MONTANA STATE UNIVERSITY

Center for Biofilm Engineering

Biofilm Regulatory Toolbox

cost EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY

Funded by the European Union

Regulating Biofilm Claims with Statistical Confidence and Power

(Should I have my product tested in an external lab?)

Al Parker, Biostatistician
Center for Biofilm Engineering
Department of Mathematical Sciences
Montana State University, Bozeman

8 Sept, 2026

Goals of this presentation

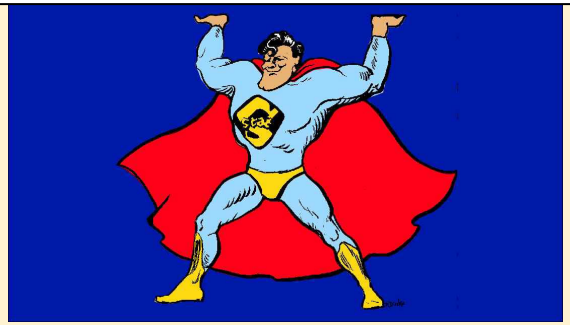
- The reproducibility (among-labs) of a regulatory method should be considered when setting a performance standard for determining product efficacy against biofilms
- Given:

  (1) a performance standard for biofilm claims
  (2) multi-lab reproducibility data
  (3) stakeholder specifications

  Estimate how often pass-errors and fail-errors occur (i.e., levels of confidence and power)

One R Module Based on This Talk

Performance Standard Assessment



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: 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

Standardized regulatory test methods against biofilms?

