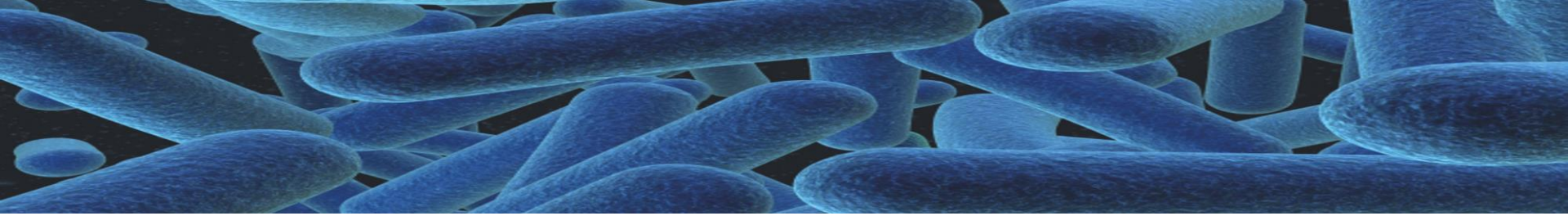
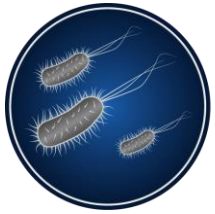


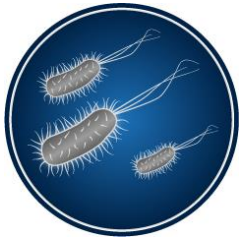
# Evaluation, Validation and Implementation of Alternative and Rapid Microbiological Methods

## Validation and Statistical Testing Strategies

Michael J. Miller, Ph.D.

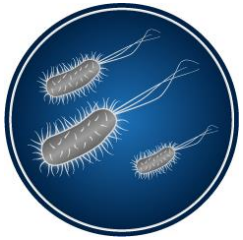


MICROBIOLOGY CONSULTANTS, LLC



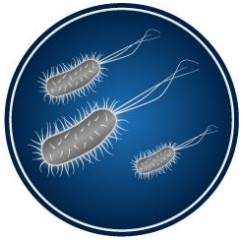
# Validation

- Validation is the act of demonstrating and documenting that a procedure operates effectively. Process validation is the means of ensuring and providing documentary evidence that processes (within their specified design parameters) are capable of consistently producing a finished product of the required quality.
  - *EMA, Note for Guidance on Process Validation. 2001*



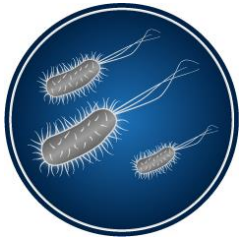
# Validation

- Should be more than a study conducted on a new rapid or alternative microbiology method or sample
- Should encompass the *entire process* that starts with the decision to change some aspect of the microbiological testing program and continues through ongoing routine use of the method
- A validated RMM *system* may actually consist of equipment or instrumentation, associated software and an analytical test method



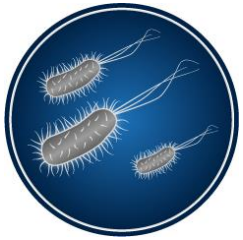
# Validation

- Provides the roadmap for all of the activities that will be required to demonstrate that the system is validated and suitable for its intended use
- Should include project deliverables, responsible parties for each phase of test execution, review and approval, and the documentation required to satisfy the expectations of the validation strategy
- Identify those features that will not be qualified/validated



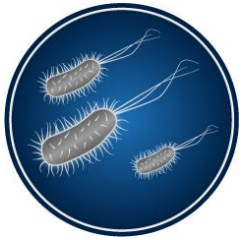
# Validation

- At the end of each phase, a summary of the results, whether the acceptance criteria have been met, and any deviations from the test plan should be documented and approved prior to initiating the next phase, unless it is acceptable to run phases in parallel



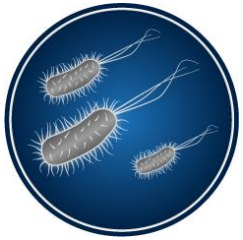
# Validation

- Pre-validation activities (e.g., proof of concept studies, supplier audit, business justification)
- Qualification of the equipment/instrumentation and the associated software/computer system
  - Can be part of the DQ, IQ, OQ and/or PQ
  - Refer to USP <1058>, *Analytical Instrument Qualification*
- Validation for the intended use
  - Meeting validation criteria using standardized microorganisms
  - Method suitability
  - Equivalency studies



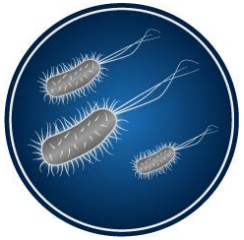
# Proof-of-Concept (POC) Testing

- Prior to purchasing a RMM system, POC or feasibility testing can be performed to determine if incompatibilities exist between the RMM and the intended product or test sample(s)
- These types of studies can also be performed in the event the RMM supplier has little or no data on testing similar product or test materials
- This can be accomplished using a rental or loaner instrument, or by sending samples directly to the supplier for evaluation
- Also provides a top level opportunity to “play” with the system



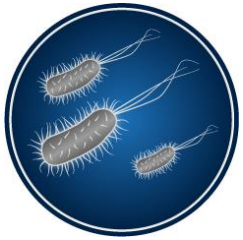
# Assessment of Supplier Capabilities

- Understand the supplier's technical capabilities and their ability to support each phase of the validation process as well as continuing assistance once the RMM is placed into service
  - Does the supplier have a robust quality, change control and manufacturing system in place?
  - Do they have appropriate documentation with regard to the design and manufacture of their instrumentation?
  - Has the supplier been audited by other companies or regulatory agencies?
  - Are they financially secure?
  - Are they the sole provider of the RMM consumables, reagents, supplies or replacement parts?



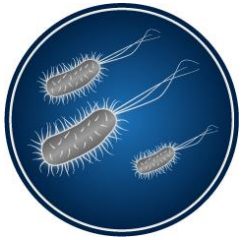
# Assessment of Supplier Capabilities

- *Continued*
  - Do they provide training programs for the end-user?
  - Do they provide on-site technical services, calibration and preventive maintenance programs?
  - Can they respond to technical issues in a timely manner?
  - Does the supplier provide validation protocols or similar documentation?
  - Have they published results of their own testing or have they submitted a Drug Master File of similar document to a regulatory agency?
  - How does the supplier manage software updates and notification to the end-user?



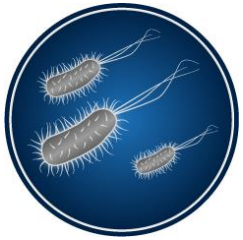
# Supplier Audit

- An assessment of whether the supplier can meet these requirements should be conducted
- A review of relevant documentation provided by the supplier
- Supplier questionnaire
- Physical audit at the supplier's manufacturing and design/development facilities
- Review of regulatory or customer audits, if available



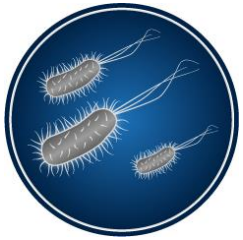
# Responsibilities

| Validation Deliverable | Supplier | User | Comments |
|---|---|---|---|
| Description of the technique (Ph. Eur. 5.1.6) | ✔ | | User shall critically review supplier information. |
| Risk benefit analysis / risk assessment | | ✔ | Supplier may list advantages over pharmacopeia. |
| Primary validation (Ph. Eur. 5.1.6) | ✔ | | User shall perform if the method is employed other than what is defined by the supplier. |
| User requirements specification (URS) | | ✔ | |
| Design qualification (DQ) | | ✔ | |
| Functional design specification (FDS) | | ✔ | |
| Requirement traceability matrix (RTM) | | ✔ | |
| SOP's and technology training | ✔ | ✔ | |
| System integration | ✔ | ✔ | |
| Installation qualification (IQ) | | ✔ | May be performed by supplier. |
| Operational qualification (OQ) | | ✔ | May be performed by supplier. |
| Performance qualification (PQ) | | ✔ | |
| Method suitability | | ✔ | |
| On-going maintenance and periodic reviews | | ✔ | |



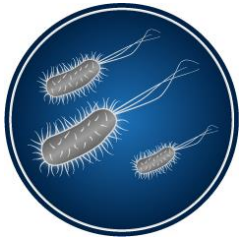
# Description of the Technique

- *Recommendation from Ph. Eur. 5.1.6*
- The supplier provides a clear description of the method
  - Demonstrated through primary validation
  - Includes conditions required for application, materials and equipment needed, and the expected signal
- The user is responsible for critically reviewing and understanding this information



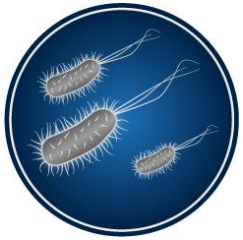
# Risk Assessment

- A risk assessment should be performed prior to the start of any RMM validation activities
- Identified risks will vary and may include:
  - The RMM supplier
  - Are RMM limitations more severe than the compendial method?
  - Product or sample requirements for evaluation
  - Alternative signal to the CFU
  - Potential for false positive or false negative results
  - Computer system capabilities and security
  - Method robustness and ruggedness
  - Equivalence to existing methods
  - Regulatory acceptance



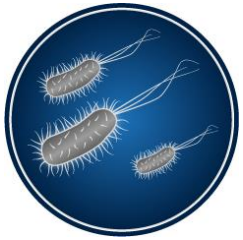
# Risk Assessment

- The evaluation of risk to quality should be based on scientific knowledge and link to the protection of the patient
- The level of effort, formality and documentation should be relative to the level of risk
- ICH Q9 describes risk management principles, approaches and QRM tools used to make risk-based decisions that are relevant to the pharmaceutical product lifecycle
- PDA Technical Report #44, *QRM for Aseptic Processes*
- Failure Modes and Effects Analysis (FMEA) or Hazard Analysis and Critical Control Points (HACCP)



