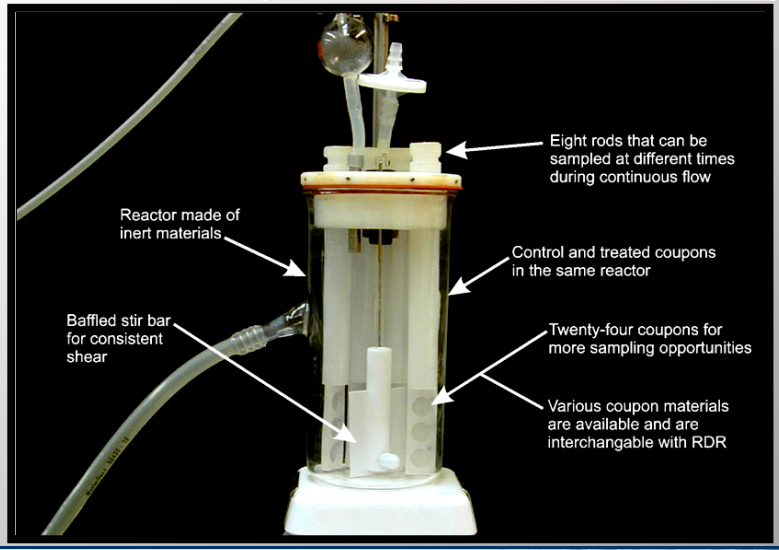
*mixed effects* regression or ANOVA

# STM Ruggedness Testing

## Parameters in the protocol:

- Sonication Power: 20, 80, 200 watts
- Sonication Duration: 25, 30, 35 minutes
- Degas: Yes or No

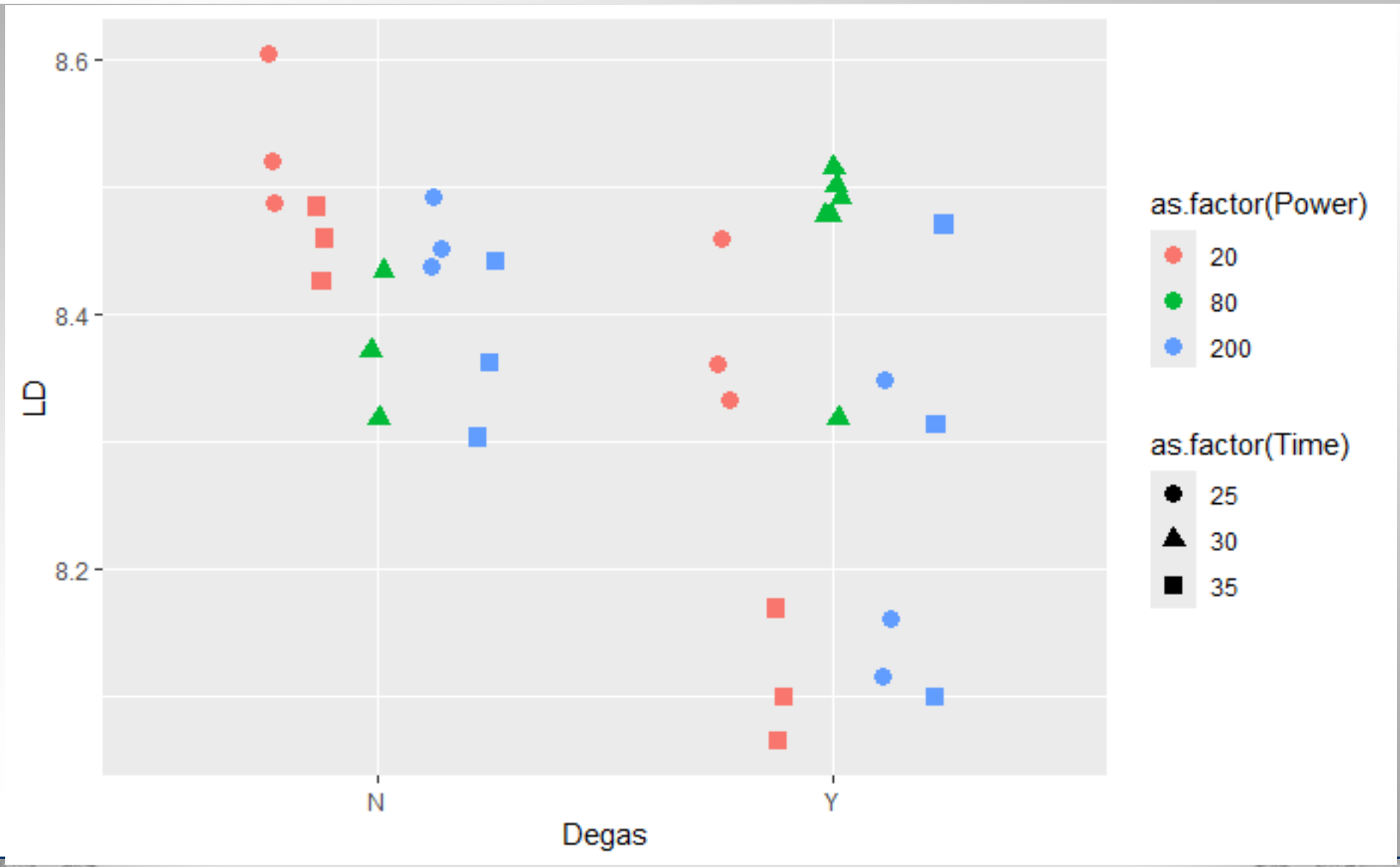




# STM Ruggedness Test Design

| Experiment # | Power (W) | Time (s) | Degas |
|---|---|---|---|
| 1 | 20 | 35 | Yes |
| 1 | 200 | 25 | Yes |
| 1 | 20 | 25 | Yes |
| **1** | **80** | **30** | **Yes** |
| 1 | 200 | 35 | Yes |
| 2 | 200 | 35 | No |
| **2** | **80** | **30** | **No** |
| 2 | 20 | 35 | No |
| 2 | 20 | 25 | No |
| 2 | 200 | 25 | No |
| **2** | **80** | **30** | **Yes** |

# STM Ruggedness Testing of the Controls



# STM Ruggedness Regression Equations

When no degassing:

$LD = 8.45 - 0.0004(Power - 80) - 0.0085(Duration - 30) - 6x10^{-6}(Power - 80)*(Duration - 30)$

When degassed:

$LD = 8.33 - 0.0002(Power - 80) - 0.0153(Duration - 30) + 0.0002(Power - 80)*(Duration - 30)$

- All terms are small and not of practical importance inside the range of values tested
- None of the terms are statistically significantly larger than zero

**So the STM is Rugged** to

- Sonication whether Degas is on or off,
- Sonication Power changes from 20W – 200W
- Sonication Time from 25s – 35s

The regression model allows one to:

- quantifiably predict how deviations from the SOP affect the control LDs
- predict STM Control Reproducibility (if the variation across Degas, Power and Time in the ruggedness tests represent variation across labs)