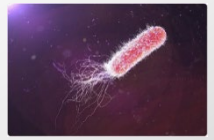
I

rdized
n slides
low
ear

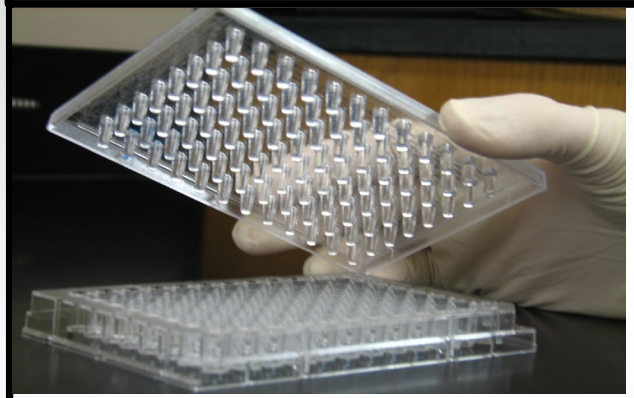
Used by researcher

CDC/STM

ASTM E2562/E2871
biofilm on discs
continuous flow
high shear

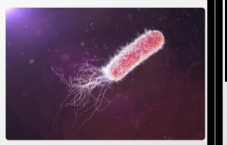


Used by US EPA

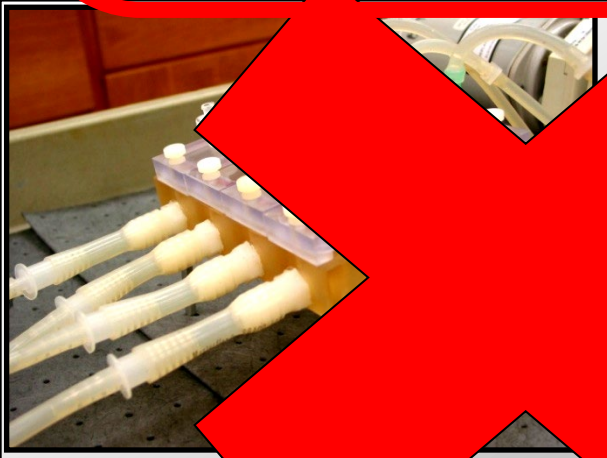


MBEC

ASTM E2799
biofilm on pegs
batch
low shear



Used by Health Canada



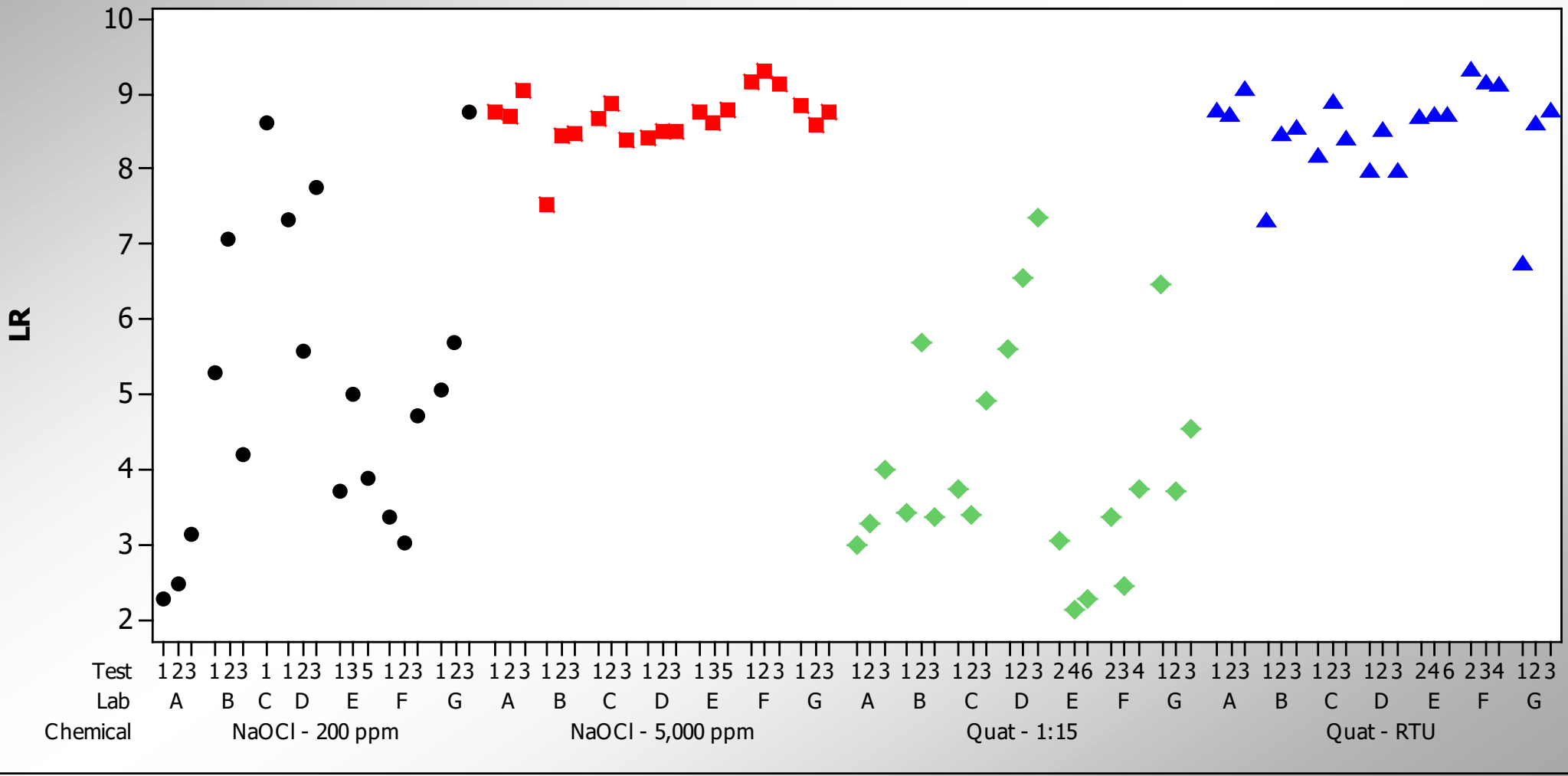
R

2647
m on slides
g flow
shear

Used by researcher

STM Reproducibility (EPA/ASTM 2015)

From a 7-lab study using *P. aeruginosa* (earlier talk showed 2012)



Reproducibility among labs for the STM

Results from a 7-lab collaborative study using *P. aeruginosa* in 2015:

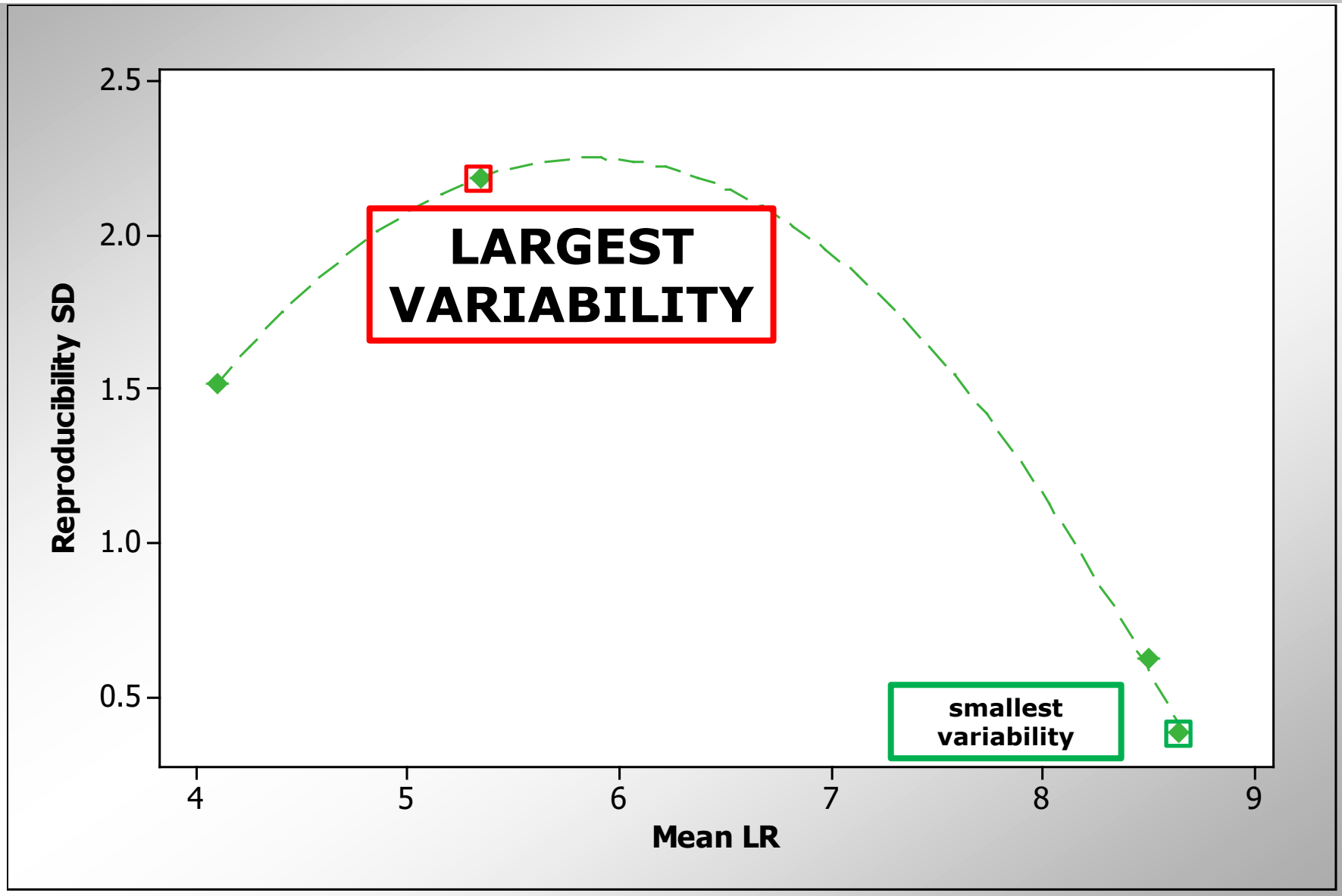
Treatment	Mean LR	S_r	S_R	Variance components		Percentage of total variance	
				Var_{Lab}	Var_{Test}	Among Lab	Within Lab
NaOCl 200ppm	5.34	1.236	2.182	3.23	1.53	68%	32%
NaOCl 5,000ppm	8.64	0.2465	0.3826	0.086	0.0607	59%	41%
Quat 1:15	4.10	0.4956	1.517	1.43	0.873	62%	38%
Quat RTU	8.50	0.5396	0.6197	0.093	0.291	24%	76%