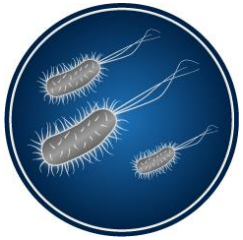
# Risk Assessment

- Identify hazards (what might go wrong). e.g.,
  - RMM instrumentation does not function or gives incorrect data
  - Availability of consumables or necessary reagents
  - Parameters for recovery of organisms are different from existing method
- Determine the likelihood of occurrence and severity of harm for each of the hazards identified
- The ability to detect the hazard may also be included
- Analyze the risk against predefined criteria
- The output is a quantitative risk score or a qualitative risk ranking (low/medium/high)



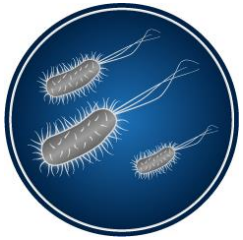
# Risk Assessment

- Risk Control: decide how the risks will be addressed
- Risk is acceptable
  - The process remains as designed
  - Take steps to further reduce the risk through a process improvement program
- Risk is unacceptable
  - Reduce the risk and control to an acceptable level
  - Change the process
  - Increase level of detection of the identified potential process failure
- Risk Review: once appropriate controls are implemented, ensure that no new risks have been introduced, and the controls are effective



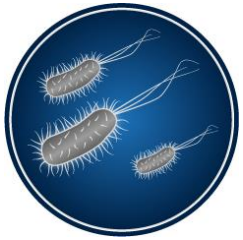
# Risk Assessment – Case Study

- 2017. Gordon, O., Goverde, M, Staerk, A., Roesti, D. Validation of Milliflex® Quantum for Bioburden Testing of Pharmaceutical Products. *PDA Journal of Pharmaceutical Science and Technology*. 71(3):206-224.
- Performed a risk analysis to determine what aspects of the RMM required validation and which aspects did not



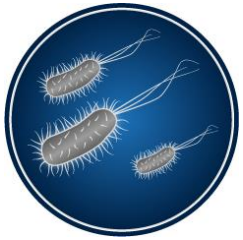
# Risk Assessment – Case Study

| Process | Existing Method | Rapid Method | Difference |
|---|---|---|---|
| Sample prep | Filtration 0.45 µm | Filtration 0.45 µm | None |
| Media | TSA, SDA | TSA, SDA | None |
| Sample volume | 1-200 mL | 1-200 mL | None |
| Incubation | TSA: 30-35° C<br>SDA: 20–25° C | TSA: 30-35° C<br>SDA: 20–25° C | None |
| Incubation time | TSA: 3-5 days<br>SDA: 5-7 days | Minimum 72 hours | **Yes; must be validated** |
| Viable count acceptance level | 10 CFU/100 ml to 100 CFU/1 ml depending on product | 10 CFU/100 ml to 100 CFU/1 ml depending on product | None |
| Evaluation of test result | Operator manually counts visible CFUs on membrane | Membranes are stained; fluorescing colonies are visually observed in Quantum system and counted by operator | **Yes; must be validated** |



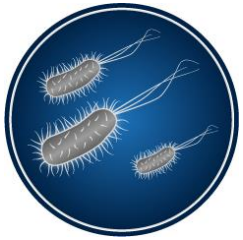
# Primary Validation (Ph. Eur. 5.1.6)

- Responsibility of the technology supplier
- Demonstrated by challenging the method with a panel of microorganisms appropriate for the method's intended use
- Criteria assessed will be dependent on whether the method is qualitative, quantitative or will provide a microbial identification
- The end-user should review data associated with the supplier's primary validation, the principle of detection, materials and equipment required and the expected output or signal



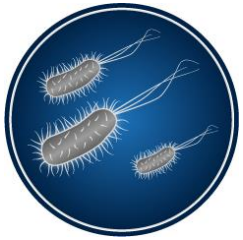
# User Requirements Specification (URS)

- Describes the functions the method must meet
- Forms the basis for ultimately selecting a method
- Defines how the method will be validated via test protocols and acceptance criteria
- For example…



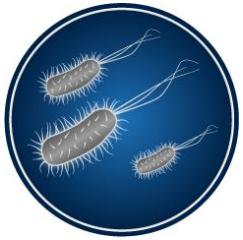
# User Requirements Specification

- The type of analysis that will be performed (e.g., qualitative, quantitative or identification)
- Scope of application
- The required level of sensitivity (i.e., limit of detection or quantitation)
- Specificity or range of microorganisms to be detected
- The number and type of samples to be evaluated
- Required time to result or detection
- Data management capabilities, including compatibility with external IT platforms and servers
- Preventive maintenance and calibration



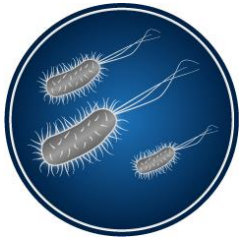
# Design Qualification (DQ)

- Documented verification that the design of the equipment is suitable for the intended purpose
- The user is responsible for verifying the equipment has been appropriately designed to meet the URS
- This should be relatively easy to perform as most RMMs are commercial off-the-shelf systems
- This activity can be completed prior to purchasing the RMM system or can be incorporated into the formal validation plan



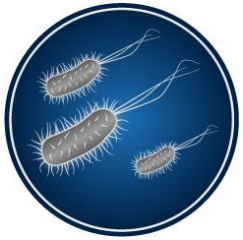
# Functional Design Specification (FDS)

- The FDS describes all of the functions and requirements for the RMM system and what will be tested to ensure that the system performs as specified in the URS
- Can cover system functionality, configuration, input/outputs, environment, utilities, computer and communication architecture, interfaces, data and security
- The FDS will point to specific test scripts where each requirement will be evaluated and verified against pre-established acceptance criteria (e.g., IQ, OQ or PQ)
- This is your validation roadmap



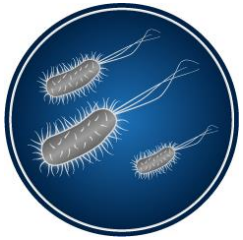
# Example of FDS

- Purpose, scope and description of the RMM
- Documentation
  - User manuals, guidelines, standards, SOPs
- Physical specifications
  - Size, electrical power, voltage frequency, operating temperature, environmental requirements, utility requirements
- Computer system specifications
  - Processor, hard drive, RAM and video graphics, network address and connections, operating software, printer ports, software and algorithms, databases, storage devices, peripherals
- Security specifications
  - User ID and password, access to data, account privileges, record retention, audit trail, administrative control, data view and print reports, data transfer to a dedicated server, 21 CFR Part 11, inactivity logoff procedure



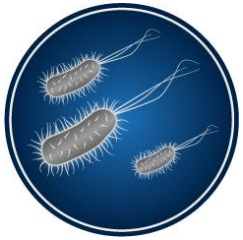
# FDS Example Sections

- Functional specifications
  - Accuracy, precision, specificity, limit of detection, limit of quantification, linearity, range, ruggedness, robustness, equivalency and all the required microbiology testing
- Databases and libraries that will be used
- Reports
  - Electronic and/or printed, report views, storage
- Functions that will not be used (or tested)
- System customization, database updates
- System availability



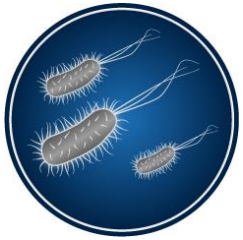
# FDS Example Sections

- Alarm configuration and error handling
  - Potential alarm messages and their respective causes, and error messages and how they are handled
- May also include identification of critical and non-critical functions that may adversely affect product quality and/or GMP records of the product, if not working properly or not available



# Requirements Traceability Matrix (RTM)

- The RTM provides traceability that all the requirements listed in the FDS have been verified and/or tested
- This is your validation checklist
- The RTM also specifies which SOPs and other documentation that needs to be in place in order to satisfy the criteria for meeting a specific function or requirement
- This is a living document during the execution of the validation test scripts or protocols
- Can also be merged with the FDS

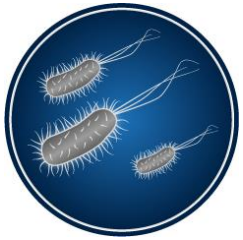


# Standard Operating Procedures

- SOPs that facilitate the proper use of the equipment/instrumentation, analytical methods and those that are required to be in place as specified in the FDS should be written and approved prior to the execution of the validation plan
- May also include:
  - Calibration and preventive maintenance
  - System and software security, data management, backup and recovery
  - Change control
  - Corrective and Preventive Actions (CAPA)
  - Business contingency plans

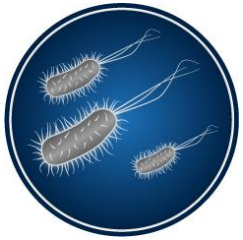
# Technology Training

- Training and the qualification of analysts are required for the effective execution of the testing protocols and are critical to the success of the overall validation plan
- Training usually occurs on-site or at the supplier’s facility
- May also be part of the method’s initial commissioning activities



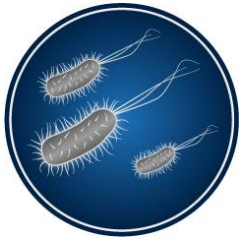
# System Integration

- Information technology (IT) and computer systems
- Bringing together all of the component subsystems into a single, operating system and ensuring that all the components function appropriately
- Laboratory Information Management System (LIMS)
- External server
- System integration testing may be required; consult with your IT group



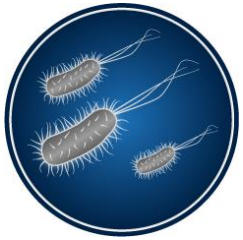
# Optional Acceptance Testing

- Some firms require Factory Acceptance Testing (FAT) or Site Acceptance Testing (SAT) prior to accepting equipment/instrumentation
  - A FAT is performed at the supplier's facility to ensure the system meets design criteria prior to the system being shipped
  - FAT may be appropriate when the end-user cannot test certain requirements, when custom made systems have been built or when the safety of the end-user may be at risk (rare for a RMM!)
  - A SAT may be performed when the system arrives at the end-user's facility to ensure that the system operates properly after shipping (e.g., it is not damaged)