# STM Ruggedness Regression in R

Start with the most complex linear model with **all interactions**

```
m = lmer(LD ~ Degas*I(Power-80)*I(Time - 30) + (1|Exp),data=d)
summary(m)

##Fixed effects:
                                      Estimate  Std. Error          df  t value Pr(>|t|)
##(Intercept)                          8.366e+00  9.861e-02  1.119e+00  84.839  0.00455 **
##DegasY                               1.319e-02  6.519e-02  2.317e+01   0.202  0.84145
##I(Power - 80)                       -3.900e-04  3.397e-04  2.400e+01  -1.148  0.26228
##I(Time - 30)                        -8.460e-03  6.517e-03  2.400e+01  -1.298  0.20655
##DegasY:I(Power - 80)                 2.532e-04  4.802e-04  2.404e+01   0.527  0.60281
##DegasY:I(Time - 30)                 -6.903e-03  9.216e-03  2.400e+01  -0.749  0.46110
##I(Power - 80):I(Time - 30)          -5.823e-06  6.869e-05  2.400e+01  -0.085  0.93315
##DegasY:I(Power - 80):I(Time - 30)    2.054e-04  9.715e-05  2.400e+01   2.114  0.04510 *
```

# STM Ruggedness Regression in R

Because of the 3-way interaction, get the Final Regression by fitting to each Degas level

```
m.DegasN = lm(LD ~ I(Power-80)*I(Time - 30),data=d[d$Degas=="N",])
m.DegasY = lm(LD ~ I(Power-80)*I(Time - 30),data=d[d$Degas=="Y",])
summary(m.DegasY)
```

```
                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 8.326e+00  3.667e-02 227.023   <2e-16 ***
## I(Power - 80)              -2.354e-04  4.735e-04  -0.497   0.6267
## I(Time - 30)               -1.536e-02  9.148e-03  -1.679   0.1152
## I(Power - 80):I(Time - 30)  1.995e-04  9.643e-05   2.069   0.0575 .
```

- **All terms are small** and not of practical importance inside the range of values tested
- **None of the terms are statistically significantly** larger than zero
- The coefficient estimaes give the regression equations provided earlier.

# 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

# Reproducibility

The **response** of interest (many times a **log reduction** measure of antimicrobial efficacy) should be repeatable within a lab and reproducible across multiple labs, quantified by a small

**reproducibility SD** *mixed effects* ANOVA

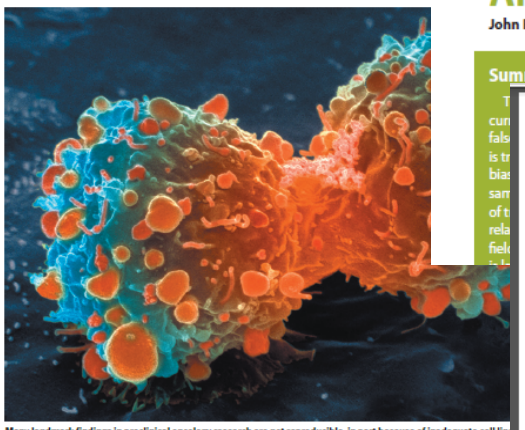
This assessment requires data from a multi-laboratory study

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

# Reproducibility

COMME

AVIAN INFLUENZA Shift expertise to track mutations where they emerge p.534 | EARTH SYSTEMS Past climates give valuable clues to future warming p.537 | HISTORY OF SCIENCE Des lost letter tracked u Google p.540



Many landmark findings in preclinical oncology research are not reproducible, in part because of inadequate cell lin

Raise standards for preclinical cancer research

C. Glenn Begley and Lee M. Ellis propose how methods, publications and incentives must change if patients are to benefit.

Efforts over the past decade to characterize the genetic alterations in human cancers have led to a better understanding of molecular drivers of this complex set of diseases. Although we in the cancer field hoped that this would lead to more effective drugs, historically, our ability to translate cancer research to clinical success has been remarkably low[1]. Sadly, clinical trials in oncology have the highest failure rate compared with other therapeutic areas. Given the high unmet need in oncology, it is understandable that barriers to clinical development may be lower than for other disease areas, and a larger number of drugs with suboptimal preclinical validation will enter oncology trials. However, this low success rate is not sustainable or acceptable, and investigators must reassess their approach to translating discovery research into greater clinical success and impact.

Many factors are responsible for the high failure rate, notwithstanding the inherently difficult nature of this disease. Certainly, the limitations of preclinical tools such as inadequate cancer-cell-line and mouse models[2] make it difficult for even ▶

29 MARCH 2012 | VOL 483 | NATURE | 531



Open access, freely available online

Essay

Why Most Published Research Findings Are False

John P. A. Ioannidis

Summary

factors that influence this problem and some corollaries thereof.

is characteristic of the field and can vary a lot depending on whether the

OPEN ACCESS Freely available online

PLOS | MEDICINE

Essay

How to Make More Published Research True

John P. A. Ioannidis[1,2,3,4]

1 Meta-Research Innovation Center at Stanford (METRICS), Stanford University, Stanford, California, United States of America, 2 Department of Medicine, Stanford Prevention Research Center, Stanford, California, United States of America, 3 Department of Health Research and Policy, Stanford University School of Medicine, Stanford, California, Unite

COMMENTARY

generates irreproducible results

t, Sarah L Vowler & Gordon B Drummond

wever, a major cause of this lack of variability in the *P* value. We explain ng analyses based predominantly on

null-hypothesis testing, this article will not revisit these issues. Rather, we concentrate on how *P* values themselves are misunderstood.

Although statistical power is a central element in reliability[18], it is often considered only when a test fails to demonstrate a real effect (such as a difference between lse negative' result (see **Box 2** y of statistical terms used in Many scientists who are not lo not realize that the power of lly relevant when considering gnificant results, that is, when othesis appears to be untenecause the statistical power of atically affects our capacity to

THIS WEEK

EDITORIALS

CONSERVATION Saving species is far from a walk in the park p.8 | WORLD VIEW Psychology gears up to check its workings p.9 | BREAKFAST Chimps plan days to ensure they nab tastiest figs p.11