LARGEST VARIABILITY

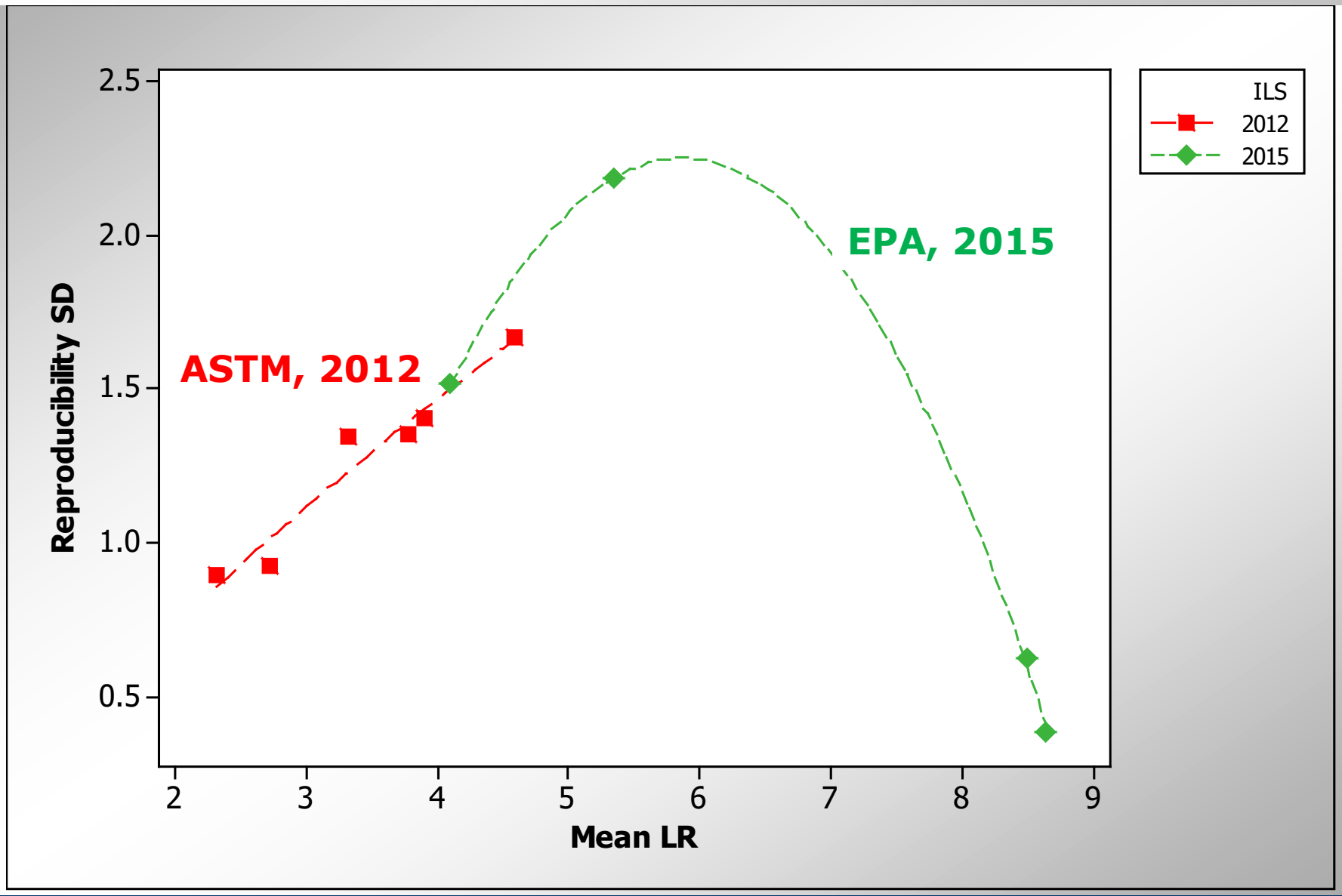
smallest variability

These are inputted into the performance standard tool

Investigating why reproducibility changes vs. LRs



Reproducibility is a quadratic function of the LR



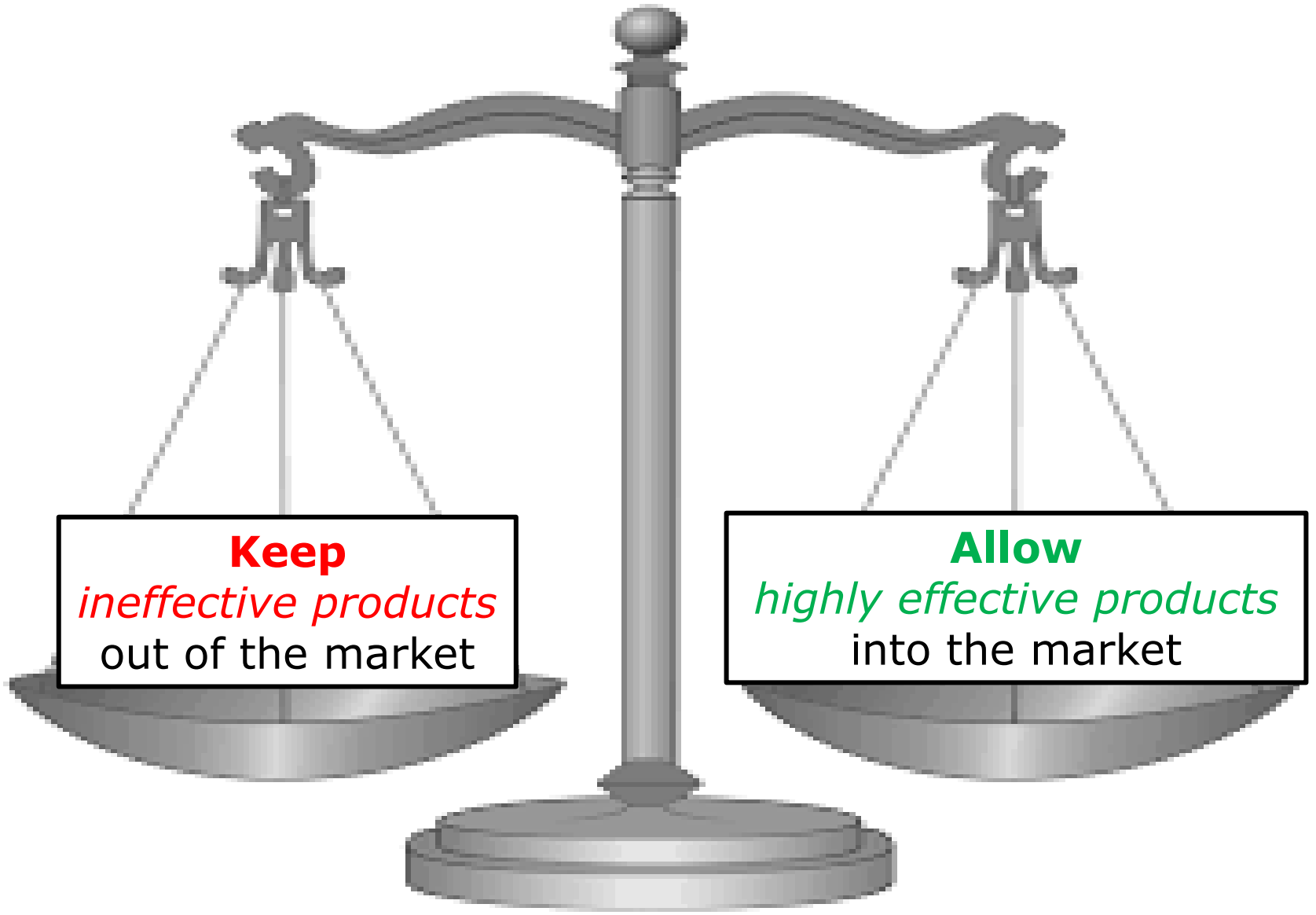
Performance standard for an antimicrobial method