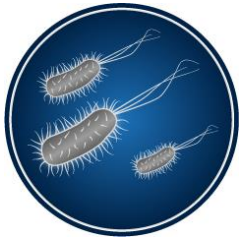
# Installation Qualification (IQ)

- The IQ establishes the equipment is received as designed and specified, that it is properly and safely installed with the correct utilities in the selected environment, and that the environment is suitable for the operation and use of the equipment
- The IQ can be carried out by the RMM supplier (during installation) and/or by the end-user (especially if a more extensive IQ is required by the user's firm)



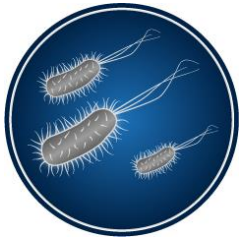
# Installation Qualification (IQ)

- Equipment descriptions
- Operating environmental conditions
- Calibration requirements
- Establishment of an equipment log book
- Safety features
- Required utilities, power and wiring
- Computer system capabilities, configuration and access
- Secure server installation and communication
- Firmware and software installation and access
- Data backup and recovery
- Supporting documentation (SOPs, manuals, blueprints)



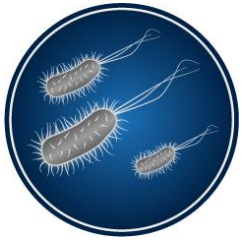
# Operational Qualification (OQ)

- The OQ provides documented verification that the equipment performs effectively and reproducibly as intended throughout the anticipated or representative operational ranges, defined limits and tolerances
- For example, verification of specified heating or cooling rates, adequate performance of optical systems or proper functioning of the user interface
- The OQ is the focal point for the majority of the computer system validation (CSV), including hardware, software and security testing
- Can be performed by the end-user and/or the supplier



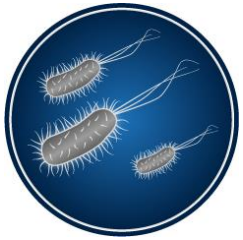
# Operational Qualification (OQ)

- Depending on the complexity of the technology and the end-user's company policies, the CSV can be quite extensive and may include:
  - Administrator control and operator access, user ID and password set up, user and system lockout, data archiving and access, audit trails, report generation, data transfer and server communication, electronic signatures, data backup and recovery, database management and integrity, interference (radio frequency, electromagnetic, wireless)
- Refer to guidance from EU Annex 11, 21 CFR Part 11, and Good Automated Manufacturing Practice (GAMP)



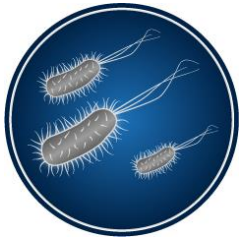
# Performance Qualification (PQ)

- Demonstrates the installed equipment and the microbiological method provide results in accordance with expected outcomes and criteria
- Verifying the results obtained by the supplier using a panel of microorganisms (e.g., validating the method)
  - Specificity and detection limit for qualitative methods
  - Accuracy, precision, quantitation limit, linearity and range for quantitative methods
- Verifying the method for its intended use
  - e.g., sterility testing, total aerobic microbial count, etc.
  - Compatibility with the test sample (method suitability testing)
  - Demonstrating equivalence with the compendial method



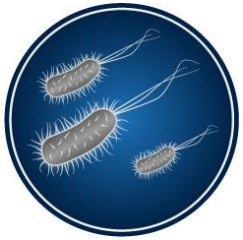
# Ongoing Maintenance and Periodic Reviews

- Following equipment verification and method validation, procedures should be established to maintain the system in a validated state
  - SOPs
  - Change control
  - Preventive maintenance and calibration
  - Software updates
  - Re-qualification, when appropriate
- Periodic review of compliance with cGMPs
  - Outputs should feed back into the initial risk assessment



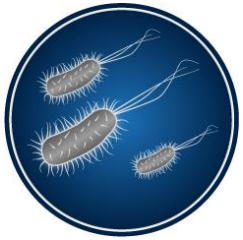
# Establishment of Method Validation Criteria

- USP <1223>, *Validation of Alternative Microbiological Methods*. Effective December 2015.
- Ph. Eur. chapter 5.1.6, *Alternative Methods for Control of Microbiological Quality*. Effective July 2017.
- PDA Technical Report No. 33 (TR33), *Evaluation, Validation and Implementation of Alternative and Rapid Microbiological Methods*. Revised 2013.



# Establishment of Method Validation Criteria

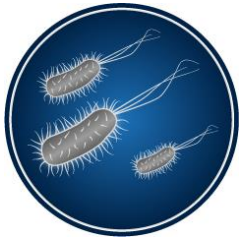
- Validation of the method is performed using standardized microbial suspensions
- There may be unique situations when this is not possible and other strategies may be considered
  - Technologies that cannot introduce liquids into their systems
- Testing may also be designed where the data from one study may be used for other validation criteria
  - e.g., Linearity and Accuracy



# Validation Criteria for Qualitative Methods

| Validation Criteria | USP <1223> | Ph. Eur. 5.1.6 | PDA TR33 |
|---|---|---|---|
| Accuracy | | * | |
| Precision | | | |
| Specificity | ✔ | ✔ | ✔ |
| Limit of detection | ✔ | ✔ | ✔ |
| Limit of quantification | | | |
| Linearity | | | |
| Operation (dynamic) range | | | |
| Ruggedness | ✔ | | ✔ |
| Robustness | ✔ | ✔ | ✔ |
| Repeatability | ✔ ** | | |
| Equivalency | ✔ (without product) | ✔ (without product or with product or do both) | ✔ (with product) |

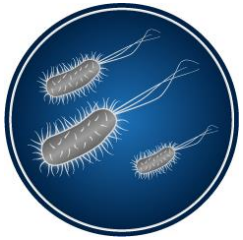
* Accuracy can be used instead of the limit of detection test. ** Same as precision?



# Validation Criteria for Quantitative Methods

| Validation Criteria | USP <1223> | Ph. Eur. 5.1.6 | PDA TR33 |
|---|---|---|---|
| **Accuracy** | ✔ | ✔ | ✔ |
| **Precision** | ✔ | ✔ | ✔ |
| **Specificity** | ✔ | ✔ | ✔ |
| **Limit of detection** | ✔ | * | ✔ |
| **Limit of quantification** | ✔ | ✔ | ✔ |
| **Linearity** | ✔ | ✔ | ✔ |
| **Operation (dynamic) range** | ✔ | ✔ | ✔ |
| **Ruggedness** | ✔ | (intermediate precision) | ✔ |
| **Robustness** | ✔ | ✔ | ✔ |
| **Repeatability** | ✔ ** | (addressed under Precision) | (addressed under Precision) |
| **Equivalency** | ✔ (without product) | ✔ (without product or with product or do both) | ✔ (with product) |

***** May be needed in some cases. ****** Same as precision?



# Validation Criteria for Identification Methods

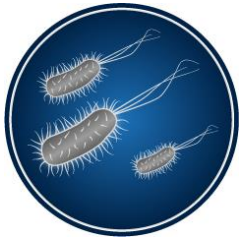
| Validation Criteria | USP <1223> | Ph. Eur. 5.1.6 | PDA TR33 |
|---|---|---|---|
| Accuracy | ✔ | ✔ | ✔ |
| Precision | ✔ | | ✔ |
| Specificity | | ✔ | |
| Limit of detection | | | |
| Limit of quantification | | | |
| Linearity | | | |
| Operation (dynamic) range | | | |
| Ruggedness | | | |
| Robustness | | ✔ | |
| Equivalency | | | |

# Validating the Method Requires the Use of Statistics

**This is why your statistician should be involved.**

# Statistics and RMM Validation

- Regulators (FDA, EMA) want the industry to prove equivalence between alternative and compendial methods using statistics
- There is guidance (and some confusion) in PDA TR33, USP 1223 and Ph. Eur. 5.1.6



# Statistics and RMM Validation

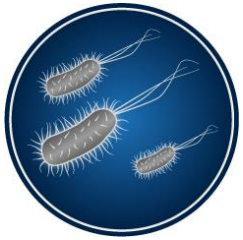
- We will use a **Hypothesis Test**
- Generally involves 4 steps:

1. Formulate the null and alternative hypotheses
2. Identify an appropriate statistical test to use
3. Calculate the p-value
4. Compare the p-value to an acceptable significance value

# What is a Hypothesis?

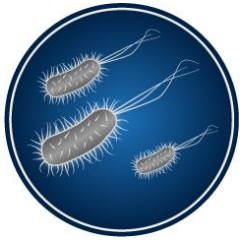
- Proposed explanation for some phenomenon
- Made on the basis of limited evidence
- We assume it is true
- Used as a starting point for further investigation

e.g., Flipping a coin will have a 50-50 chance of Heads or Tails



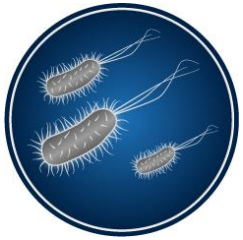
# 1. Formulate Your Hypothesis

- Formulate the null hypothesis ($H_0$)
  - Assume the % of Heads and Tails is the same
- Formulate an alternative hypothesis ($H_1$ or $H_a$)
  - e.g., the % of Heads and Tails is statistically different
- Statisticians want to prove the alternative hypothesis
- Simply accepting the null as being true requires no effort, but to disprove it requires experimentation



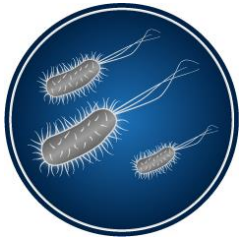
# Example: Null Hypothesis

- Consider a trial. The trial is a hypothesis test.
  - $H_0$ = defendant is innocent
  - $H_1$ = defendant is guilty
- The prosecution needs to prove, beyond a reasonable doubt, the defendant is guilty
- If this cannot be proven, it does not mean the defendant is innocent. But, based on the evidence, we cannot reject that possibility.
- Therefore, the verdict is “the defendant is not guilty”



# Two Outcomes are Possible

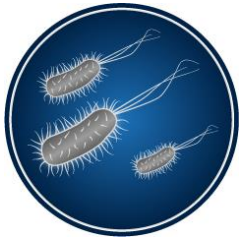
- Reject the null hypothesis and accept the alternative, or
- Fail to reject the null hypothesis



# 2. Identify a Statistical Test

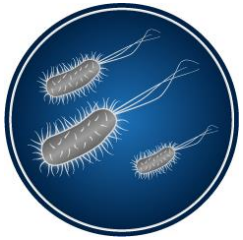
- T-test, Chi-square, Fisher's Exact Test, ANOVA, etc.
- Each test will provide a test statistic, which is a standardized value calculated from the sample data
- The test statistic will be used to reject or fail to reject the null hypothesis

| Hypothesis Test | Test Statistic |
|---|---|
| t-tests | t-statistic |
| ANOVA | F-statistic |
| Chi-square tests | Chi-square statistic |



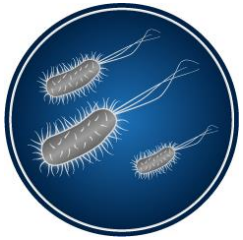
# t-Test

- Determines if the mean of a population significantly differs from the mean of another population or a specific value
- Used on two samples
- The “t-statistic” measures how far the data are spread out from their mean
- For example, does the mean microbial recovery of a suspension of *S. auerus* significantly differ when a RMM is used vs. a standard plate count?