Journals unite for reproducibility

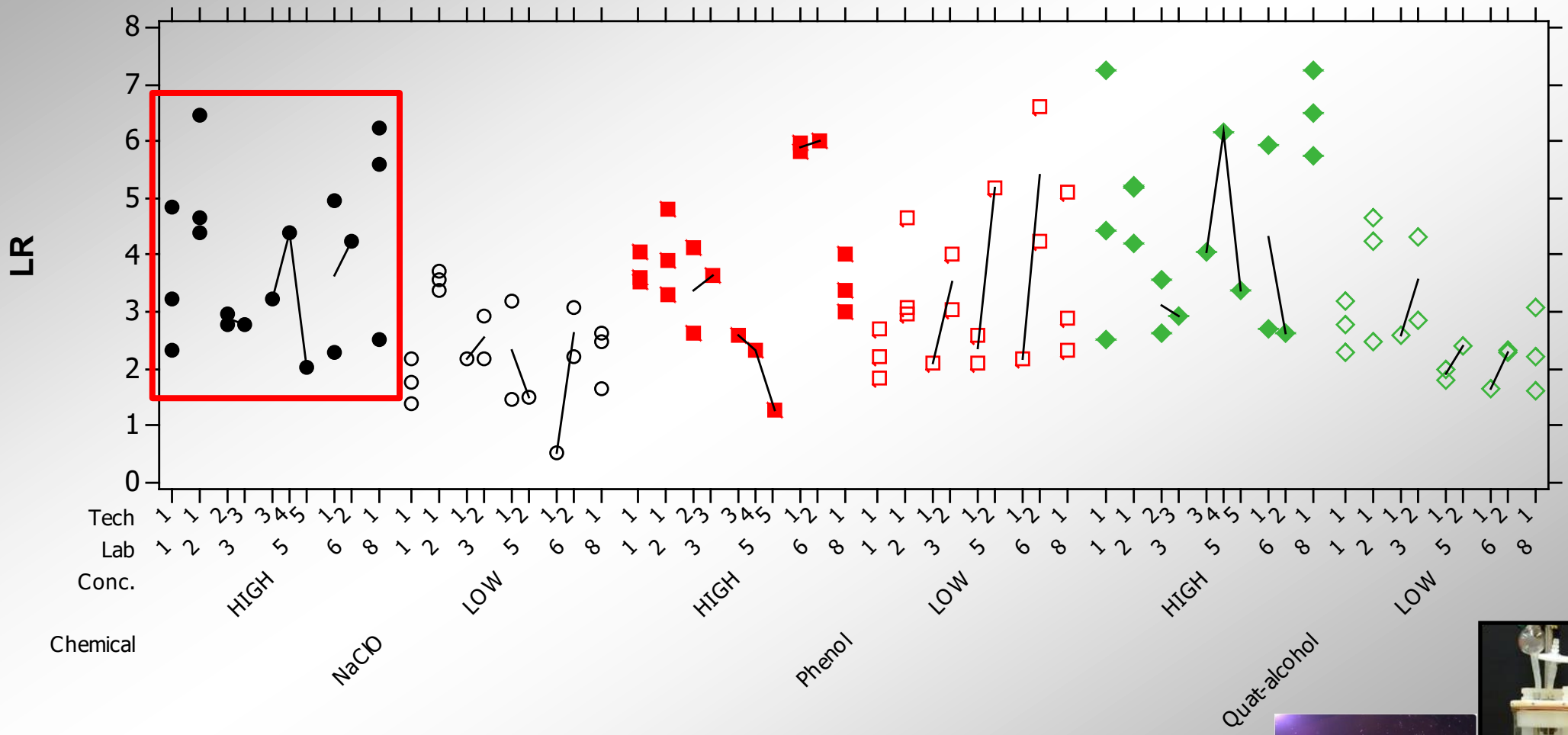
*Consensus on reporting principles aims to improve quality control in biomedical research and encourage public trust in science.*

Reproducibility, rigour, transparency and independent verification are cornerstones of the scientific method. Of course, just because a result is reproducible does not make it right, and just because it is not reproducible does not make it wrong. A transparent and rigorous approach, however, will almost always shine a light on issues of reproducibility. This light ensures that science moves forward, through independent verifications as well as the course corrections that come

the sample size was determined and what criteria were used to include or exclude any data. Journals should recommend deposition of data in public repositories, where available, and link data bidirectionally when the paper is published. Journals should strongly encourage, as appropriate, that all materials used in the experiment be shared with those who wish to replicate the experiment. Once a journal publishes a paper, it assumes the obligation to consider publication of a refutation of that

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

# STM Reproducibility (CBE/ASTM 2012)




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

# Reproducibility of STM applied to NaOCl in R

```
m = lmer(LR ~ (1|Lab),
      data=d[d$Chemical=="Chlorine"&d$Conc=="HIGH",])
summary(m)
## Random effects:
##  Groups   Name        Variance
##  Lab                  0.2671
##  Residual             1.7055
## Number of obs: 18, groups:  Lab, 6
## Fixed effects:
##             Estimate Std. Error
## (Intercept)   3.8906     0.3732

confint(m)
Computing profile confidence intervals
                2.5 %    97.5 %
(Intercept) 3.1002697 4.680862
```

$S_R = [0.27 + 1.71]^{1/2}$
$= 1.404$

**mean LR = 3.89**

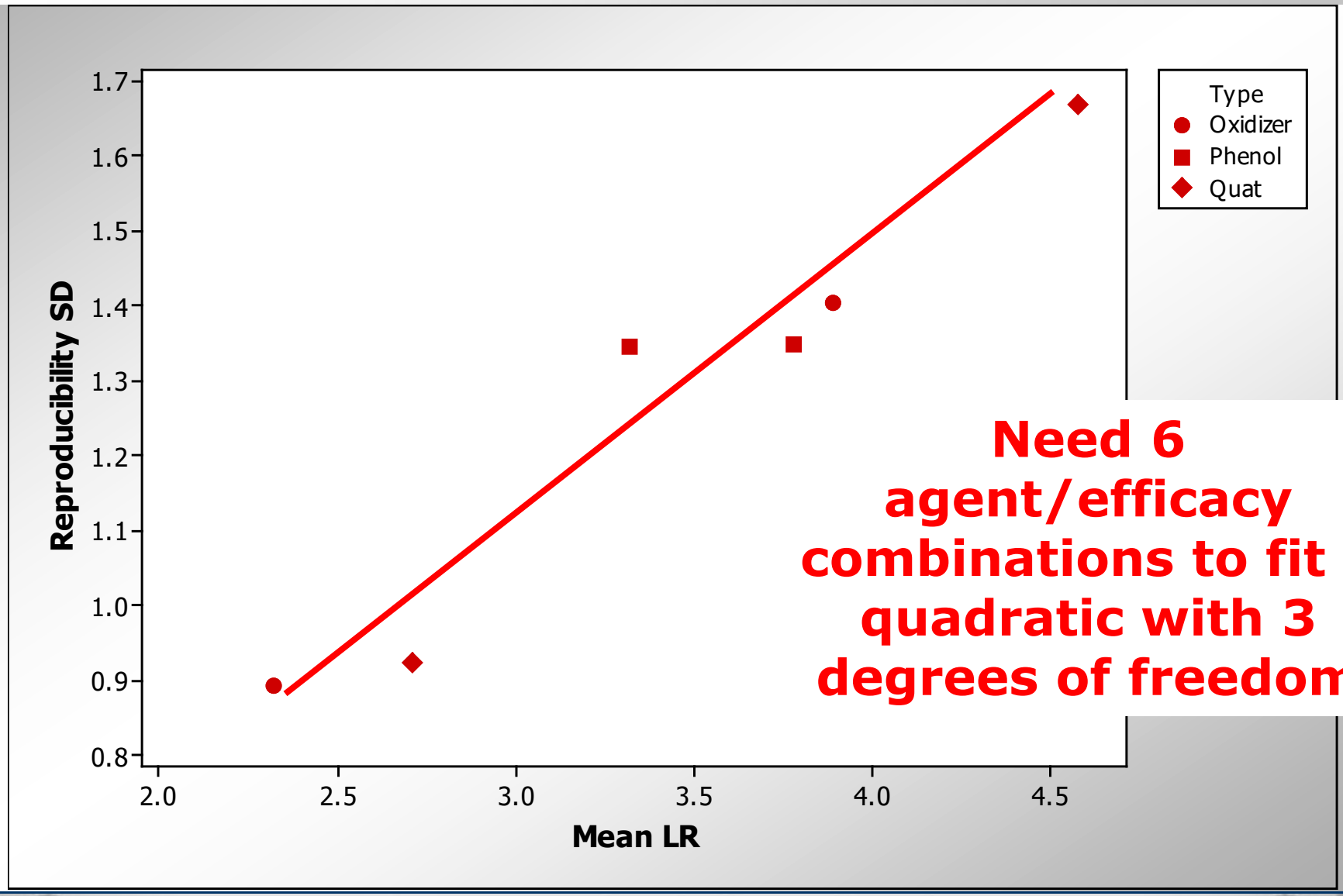
**95% Confident that the true mean LR for Chlorine when applied via the STM is larger than 3**