A performance standard (PS) specifies how a product should be tested:

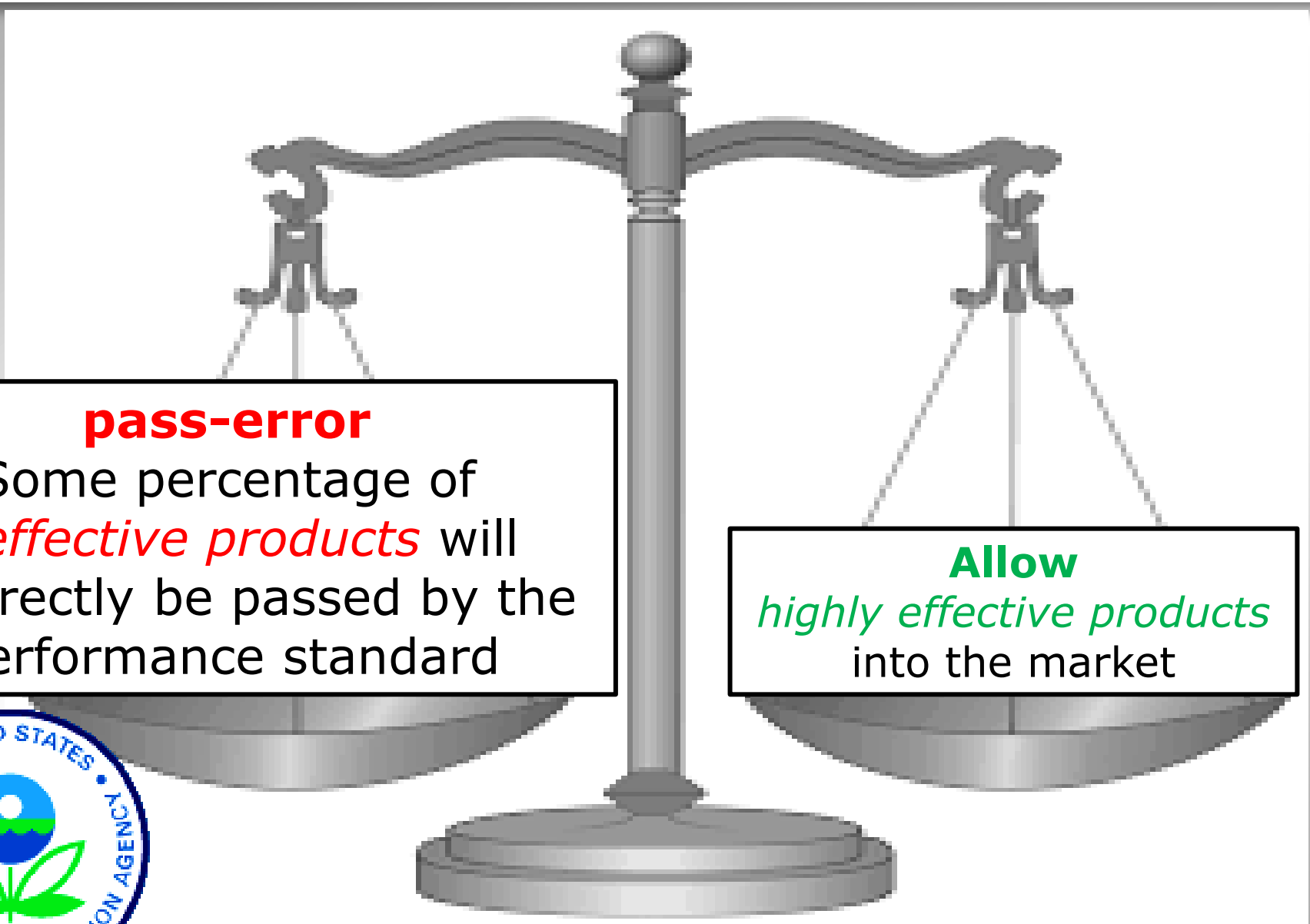
- the number of tests that must be performed
- the number of laboratories to conduct the tests (usually 1)
- the number of batches/lots of product to use
- the test microbe(s)
- an acceptable LR, $LR_{required}$. EPA typically specifies that the product attains $LR \geq LR_{required}$ to pass each test, and that the product passes all tests

If the product meets the specifications of a regulatory PS, then the product "**passes**" and can be sold in the marketplace. Otherwise, the product **fails**.

Regulating antimicrobial products



Regulating antimicrobial products



pass-error
Some percentage of *ineffective products* will incorrectly be passed by the performance standard

Allow
highly effective products
into the market

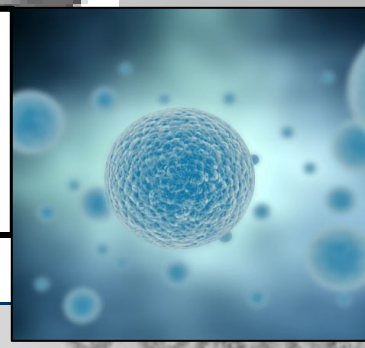
Regulating antimicrobial products

In scientific studies, investigators try to minimize both errors; they cannot be eliminated entirely due to a method's variability

pass-error
Some percentage of *ineffective products* will incorrectly be passed by the performance standard

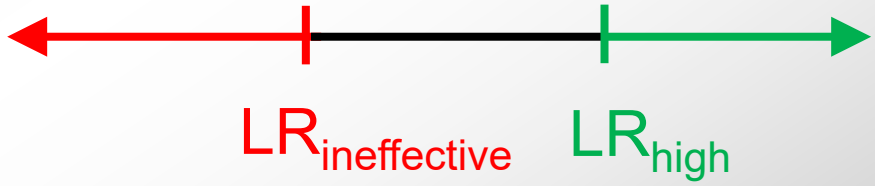
fail-error
Some percentage of *highly effective products* will incorrectly be failed by the performance standard

Manufacturers of highly effective antimicrobials



Pass-errors and statistical confidence!