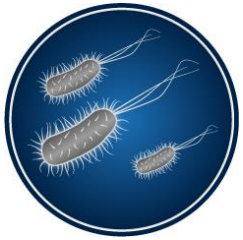
# Chi-Square Test

- Determines whether there is a significant association or relationship between two variables
- For example, does the number of positive to negative results in a presence/absence test significantly differ when a RMM is used vs. a conventional method?



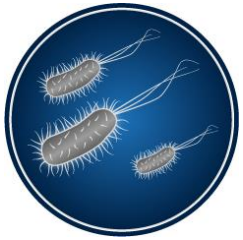
# Chi-Square Test

- The Chi-square test is an *approximation*
- Results can be inaccurate when data values are very small
- General rule of thumb: do not use Chi-square when the expected values are below 5



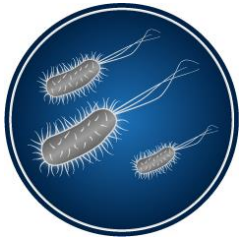
# Fisher's Exact Test

- For small values, use a Fisher's Exact Test instead
- The test is accurate for all sample sizes and can be calculated *exactly*



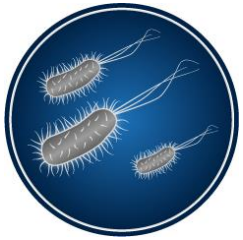
# Analysis of Variance (ANOVA)

- ANOVA determines whether the means of three or more groups are different
- Uses the F-statistic to assess equality of the means (how far the data are scattered from the mean)
- It is similar to a two-sample t-test
  - If you only have 2 groups, just use the t-test
- For example, is the mean RMM count significantly different than the mean plate count when 3 analysts perform the test?



# 3. Calculate the p-Value

- The test statistic is used to calculate a p-value
- P-value is the probability of rejecting the null hypothesis when it is actually true
- The lower the p-value, the stronger the evidence the null hypothesis is not true (i.e., we reject the null and accept the alternative)
- The greater the p-value, the null hypothesis is likely to be true (i.e., we fail to reject the null)
- What defines a low vs. a high p-value?

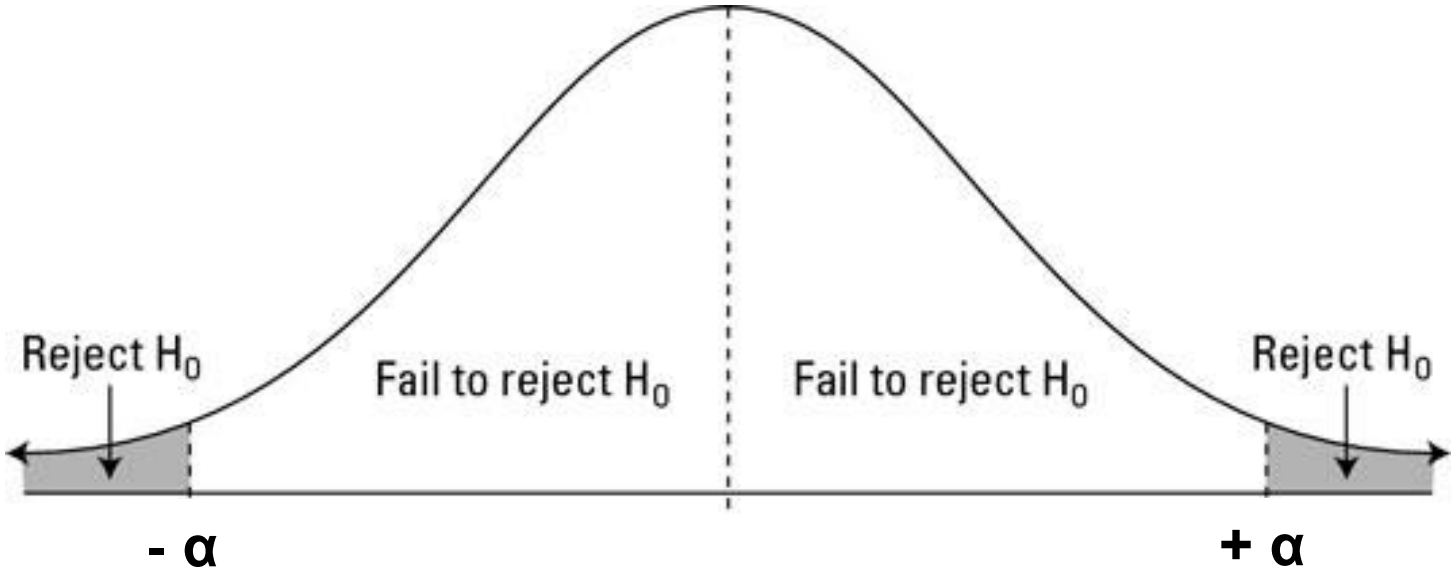


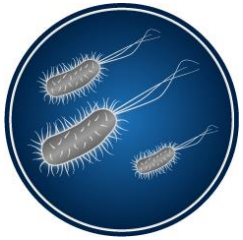
# 4. Compare the P-Value to an Acceptable Significance Value

- We need a cutoff point that will determine when the p-value rejects or fails to reject the null hypothesis
- This cutoff point is a pre-defined threshold value known as **α** (alpha) or the level of significance associated with a probability distribution of data
- Assume a normal distribution data curve …

# Compare the P-Value to an Acceptable Significance Value

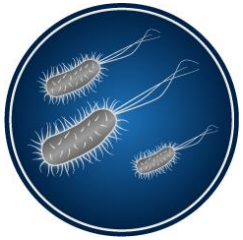
- The p-value is the area in the tails of the curve
- **α** is the cutoff point in the tails
- So, what is an acceptable **α** level?





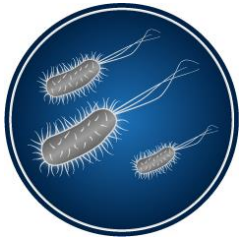
# Compare the P-Value to an Acceptable Significance Value

- Most of us use a 95% level of significance ($\alpha = 0.05$)
- We are willing to accept a 5% chance that we are wrong when rejecting the null hypothesis
- If the p-value is ≤ 0.05, we reject the null hypothesis and accept the alternative hypothesis
- If the p-value is > 0.05, we fail to reject the null hypothesis



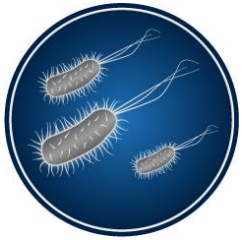
# Hypothesis Testing and RMMs

- In traditional hypothesis testing, the null hypothesis typically implies the sample populations are the same
  - $H_0$ = Both methods detect a similar number of positive/negative results in a sterility test, ***or***
  - $H_0$ = Both methods recover a similar number of cells in a bioburden assay
  - $H_a$ = The methods are statistically different
- If the p-value ≤ 0.05, we reject the null
- If the p-value > 0.05, we fail to reject the null
- **Does this prove the methods are the same?**



# Hypothesis Testing and RMMs

- No, at least not statistically. Remember the trial.
- We haven’t proven the two methods are the same; we just do not have enough evidence to prove they are different
- But, don’t statisticians want to prove the alternative hypothesis?
- Yes, and an “Equivalence” test solves this problem



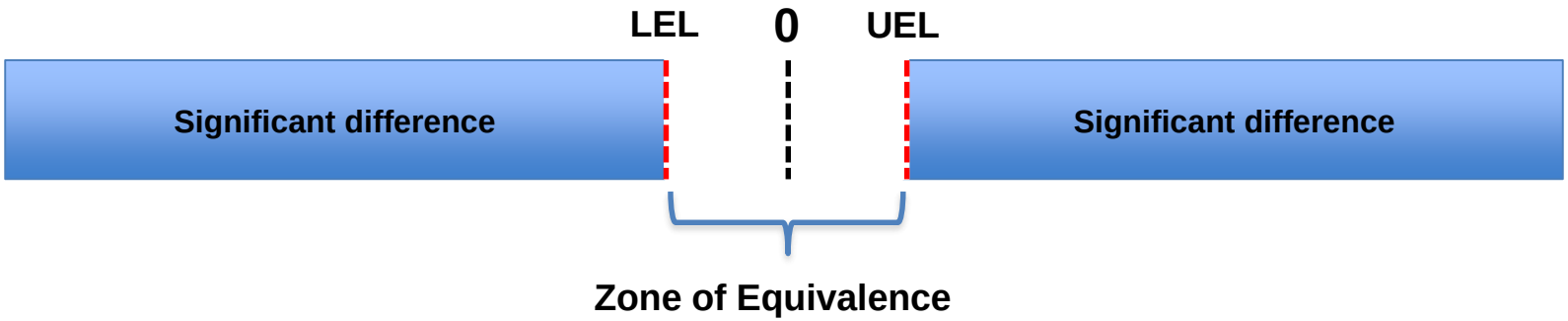
# Equivalence Test

- In an equivalence test, $H_0$ and $H_1$ are reversed:
  - **$H_0$**: both methods <u>are different</u>
  - **$H_1$**: both methods <u>are the same</u>
- This provides a more rigorously statistical test ***proving*** the alternative (both methods are the same)
- An equivalence test requires additional steps: we must decide on the size of an acceptable difference between the two populations
- This is called the “zone of equivalence”

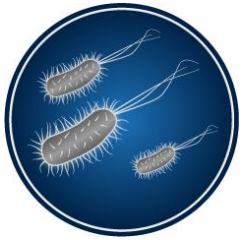
# Zone of Equivalence

LEL: Lower equivalence level

UEL: Upper equivalence level




- If the confidence interval (e.g., 95%) for the difference in the populations is within LEL and UEL, we claim equivalence
- Also, if the p-value for the upper and lower bounds is ≤ 0.05, we claim equivalence
- So, what is an acceptable LEL and UEL?