# Reproducibility among labs for the STM

## Quantification of the 6-lab collaborative study in 2012:

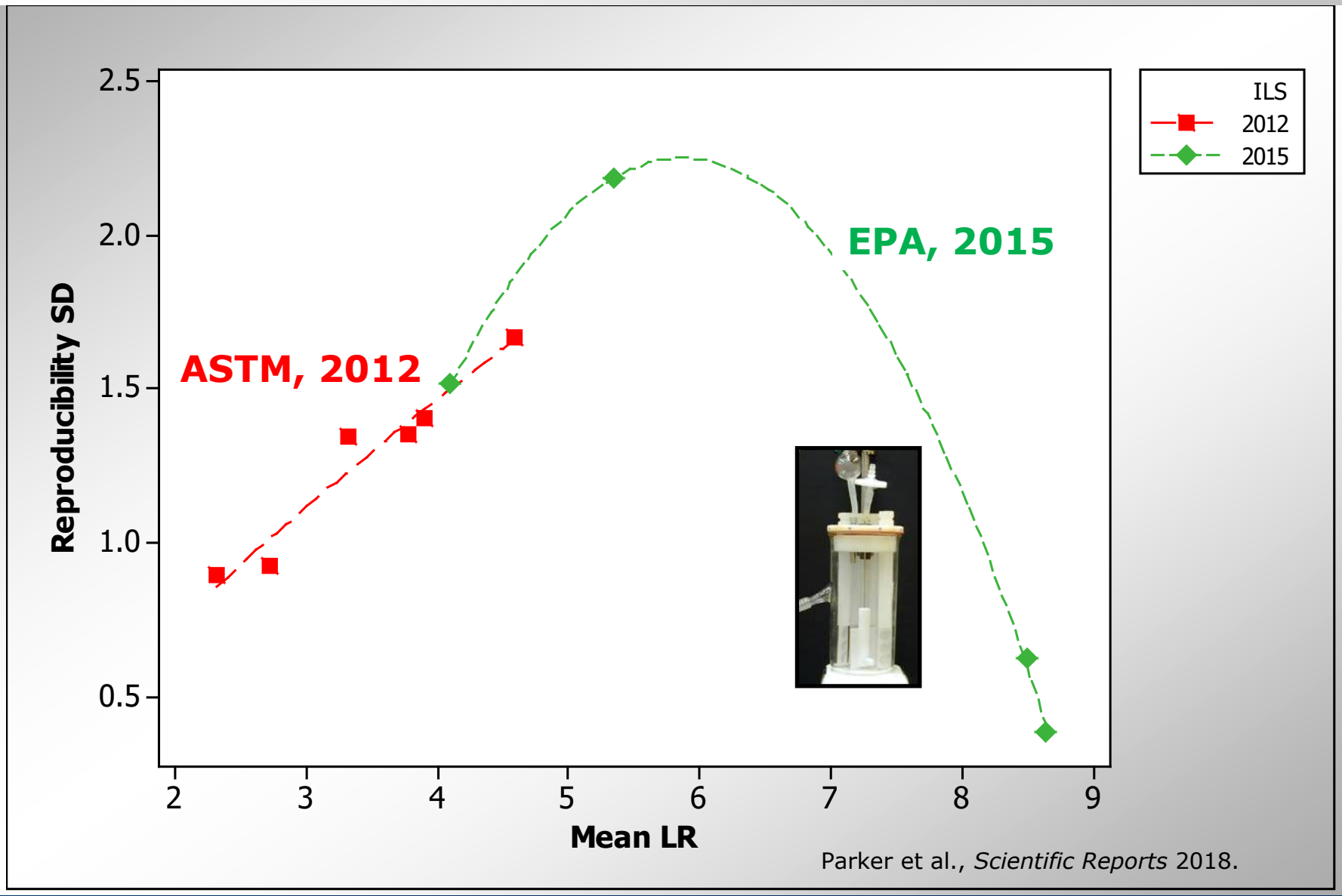
| Chemical | Conc. | Mean LR | Within lab % | Among lab % | Reproducibility SD |
|---|---|---|---|---|---|
| NaOCl | HIGH | 3.89 | 86% | 14% | 1.404 |
| | LOW | 2.32 | 71% | 29% | 0.891 |
| Phenol | HIGH | 3.78 | 18% | 82% | 1.349 |
| | LOW | 3.32 | 100% | 0% | 1.345 |
| Quat-alcohol | HIGH | 4.58 |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