in-effective products | **effective products** | **highly effective products**



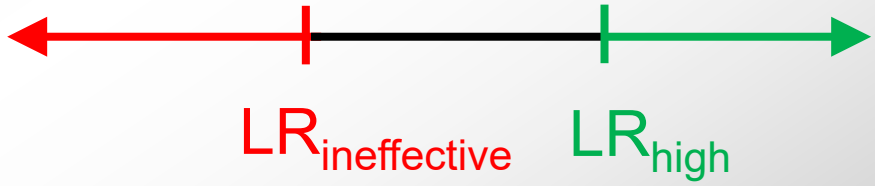
The following statements are equivalent

A performance standard …

… has a 5% pass-error percentage
… incorrectly passes 5% of ineffective products
… correctly fails 95% of ineffective products
… has 95% confidence that products that pass are truly effective

Fail-errors and statistical power!

in-effective products	effective products	highly effective products



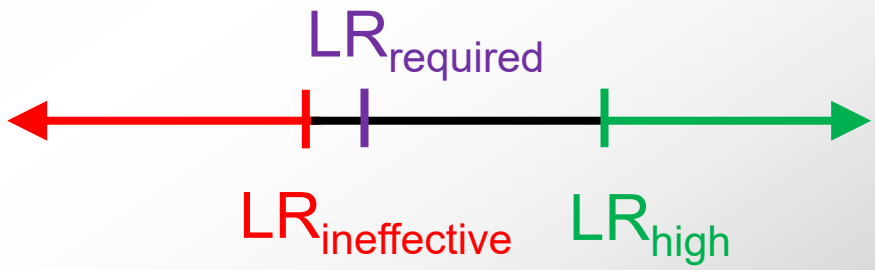
The following statements are equivalent

A performance standard …

… has a 5% fail-error percentage
… incorrectly fails 5% of highly effective products
… correctly passes 95% of highly effective products
… has 95% power to pass highly effective products

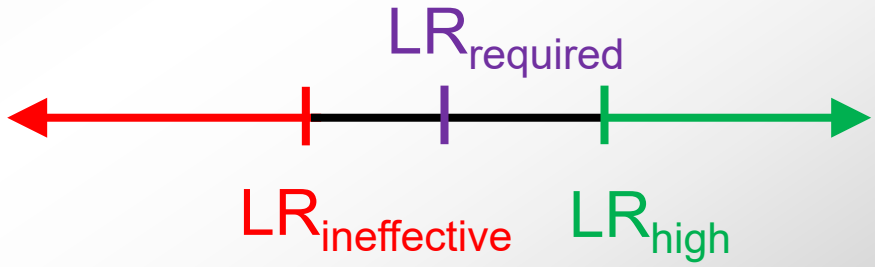
How to design a performance standard?

in-effective products | **effective products** | **highly effective products**



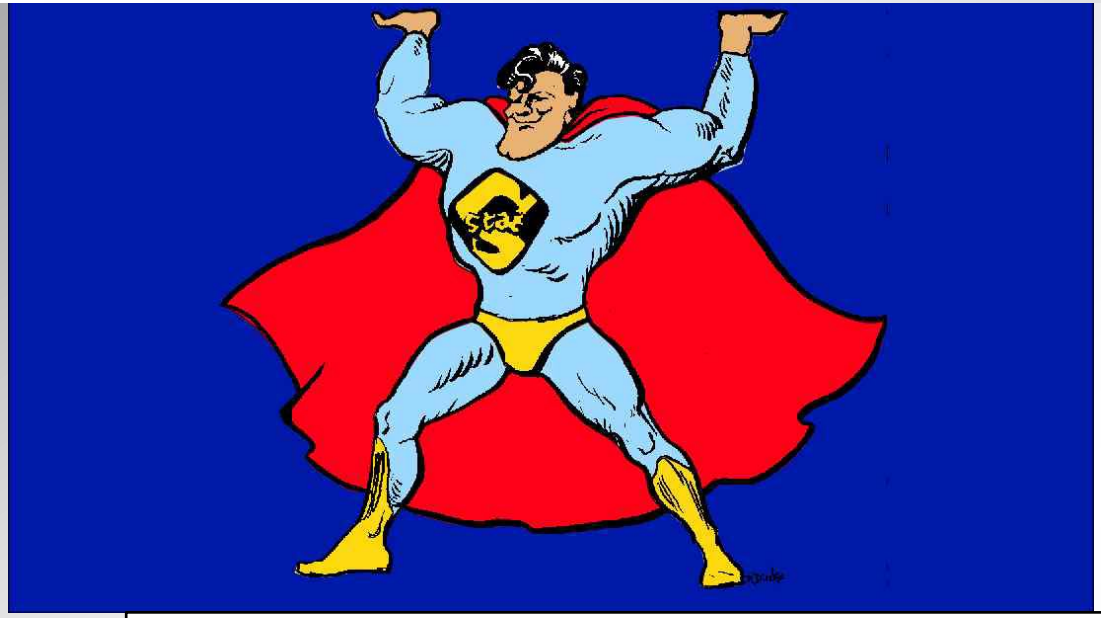
How to design a performance standard?

in-effective products | **effective products** | **highly effective products**

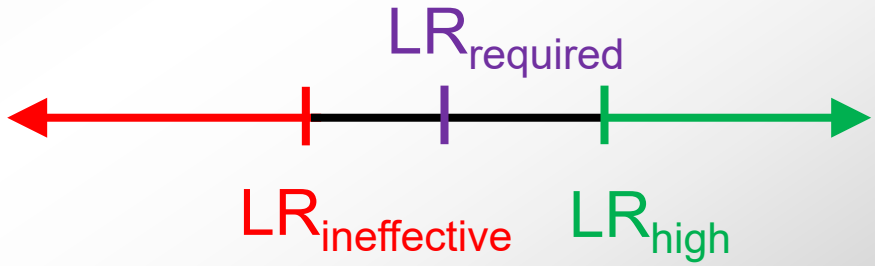


Where to set the required outcome ($LR_{required}$) per test?
How many carriers/coupons per test?
How many tests?
How many labs?

How to design a performance standard?



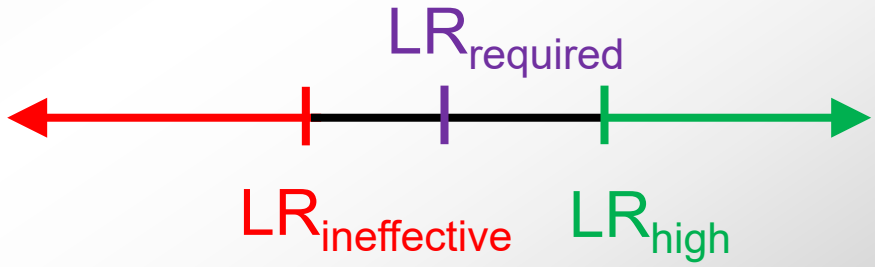
in-effective products **effective products** **highly effective products**



The *reproducibility* of the method determines how to set $LR_{required}$, the number of tests, and the number of labs that the *performance standard* must require in order to assess products' antimicrobial efficacy with low pass-error and low fail-error percentages (high confidence and high power)

How to design a performance standard?

in-effective products | **effective products** | **highly effective products**



Given a **performance standard** for a method there is a **statistical tool** to determine how often pass-errors and fail-errors occur.