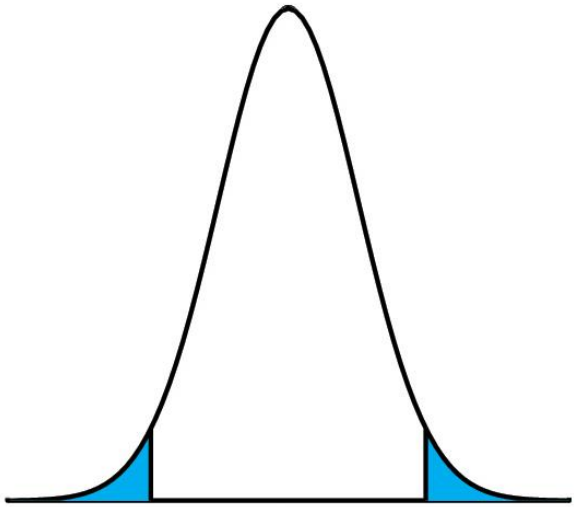


# Selecting the LEL and UEL (Δ)

- USP <1223> suggests using a Δ = 0.2 for presence/absence tests (e.g., Limit of Detection)
- Δ = 0.3 for a quantitative test (e.g., for Accuracy counts based on 70% recovery)
  - 2015. Evaluation of PDA Technical Report No 33. Statistical Testing Recommendations for a Rapid Microbiological Method Case Study. Murphy T, Schwedock J, Nguyen K, Mills A, Jones D. PDA J Pharm Sci Tech. 69(4):526-39.
- Δ = 0.1 for false positive rates of a RMM (e.g., the limit for false positives should not exceed 10%)
  - 2015. IJzerman-Boon PC, van den Heuvel ER. Validation of qualitative microbiological methods. Pharmaceutical Statistics. 44(2):120-128.

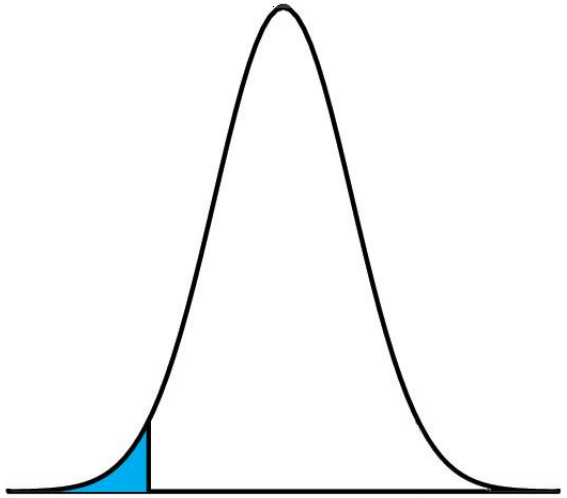
# Equivalence Test

- What happens if the RMM is more sensitive than the compendial method (higher counts)?
- The equivalence test may be penalized at the upper tail due to better recovery of microorganisms
- Therefore, the methods would not be statistically equivalent
- In this case, it is better to use a non-inferiority test



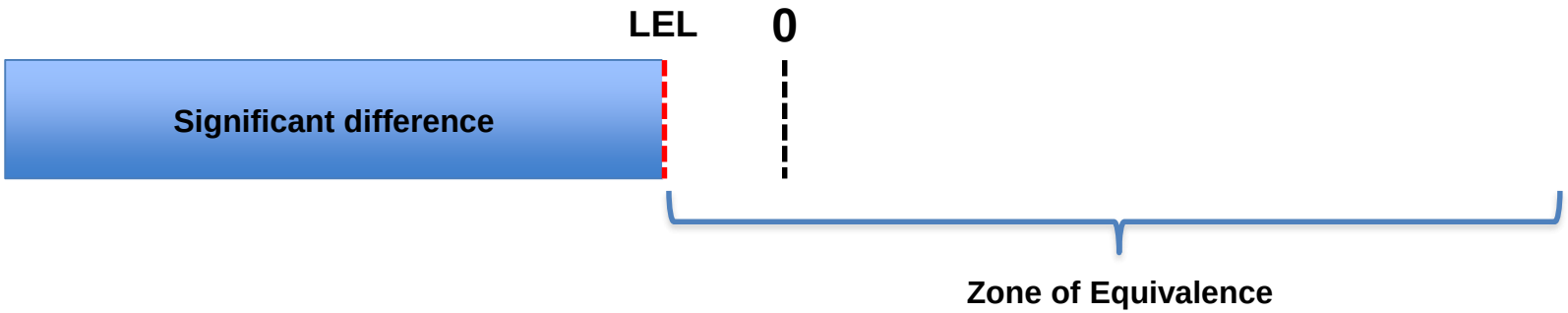
# Non-Inferiority Test

- Demonstrates one method is not worse (non-inferior) than another method
- Uses a one-tailed analysis
- For example, the mean count in a RMM can be higher than in the compendial method; however, it cannot be statistically lower

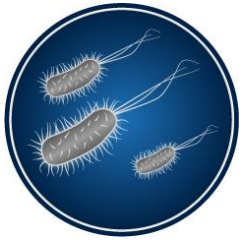


# Zone of Equivalence

LEL: Lower equivalence level

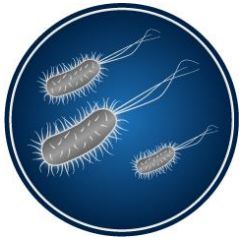



- If the confidence interval (e.g., 95%) for the difference in the populations is above the LEL, we claim non-inferiority
- Also, if the p-value for the lower bound is ≤ 0.05, we claim non-inferiority



# Mean Count Examples

- Compare the mean recovery in a RMM with the standard plate count during Accuracy studies
- Directly demonstrate at least a 70% recovery
  - Recommended in USP 1227 for method suitability
  - Represents a 30% or 0.3 margin of variability, or ~ 0.5 logs
- Or, we can statistically compare the data from both methods



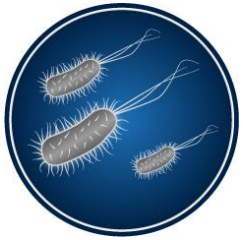
# Accuracy Example 1

- Assume the following data for a bacterial suspension:

**RMM counts = 20, 19, 14, 14, 18 (mean = 17.0)**

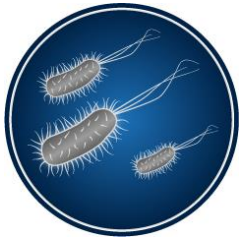
**Plate count (CFU) = 18, 19, 15, 12, 15 (mean = 15.8)**

- The RMM recovers 107.59% of the plate count (> 70%)
- **Question:** should we perform a traditional 2-sample t-test or a 2-sample equivalency or non-inferiority test?



# Accuracy Example 1: Traditional 2-sample t-test

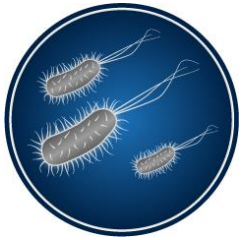
- $H_0$: mean count for RMM and plate count method are the same
- $H_1$: mean count for RMM and plate count method are different
- alpha = 0.05 (95% confidence level)
- If $p \leq 0.05$, reject $H_0$
- If $p > 0.05$, we fail to reject $H_0$
- Software package used: Minitab 17



# Accuracy Example 1: Traditional 2-sample t-test

```
      N   Mean  StDev  SE Mean
RMM   5  17.00   2.83      1.3
PC    5  15.80   2.77      1.2

Difference = μ (RMM) - μ (PC)
Estimate for difference:  1.20
95% CI for difference:  (-2.89, 5.29)
T-Test of difference = 0 (vs ≠): T-Value = 0.68 P-Value = 0.517 DF =
8
Both use Pooled StDev = 2.8018
```

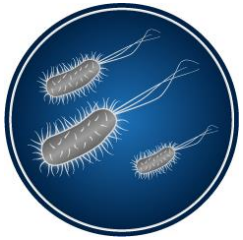
- p-value > 0.05; we fail to reject the null hypothesis
- But to prove the methods are the same, we must use an equivalency or non-inferiority test



# Accuracy Example 1: 2-sample Equivalency Test

- $H_0$: mean count for RMM and plate count method are different
- $H_1$: mean count for RMM and plate count method are the same

- If the difference in the mean count (at a 95% confidence interval) is within the LEL and UEL, claim equivalence
- If the p-value for the upper and lower tails ≤ 0.05, reject the null hypothesis and claim equivalence
- Use a $\Delta = 0.3$



# Accuracy Example 1: 2-sample Equivalency Test

| Variable | N | Mean | StDev | SE Mean |
|---|---|---|---|---|
| RMM | 5 | 17.000 | 2.8284 | 1.2649 |
| PC | 5 | 15.800 | 2.7749 | 1.2410 |