# STM: Investigating how reproducibility changes vs. LRs …



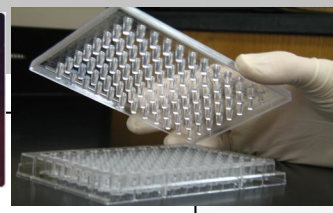

**Need 6 agent/efficacy combinations to fit a quadratic with 3 degrees of freedom**

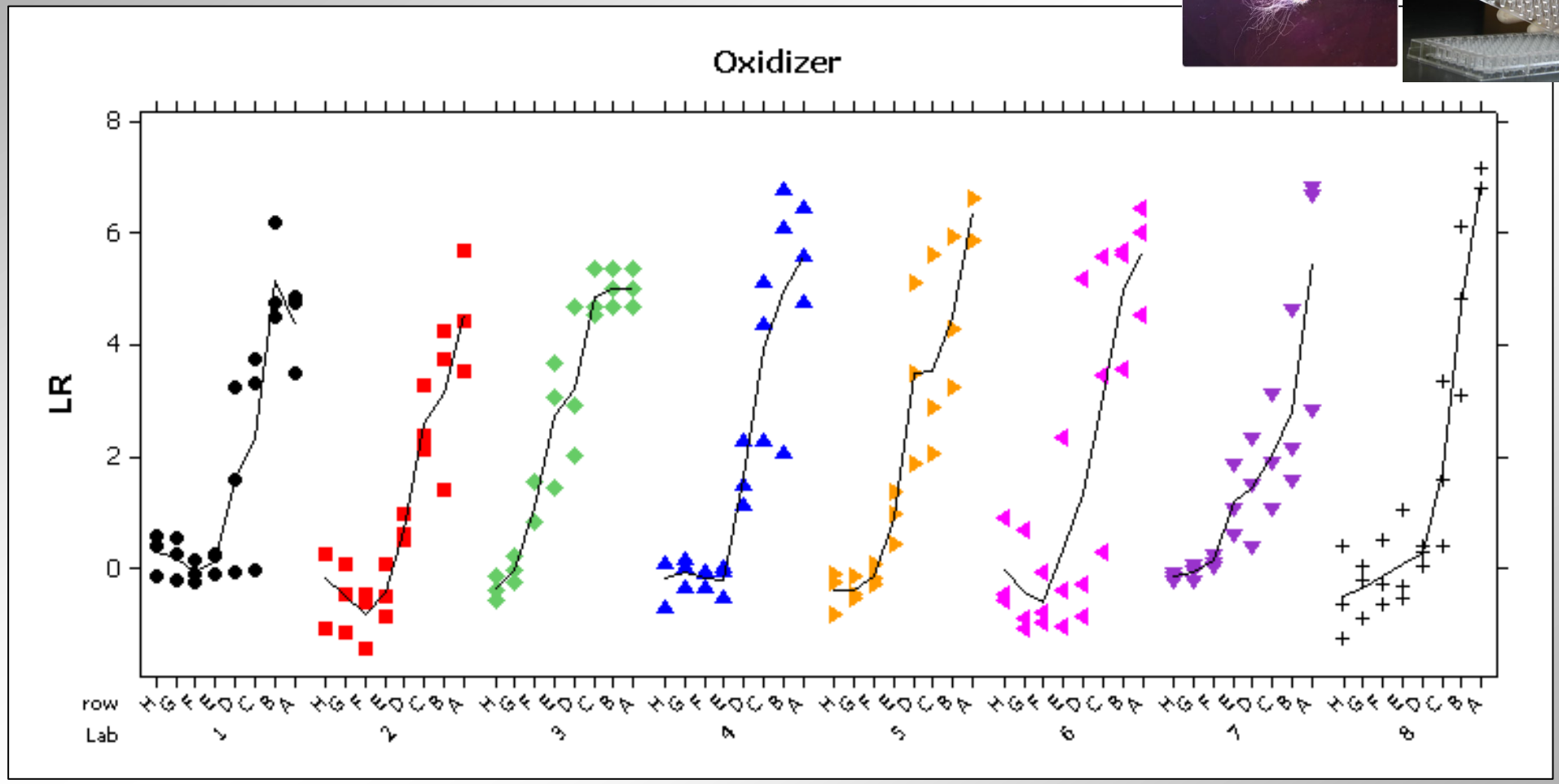
# STM Reproducibility in a graph …




Parker et al., *Scientific Reports* 2018.

# MBEC Reproducibility of the LRs





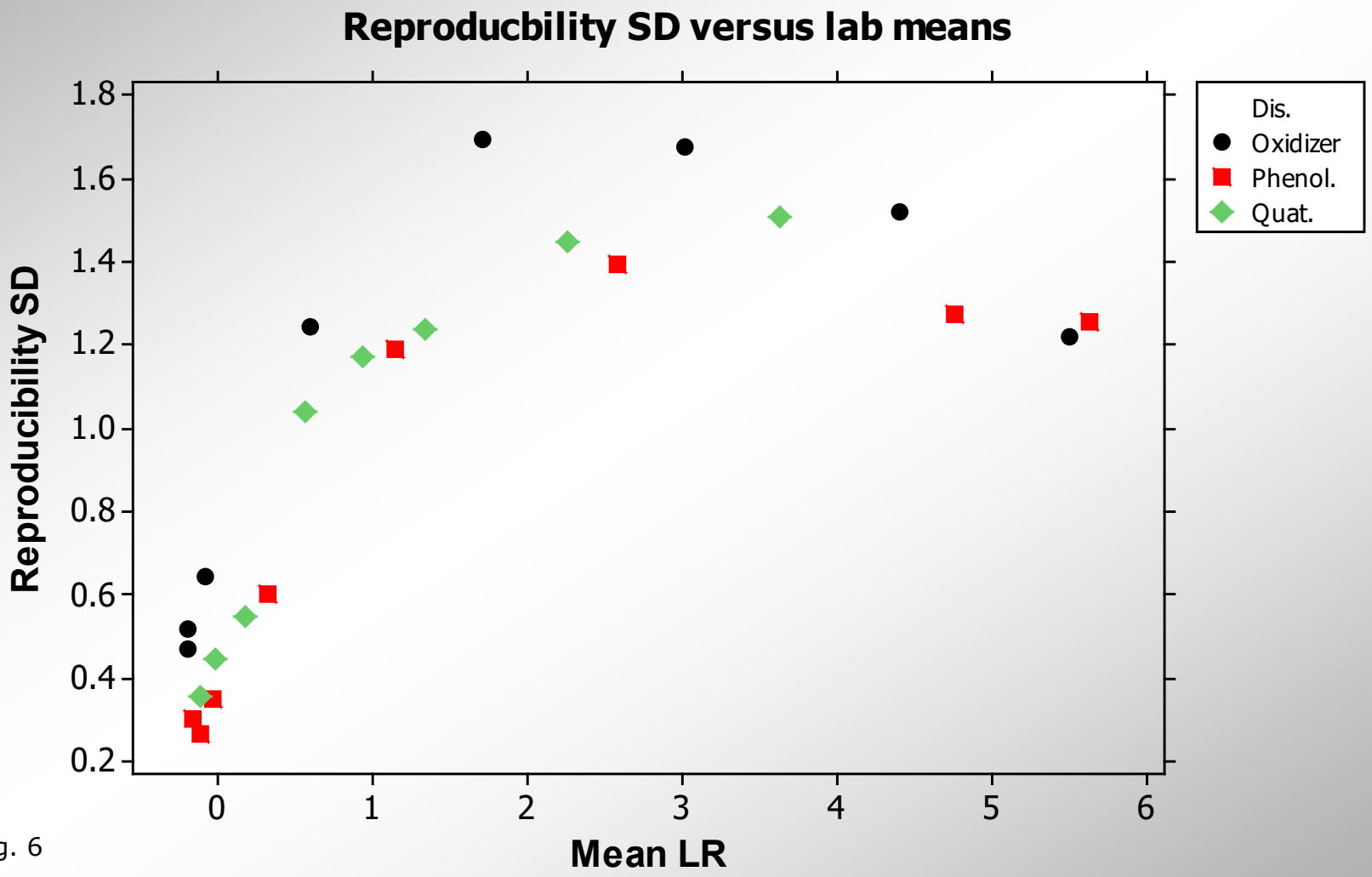

Parker et al. *JMM*, 2014; ASTM RR:E35-1006, 2011

# MBEC Reproducibility of the LRs

| Treatment | Efficacy (x2) | Mean LR | Reproducibility SD |
|---|---|---|---|
| Oxidizer | 8 | 5.5 | 1.2205 |
| | 7 | 4.41 | 1.5196 |
| | 6 | 3.03 | 1.6771 |
| | 5 | 1.72 | 1.6986 |
| | 4 | 0.6 | 1.2488 |
| | 3 | -0.08 | 0.6453 |
| | 2 | -0.19 | 0.4687 |
| | 1 | -0.18 | 0.5223 |
| Phenol | 8 | 5.64 | 1.2578 |
| | 7 | 4.76 | 1.2747 |
| | 6 | 2.59 | 1.3979 |
| | 5 | 1.15 | 1.1905 |
| | 4 | 0.34 | 0.6059 |
| | 3 | -0.02 | 0.3509 |
| | 2 | -0.11 | 0.2683 |
| | 1 | -0.15 | 0.3009 |
| Quat | 8 | 3.64 | 1.512 |
| | 7 | 2.26 | 1.4522 |
| | 6 | 1.34 | 1.2406 |
| | 5 | 0.95 | 1.1715 |
| | 4 | 0.58 | 1.0394 |
| | 3 | 0.18 | 0.5501 |
| | 2 | -0.01 | 0.4472 |
| | 1 | -0.11 | 0.3598 |