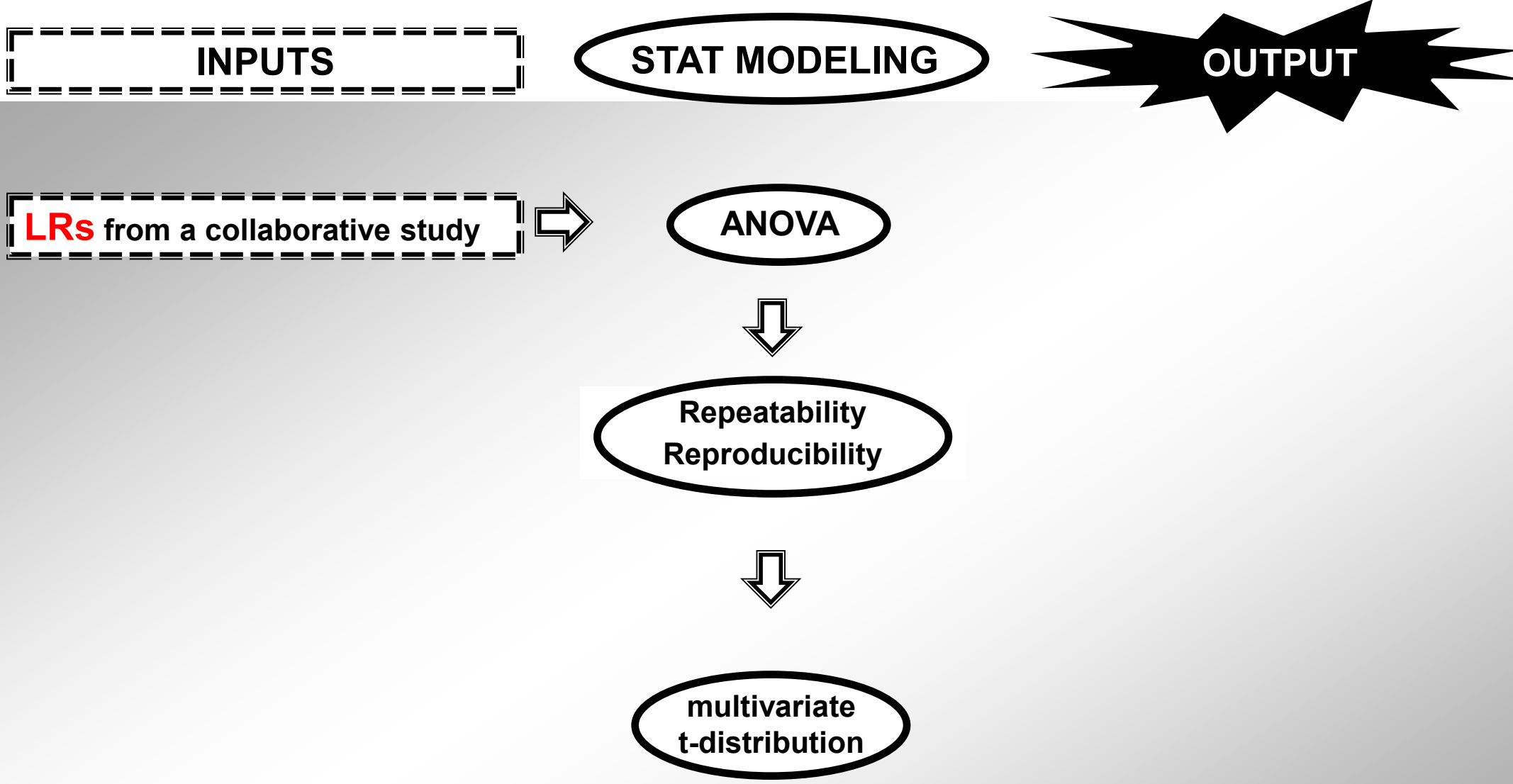
What is the Statistical Tool that Assesses a performance standard?

INPUTS

STAT MODELING

OUTPUT

LRs from a collaborative study


INPUTS
STAT MODELING
OUTPUT
LRs from a collaborative study
ANOVA
Repeatability
Reproducibility
multivariate
t-distribution

LRs from a collaborative study

ANOVA

Repeatability
Reproducibility

Stakeholder Specifications:

(a) **PASSING CRITERION**

all tests have LRs ≥ $LR_{required}$

(b) **INEFFECTIVE PRODUCTS**

have mean LR ≤ $LR_{ineffective}$

and SD = $SD_{ineffective}$

(c) **HIGHLY EFFECTIVE PRODUCTS**

have mean LR ≥ LR_{high}

and SD = SD_{high}

multivariate
t-distribution

INPUTS

STAT MODELING

OUTPUT

LRs from a collaborative study

ANOVA

Repeatability
Reproducibility

Stakeholder Specifications:

(a) **PASSING CRITERION**

all tests have LRs ≥ $LR_{required}$

(b) **INEFFECTIVE PRODUCTS**

have mean LR ≤ $LR_{ineffective}$

and SD = $SD_{ineffective}$

(c) **HIGHLY EFFECTIVE PRODUCTS**

have mean LR ≥ LR_{high}

and SD = SD_{high}

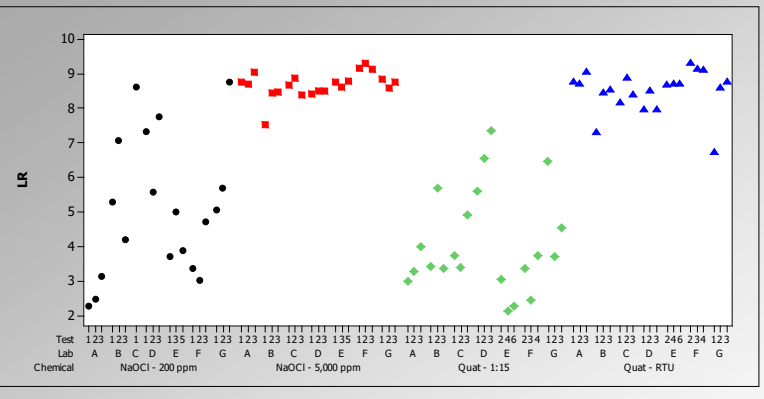
multivariate t-distribution

Goal:
low pass-error
and
low fail-error
percentages

INPUTS

STAT MODELING

OUTPUT



ANOVA

Repeatability
Reproducibility

multivariate
t-distribution

Goal:
low pass-error
and
low fail-error
percentages

Stakeholder Specifications:

(a) **PASSING CRITERION**

all tests have LRs ≥ $LR_{required}$

(b) **INEFFECTIVE PRODUCTS**

have mean LR ≤ $LR_{ineffective}$

and SD = $SD_{ineffective}$

(c) **HIGHLY EFFECTIVE PRODUCTS**

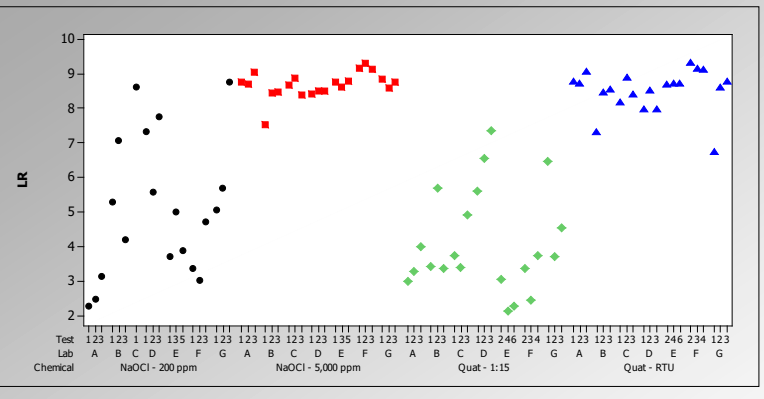
have mean LR ≥ LR_{high}

and SD = SD_{high}

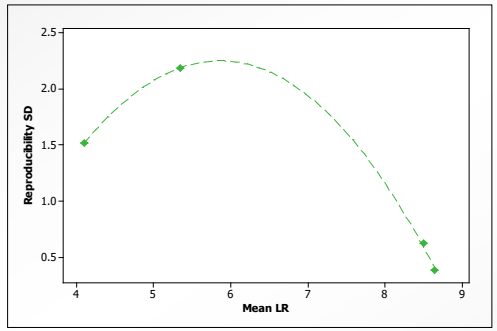
INPUTS

STAT MODELING

OUTPUT



ANOVA



Stakeholder Specifications:

(a) **PASSING CRITERION**

all tests have LRs ≥ $LR_{required}$

(b) **INEFFECTIVE PRODUCTS**

have mean LR ≤ $LR_{ineffective}$

and SD = $SD_{ineffective}$

(c) **HIGHLY EFFECTIVE PRODUCTS**

have mean LR ≥ LR_{high}

and SD = SD_{high}

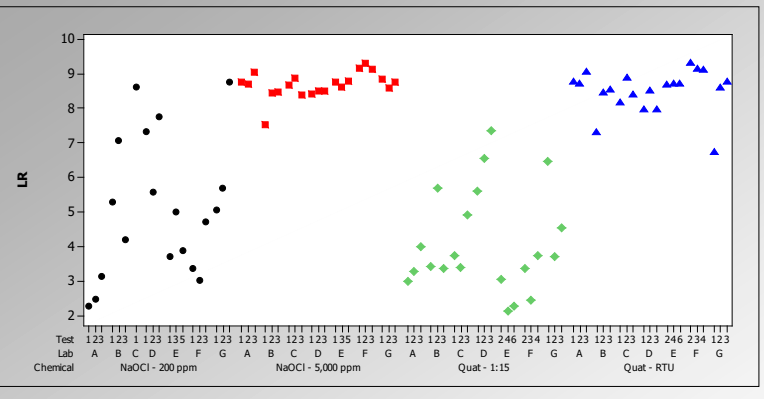
multivariate t-distribution

Goal:
low pass-error
and
low fail-error
percentages

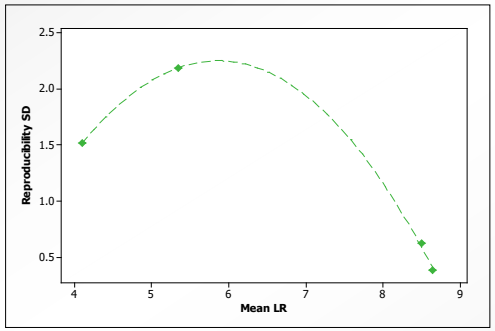
INPUTS

STAT MODELING

OUTPUT



ANOVA



Stakeholder Specifications:

(a) **PASSING CRITERION**

3 *P.a.* test LRs ≥ **6.0**

(b) **INEFFECTIVE PRODUCTS**

have mean LR ≤ **5**

SD = 2.07

(c) **HIGHLY EFFECTIVE PRODUCTS**

have mean LR ≥ **8.5**

SD = 0.784

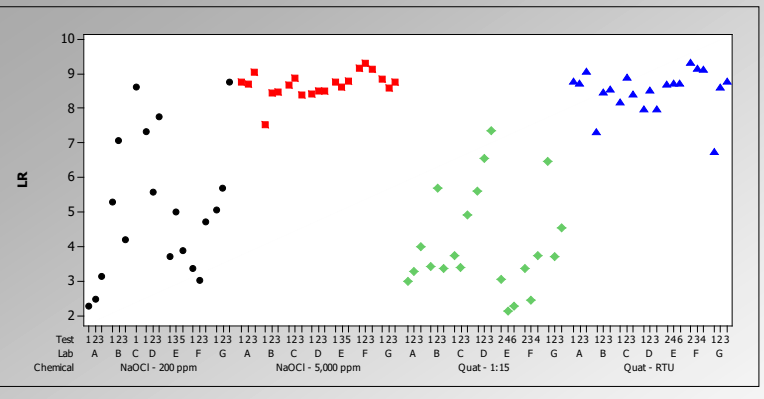
multivariate t-distribution

Goal:
low pass-error and low fail-error percentages

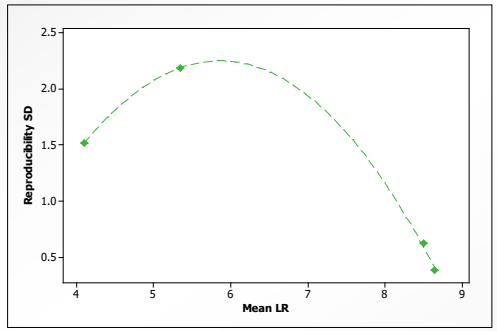
INPUTS

STAT MODELING

OUTPUT



ANOVA



Stakeholder Specifications:

(a) **PASSING CRITERION**

3 *P.a.* test LRs ≥ **6.0**

(b) **INEFFECTIVE PRODUCTS**

have mean LR ≤ **5**

SD = 2.07

(c) **HIGHLY EFFECTIVE PRODUCTS**

have mean LR ≥ **8.5**

SD = 0.784

multivariate t-distribution

Requiring 3 *P.a.* tests with LRs ≥ 6.0 achieves: 15% pass-errors and 0.3% fail-errors (for LR_{high} = 8.5)

Estimate Reproducibility off the parabola …

Stakeholder Specifications:

(a) **PASSING CRITERION**

3 *P.a.* test LRs ≥ **6.0**

(b) **INEFFECTIVE PRODUCTS**

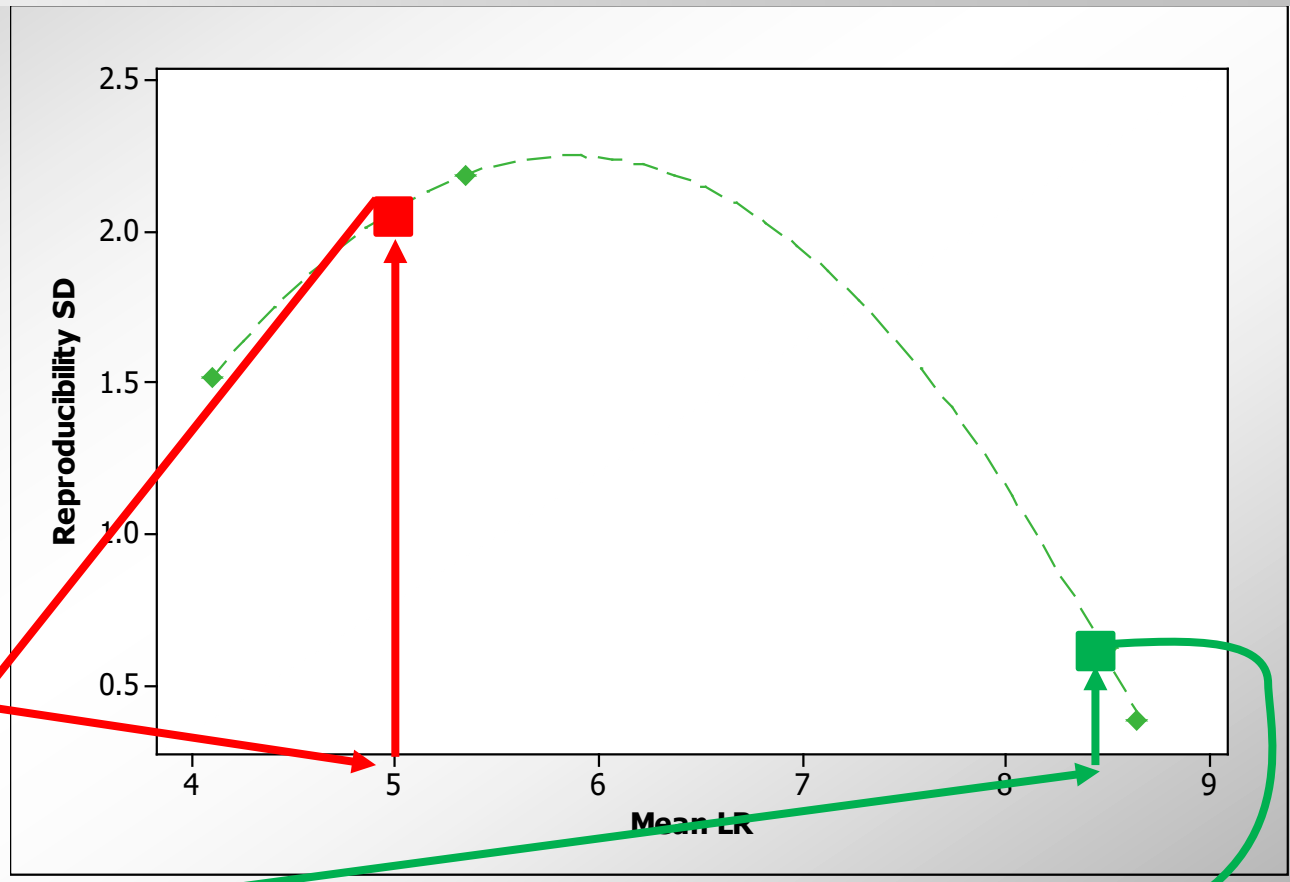
have mean LR ≤ **5**

SD = 2.07

(c) **HIGHLY EFFECTIVE PRODUCTS**

have mean LR ≥ **8.5**

SD = 0.784



Error Rates

1 and 2 bug performance STM standard scenarios

Microbes	Num Labs (*L*)	Num Tests per Lab and Microbe (*K*)	Total Num Tests	$LR_{required}$	Pass Error % (α) for products with mean LR = 5.0	Fail Error % (β) for products with mean LR = 8.5
P.a.	1	1	1	6.0	32.0%	0.1%
P.a.	1	3	3	6.0	15.2%	0.3%
P.a.	1	4	4	6.0	12.4%	0.4%
P.a.	2	1	2	6.0	10.4%	0.2%
P.a.	1	6	6	6.0	9.2%	0.5%
P.a.	1	15	15	6.0	4.8%	1.0%
P.a.	2	2	4	6.0	4.2%	0.4%
P.a. & *S.a.*	1	3	6	6.0	4.3%	0.5%
P.a. & *S.a.*	2	1	4	6.0	2.0%	0.4%

A single STM test vs *P.a.* incorrectly passes 32% of ineffective products

Even 3 STM tests vs. *P.a.* fails to maintain <10% pass-errors

EPA requires 3 STM tests vs *P.a.* and 3 tests vs. *S.a.* to maintain <5% pass-errors

Summary

- Regulators should set performance standards for antimicrobial test methods used for product registration so that:
  - a low percentage of ineffective products incorrectly pass
  - a low percentage of highly effective products incorrectly fail
- Given stakeholder specifications, a regulator can assess a method's performance standard with respect to:
  - pass-error percentages (statistical confidence)
  - fail-error percentages (statistical power)
  - the amount of effort (number of tests and labs) required to attain acceptably low error percentages
- Performance standards for more *reproducible* antimicrobial methods require less effort
- Do not send your product out to an external lab until you estimate the fail-error percentage for your product when tested by the regulatory performance standard

For additional statistical resources to assess antimicrobial methods, check out:
https://biofilm.montana.edu/documents-reports/knowledge-sharing-articles.html

Any questions?