Pooled StDev = 2.80179

Difference: Mean(RMM) - Mean(PC)

| Difference | SE | 95% CI for Equivalence | Equivalence Interval |
|---|---|---|---|
| 1.2000 | 1.7720 | (-2.0951, 4.4951) | (-4.74, 4.74) |

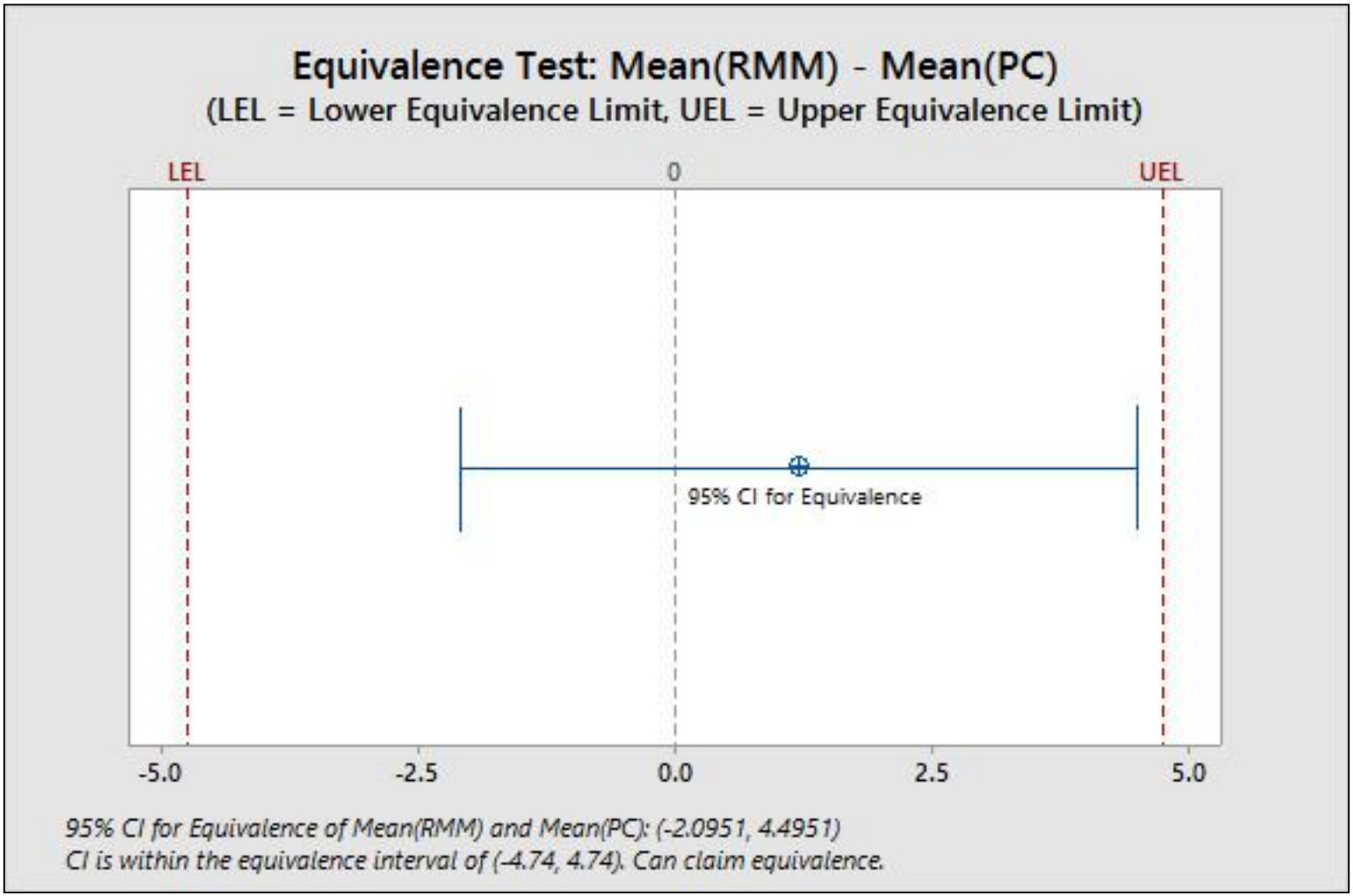
CI is within the equivalence interval. Can claim equivalence.

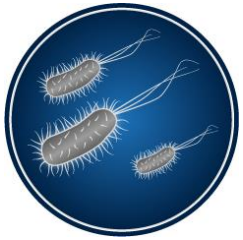
| | |
|---|---|
| Null hypothesis: | Difference ≤ -4.74 or Difference ≥ 4.74 |
| Alternative hypothesis: | -4.74 < Difference < 4.74 |
| α level: | 0.05 |

| Null Hypothesis | DF | T-Value | P-Value |
|---|---|---|---|
| Difference ≤ -4.74 | 8 | 3.3521 | 0.005 |
| Difference ≥ 4.74 | 8 | -1.9977 | 0.040 |

The greater of the two P-Values is 0.040. Can claim equivalence.

# Accuracy Example 1: 2-sample Equivalency Test





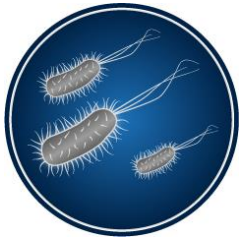
# Accuracy Example 2

- Assume the following data for a bacterial suspension:

**RMM counts = 15, 19, 12, 12, 13 (mean = 14.2)**

**Plate count (CFU) = 18, 19, 15, 12, 15 (mean = 15.8)**

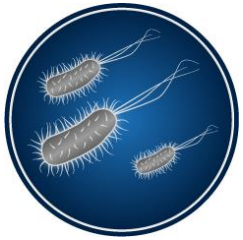
- The RMM recovers 89.87% of the plate count (> 70%)



# Accuracy Example 2: Traditional 2-sample t-test

```
        N   Mean  StDev  SE Mean
RMM_1   5  14.20   2.95      1.3
PC_1    5  15.80   2.77      1.2

Difference = μ (RMM_1) - μ (PC_1)
Estimate for difference:  -1.60
95% CI for difference:  (-5.78, 2.58)
T-Test of difference = 0 (vs ≠): T-Value = -0.88 P-Value = 0.403 DF =
 8
Both use Pooled StDev = 2.8636
```

- p-value > 0.05; fail to reject null hypothesis



# Accuracy Example 2: 2-sample Equivalency Test

| Variable | N | Mean | StDev | SE Mean |
|---|---|---|---|---|
| RMM_1 | 5 | 14.200 | 2.9496 | 1.3191 |
| PC_1 | 5 | 15.800 | 2.7749 | 1.2410 |

Pooled StDev = 2.86356

Difference: Mean(RMM_1) - Mean(PC_1)

| Difference | SE | 95% CI for Equivalence | Equivalence Interval |
|---|---|---|---|
| -1.6000 | 1.8111 | (-4.9678, 1.7678) | (-4.74, 4.74) |

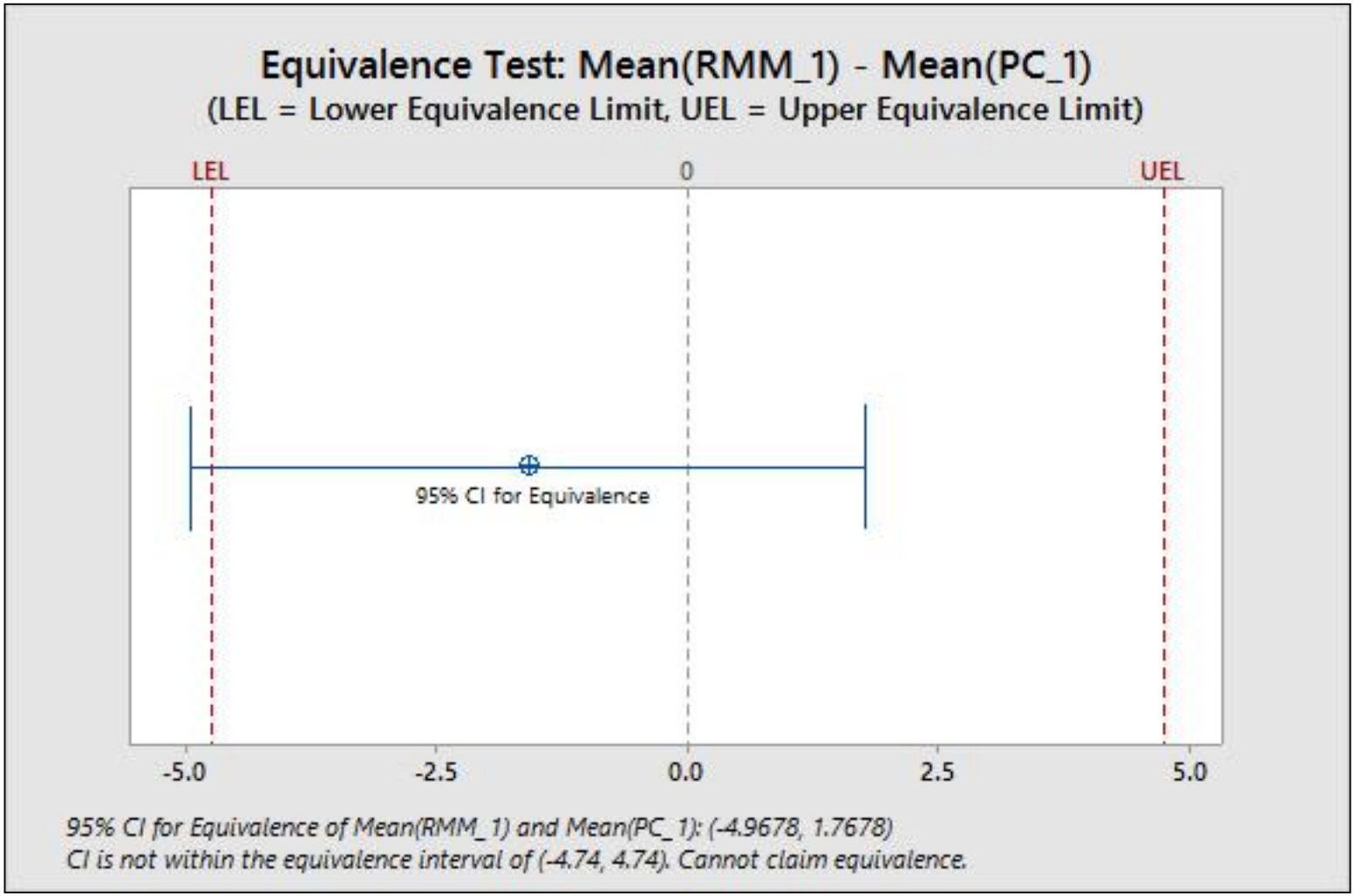
CI is not within the equivalence interval. Cannot claim equivalence.