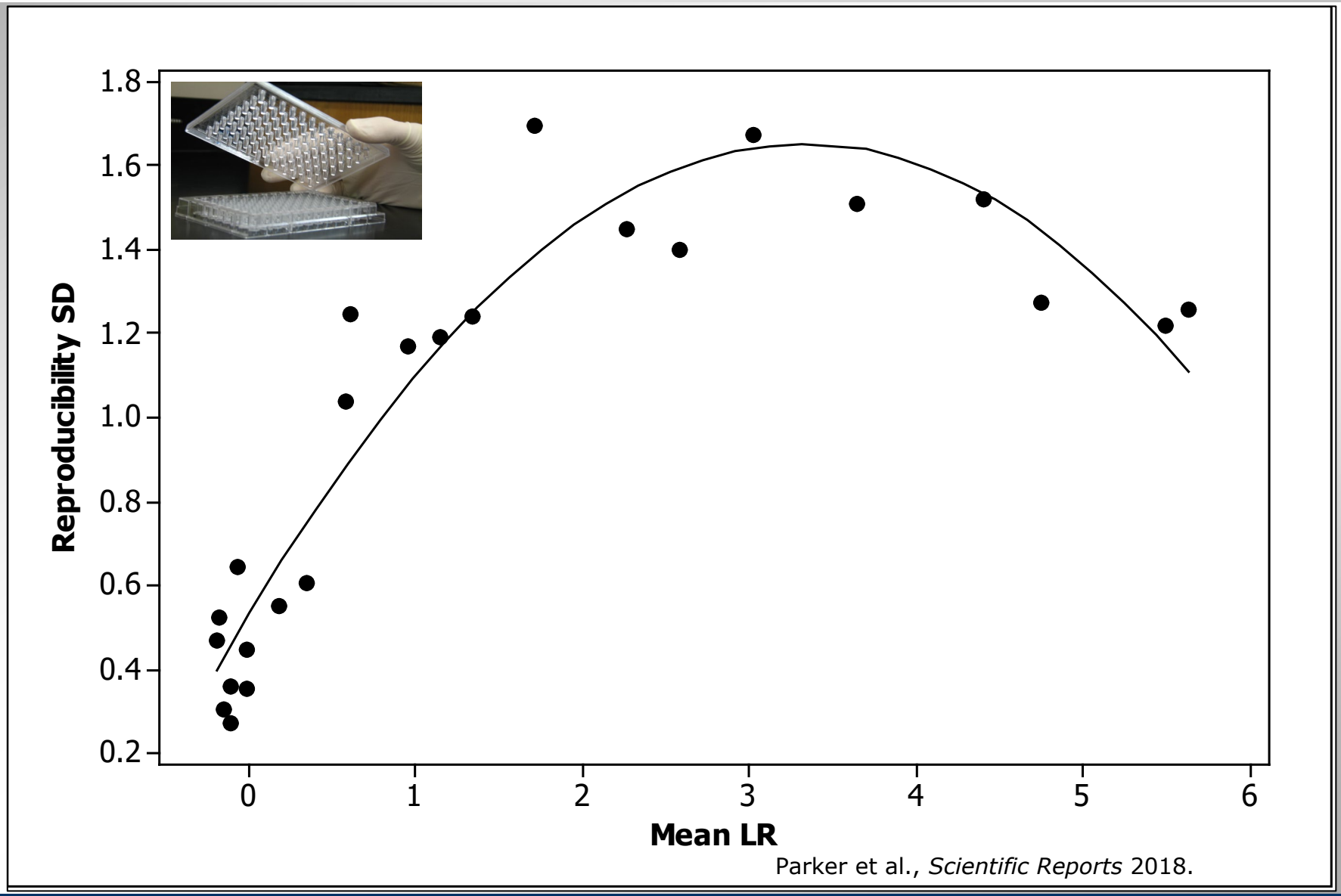
Parker et al. *JMM*, 2014; ASTM RR:E35-1006, 2011

# MBEC Reproducibility in a graph …



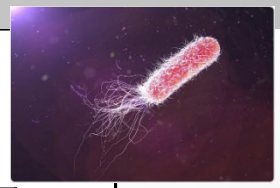

ASTM E2799, Fig. 6

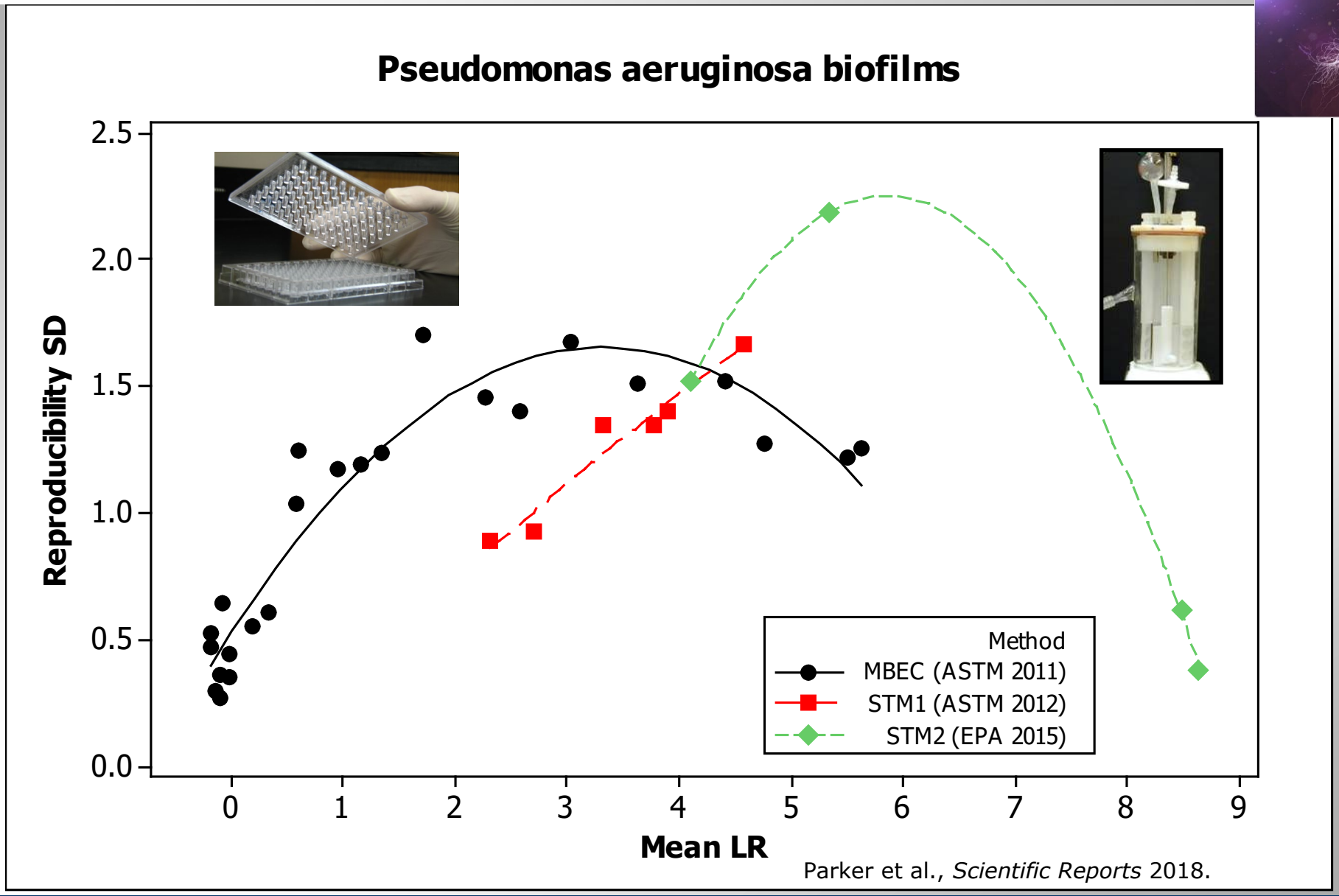
# MBEC: Investigating how reproducibility changes vs. LRs …




Parker et al., *Scientific Reports* 2018.

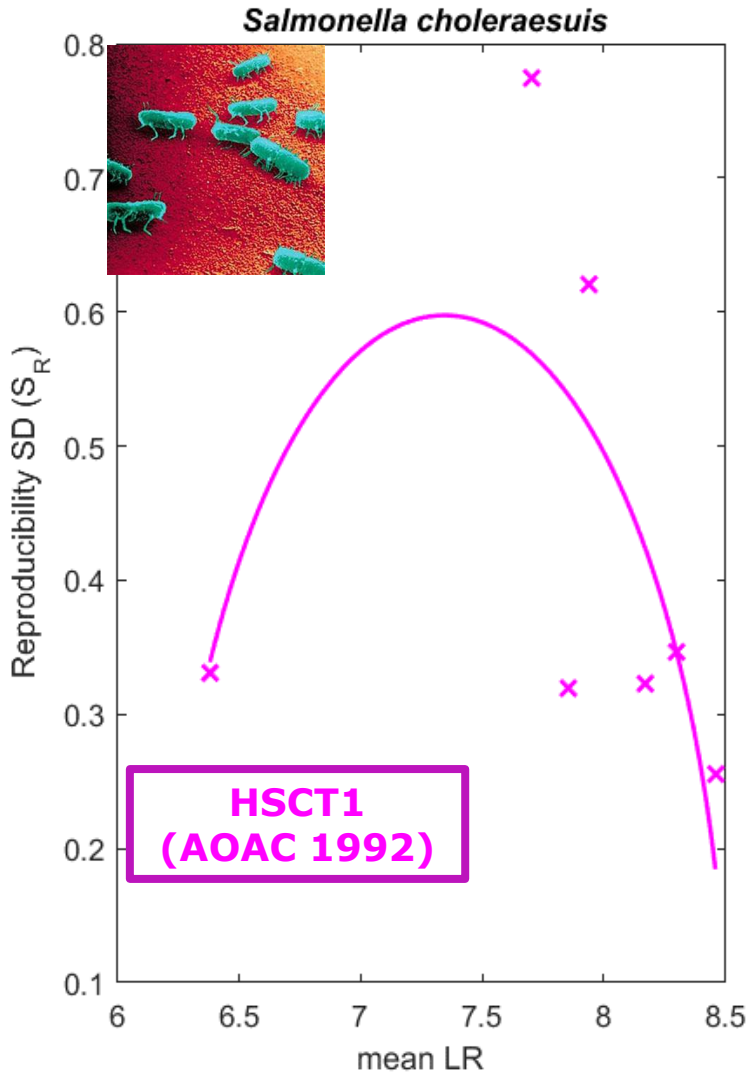
# ASTM biofilm methods: Reproducibility is a quadratic function of the LR

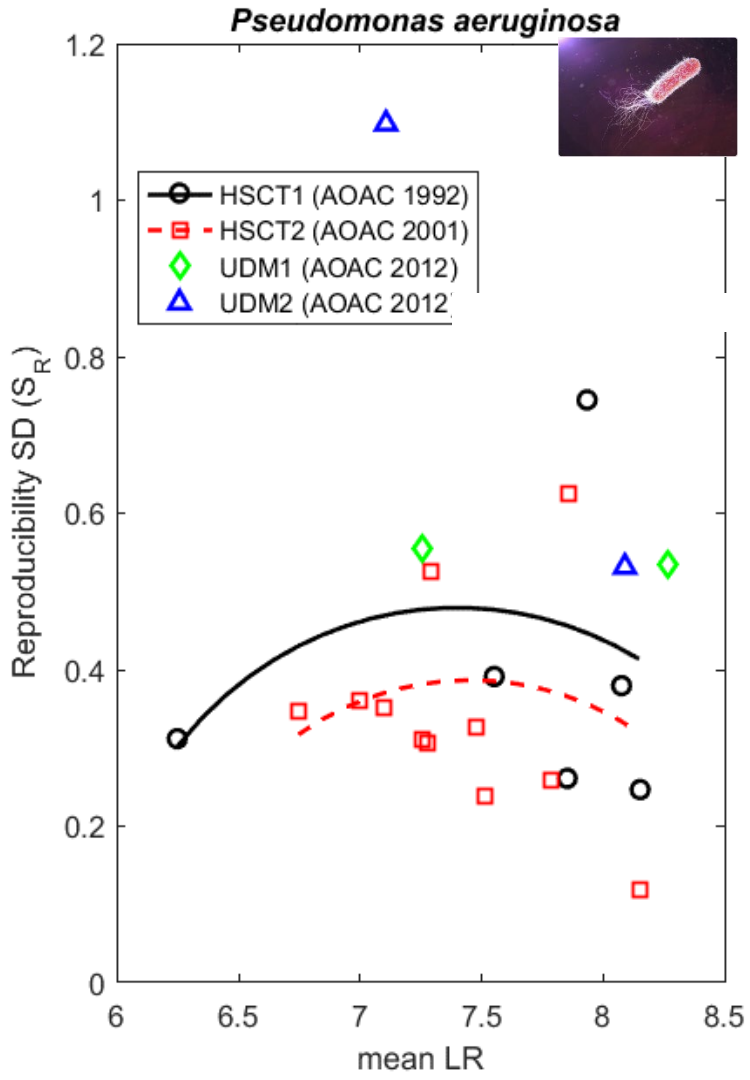





Parker et al., *Scientific Reports* 2018.

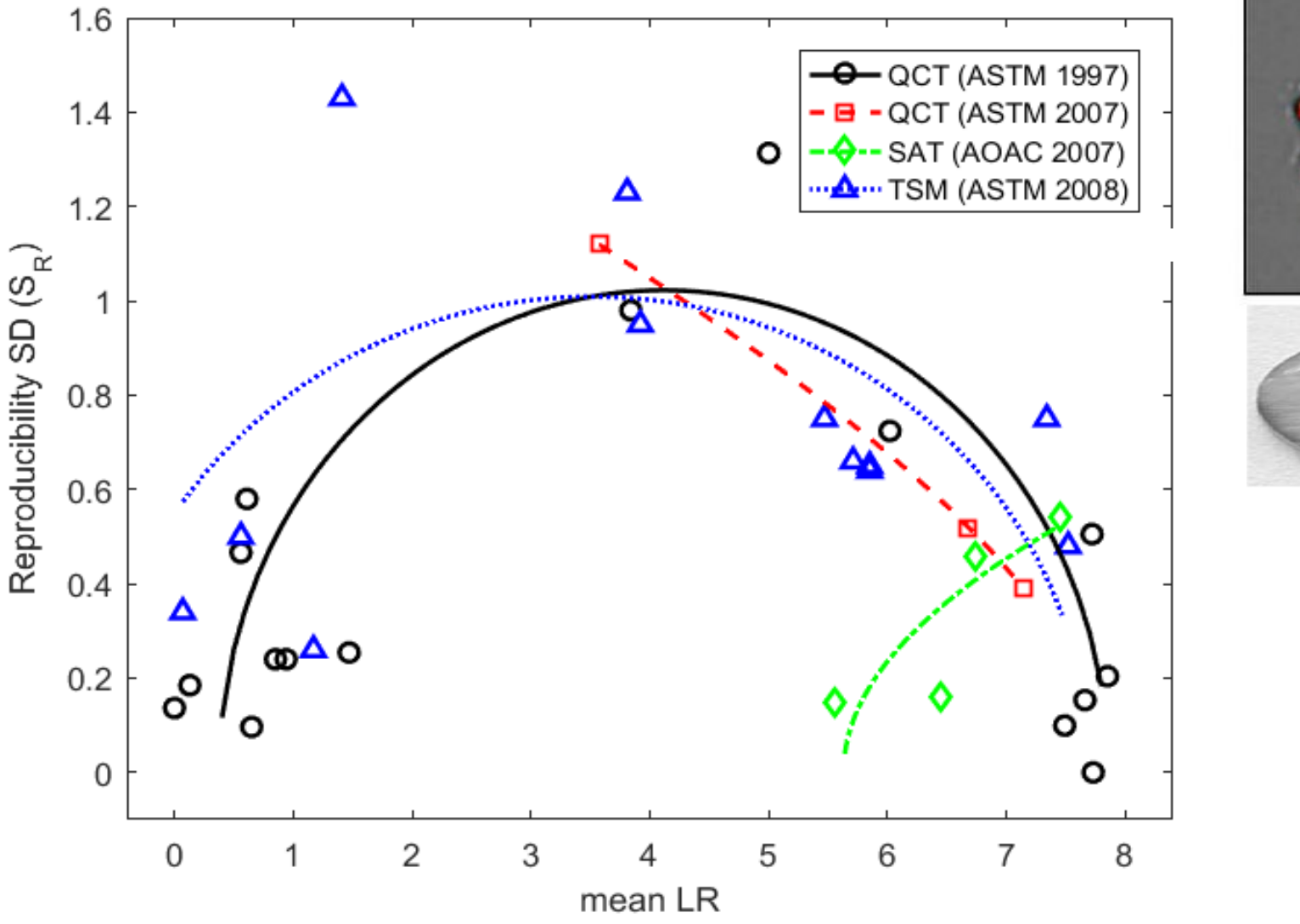
# Dried surface methods: Reproducibility is a quadratic function of the LR