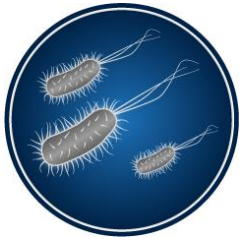
| | |
|---|---|
| Null hypothesis: | Difference ≤ -4.74 or Difference ≥ 4.74 |
| Alternative hypothesis: | -4.74 < Difference < 4.74 |
| α level: | 0.05 |

| Null Hypothesis | DF | T-Value | P-Value |
|---|---|---|---|
| Difference ≤ -4.74 | 8 | 1.7338 | 0.061 |
| Difference ≥ 4.74 | 8 | -3.5007 | 0.004 |

The greater of the two P-Values is 0.061. Cannot claim equivalence.

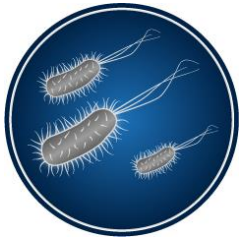
# Accuracy Example 2: 2-sample Equivalency Test





# Accuracy Example 2

- **Question:** why would the standard hypothesis test (t-test) conclude there was no difference in the two methods when the equivalency test concluded the two methods were not equivalent?
- Most likely due to the number of samples used
  - Without enough samples, it is easier to obtain a p-value > 0.05
  - In our example, we only used 5 replicates



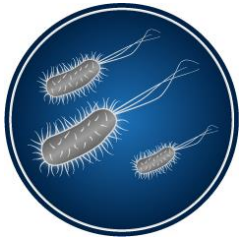
# Accuracy Example 3

- Assume the following data for a bacterial suspension:

**RMM counts = 20, 22, 17, 18, 18 (mean = 19.0)**

**Plate count (CFU) = 18, 19, 15, 12, 15 (mean = 15.8)**

- The RMM recovers 120.25% of the plate count (> 70%)

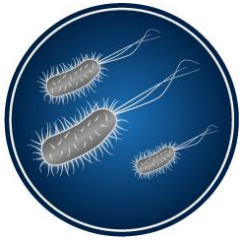


# Accuracy Example 3: Traditional 2-sample t-test

| | N | Mean | StDev | SE Mean |
|---|---|---|---|---|
| RMM_2 | 5 | 19.00 | 2.00 | 0.89 |
| PC_2 | 5 | 15.80 | 2.77 | 1.2 |

Difference = μ (RMM_2) - μ (PC_2)
Estimate for difference: 3.20
95% CI for difference: (-0.33, 6.73)
T-Test of difference = 0 (vs ≠): T-Value = 2.09 **P-Value = 0.070** DF = 8
Both use Pooled StDev = 2.4187

- p-value > 0.05; fail to reject null hypothesis



# Accuracy Example 3: 2-sample Equivalency Test

| Variable | N | Mean | StDev | SE Mean |
|---|---|---|---|---|
| RMM_2 | 5 | 19.000 | 2.0000 | 0.89443 |
| PC_2 | 5 | 15.800 | 2.7749 | 1.2410 |

Pooled StDev = 2.41868

Difference: Mean(RMM_2) - Mean(PC_2)

| Difference | SE | 95% CI for Equivalence | Equivalence Interval |
|---|---|---|---|
| 3.2000 | 1.5297 | (0, 6.0446) | (-4.74, 4.74) |

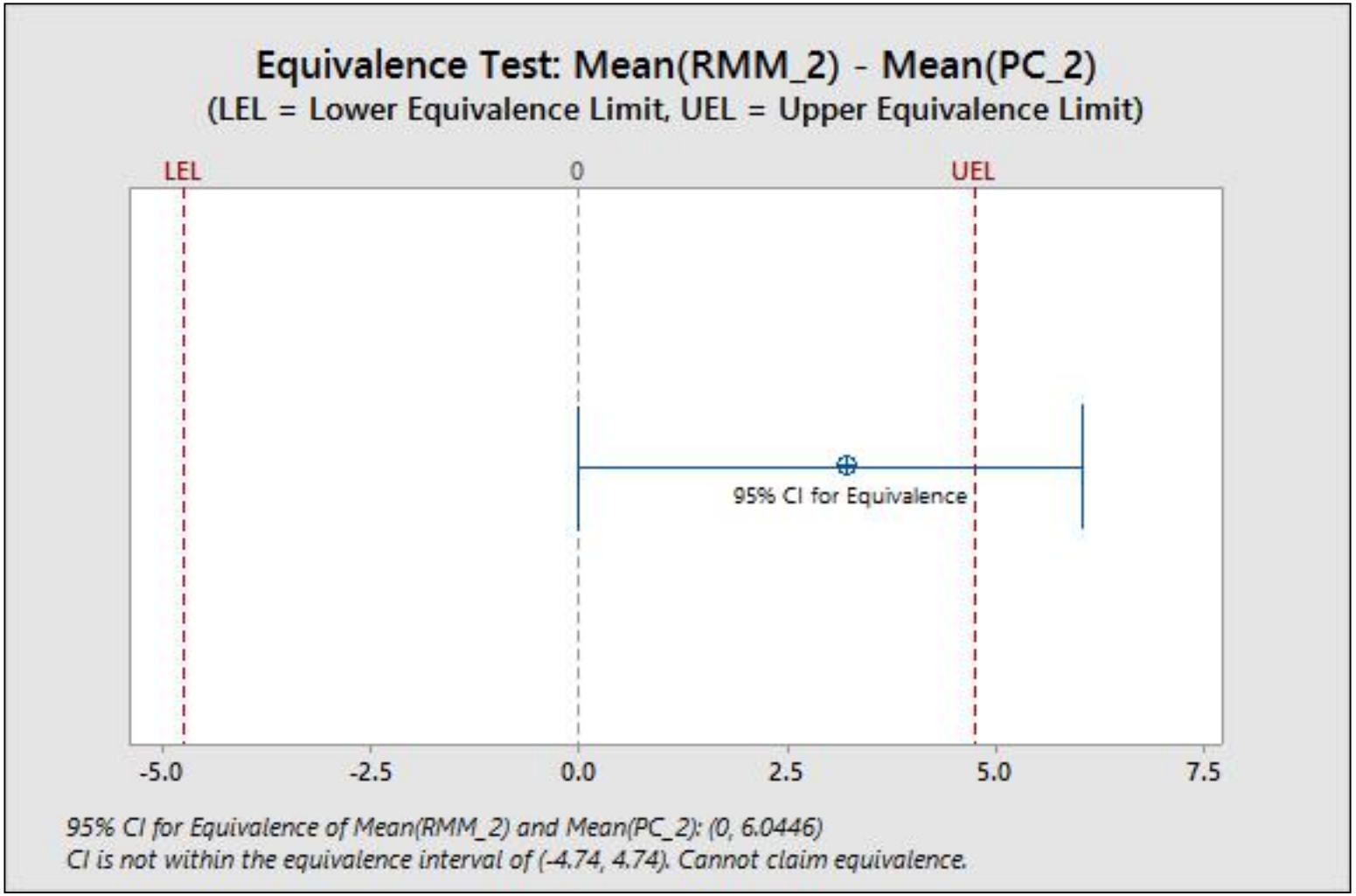
CI is not within the equivalence interval. Cannot claim equivalence.

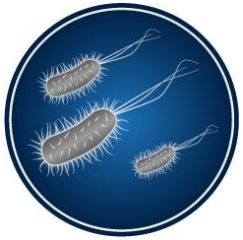
| | |
|---|---|
| Null hypothesis: | Difference ≤ -4.74 or Difference ≥ 4.74 |
| Alternative hypothesis: | -4.74 < Difference < 4.74 |
| α level: | 0.05 |

| Null Hypothesis | DF | T-Value | P-Value |
|---|---|---|---|
| Difference ≤ -4.74 | 8 | 5.1905 | 0.000 |
| Difference ≥ 4.74 | 8 | -1.0067 | 0.172 |

The greater of the two P-Values is 0.172. Cannot claim equivalence.