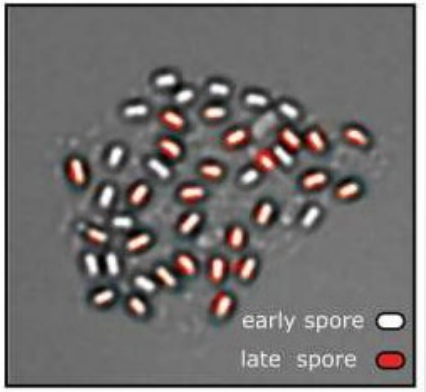



Parker et al., *Scientific Reports* 2018.

# Sporicide methods: Reproducibility is a quadratic function of the LR



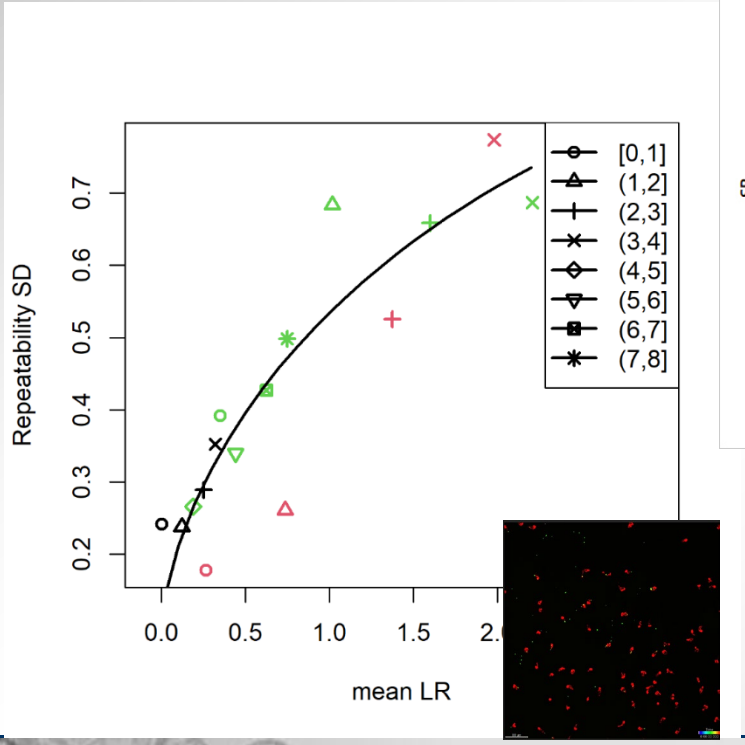




Parker et al., *Scientific Reports* 2018.

# Reproducibility is a quadratic function of the LR for any microbial response

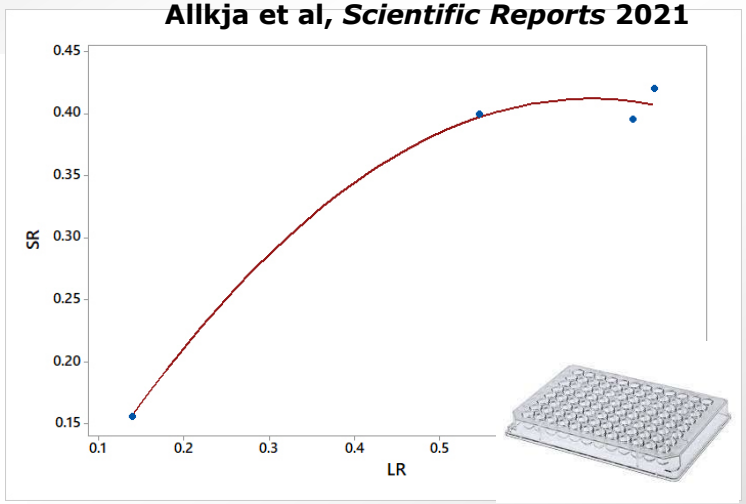
## Early *Staph* biofilms attacked by neutrophils, assessed by surface coverage

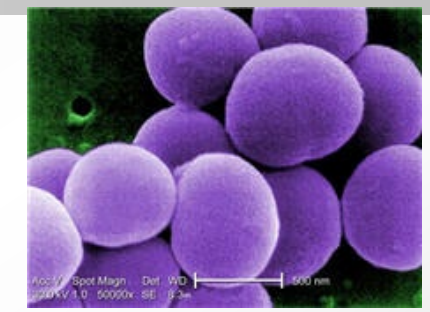
Pettygrove et al, *Frontiers Micro* 2022




## *Staph* biofilms in 96 well plate assessed by resazurin

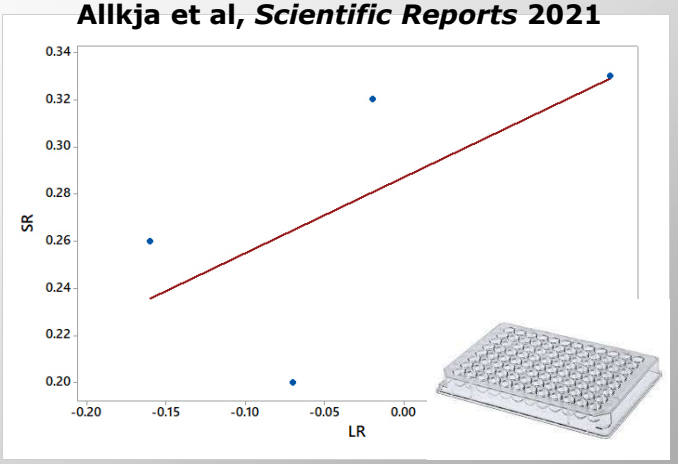
Allkja et al, *Scientific Reports* 2021



## *Staph* biofilms in 96 well plate assessed by crystal violet

Allkja et al, *Scientific Reports* 2021



# Any questions?