# Accuracy Example 3: 2-sample Equivalency Test

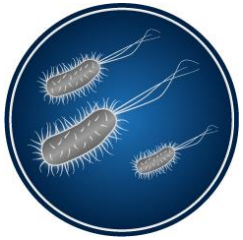




# Accuracy Example 3: 2-sample Non-Inferiority Test

- $H_0$: mean count for RMM is inferior to the plate count
- $H_1$: mean count for RMM is non-inferior to the plate count

- If the difference in the mean count (at a 95% confidence interval) is greater than the LEL, claim non-inferiority
- If the p-value for the lower tail ≤ 0.05, reject the null hypothesis and claim non-inferiority
- Use a $\Delta = 0.3$



# Accuracy Example 3: 2-sample Non-Inferiority Test

| Variable | N | Mean | StDev | SE Mean |
|---|---|---|---|---|
| RMM_2 | 5 | 19.000 | 2.0000 | 0.89443 |
| PC_2 | 5 | 15.800 | 2.7749 | 1.2410 |

Pooled StDev = 2.41868

Difference: Mean(RMM_2) - Mean(PC_2)

| Difference | SE | 95% Lower Bound | Lower Limit |
|---|---|---|---|
| 3.2000 | 1.5297 | 0.35544 | -4.7400 |

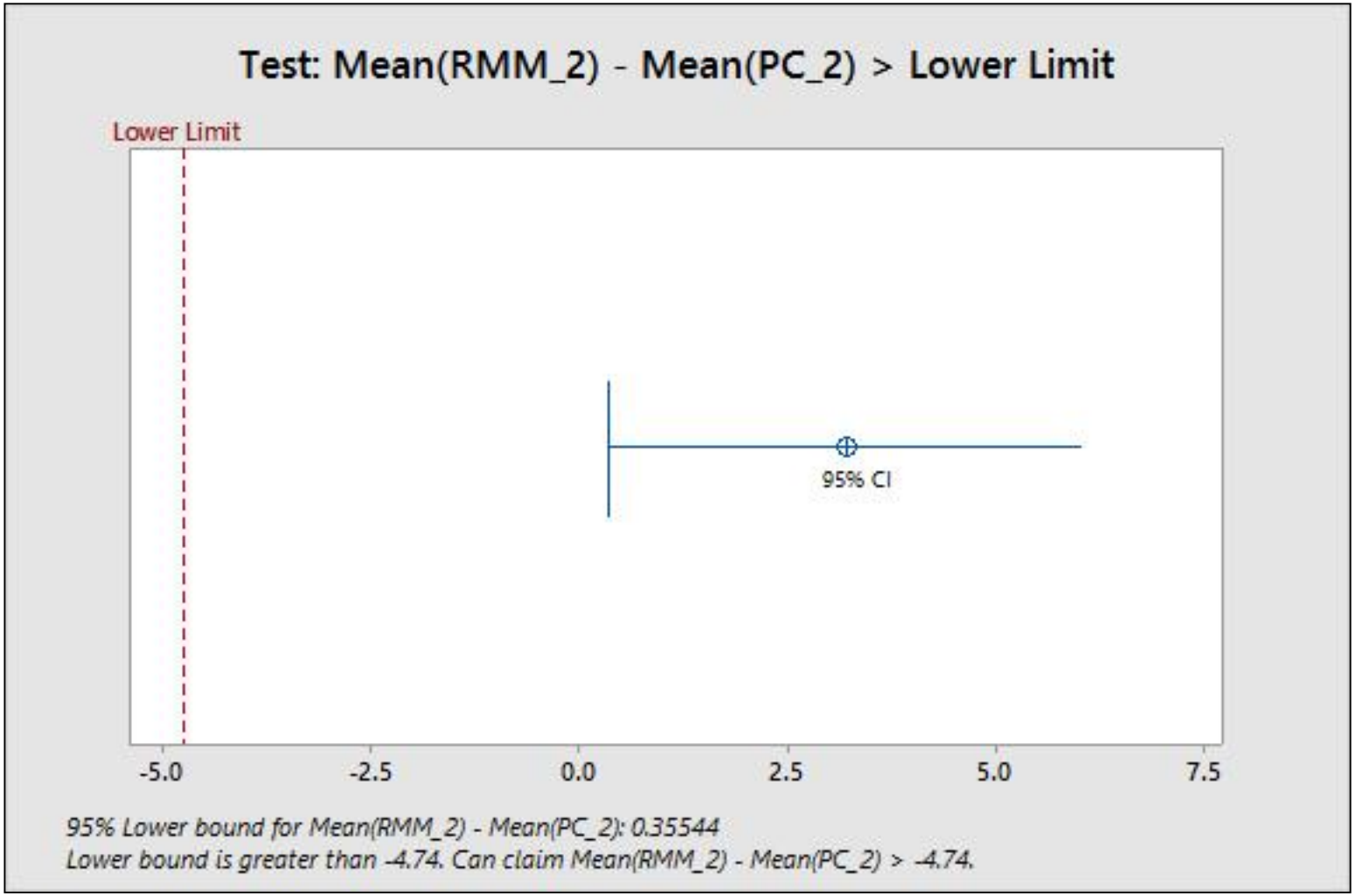
Lower bound is greater than -4.74. Can claim Mean(RMM_2) - Mean(PC_2) > -4.74.

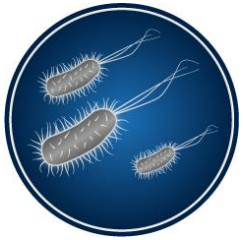
| | |
|---|---|
| Null hypothesis: | Mean(RMM_2) - Mean(PC_2) ≤ -4.74 |
| Alternative hypothesis: | Mean(RMM_2) - Mean(PC_2) > -4.74 |
| α level: | 0.05 |

| DF | T-Value | P-Value |
|---|---|---|
| 8 | 5.1905 | 0.000 |

P-Value ≤ 0.05. Can claim Mean(RMM_2) - Mean(PC_2) > -4.74.

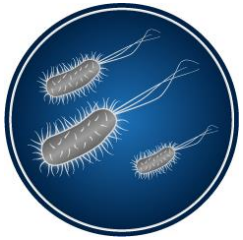
# Accuracy Example 3: 2-sample Non-Inferiority Test





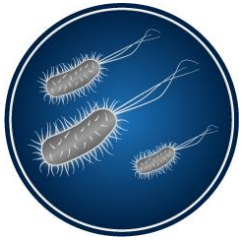
# Review

- Use an Equivalency or Non-Inferiority test when you *want to prove* a RMM is statistically the same (or not worse) than the conventional method
- If you want to use traditional hypothesis tests (t-test, chi-square, etc.), make certain an appropriate number of replicates are included
- Engage your statistician
- Use a suitable statistical software package (e.g., Minitab 17)



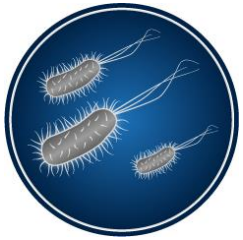
# Stats Reference


- 2016. Miller, M.J., van den Heuvel, E.R., Roesti, D. The Role of Statistical Analysis in Validating Rapid Microbiological Methods. European Pharmaceutical Review. 21(6): 46-53.



# Validation Criteria

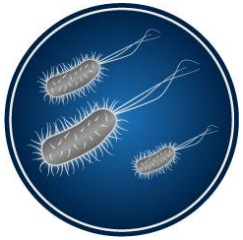
- The following sections will discuss validation criteria for qualitative and quantitative methods
- Identification methods will be discussed later in this course
- The criteria are used to validate the analytical method but can also be used during Equivalency testing



# Accuracy

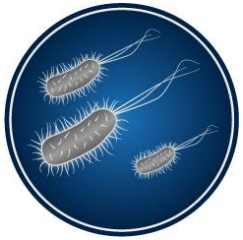
| Method | USP <1223> | Ph. Eur. 5.1.6 | PDA TR33 |
|---|---|---|---|
| **Qualitative** | | * | |
| **Quantitative** | ✔ | ✔ | ✔ |

* Accuracy can be used instead of the limit of detection (LOD) test.



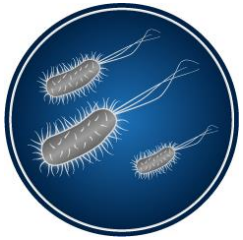
# Accuracy – USP 1223

- Closeness of the test results obtained by the alternative test method to the value obtained by the compendial method, to be demonstrated across the dynamic (operational) range of the method
- USP does not provide a specific procedure for testing Accuracy



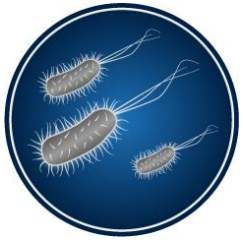
# Accuracy – Ph. Eur. 5.1.6

- Closeness of the test results obtained by the alternative method to those obtained by the pharmacopoeial method
- Demonstrated across the practical range of the test
- Usually expressed as the % recovery of microorganisms by the alternative method compared to the % recovery using the pharmacopoeial method, taking into account statistical analysis



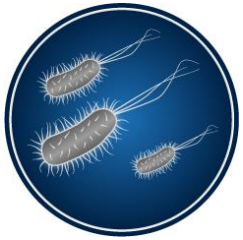
# Accuracy – Ph. Eur. 5.1.6

- Prepare a suspension of microorganisms at the upper end of the range of the test and serially dilute down to the lower end of the range
  - If the plate count method will be replaced, the range might be $10^0$-$10^6$ CFU/mL
  - If the MPN method will be replaced, a narrower range can be used
- Analyze at least 1 suspension for each test microorganism dilution
- Suspensions for both methods are counted at the same time



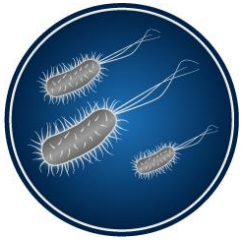
# Accuracy – Ph. Eur. 5.1.6

- The alternative method should recover at least as many organisms as the pharmacopoeial method using appropriate statistical analysis
- The protocol used to test Linearity may also be used for Accuracy
- For a qualitative method, Accuracy may be used in place of LOD by comparing the rate of positive and negative results produced by both methods for a statistically suitable number of identical samples using a standardized, low-level inoculum



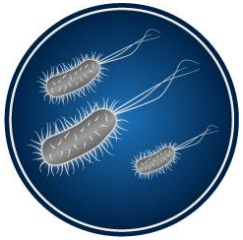
# Accuracy – PDA TR33

- The closeness of the actual test results obtained by the new method to the actual test results obtained by the existing method
- Prepare a suspension of organisms in a suitable diluent at the upper end of the range of the test and serially dilute down to the lower end of the range
- Analyze at least 5 suspensions
- Test suspensions with a suitable number of replicates (e.g., in triplicate), especially at lower concentrations where variability may be more pronounced



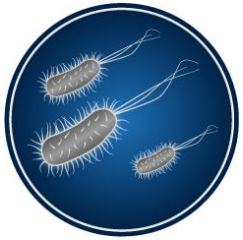
# Accuracy – PDA TR33

- Counts are obtained for each organism suspension and in each method
- Determine the % recovery by comparing the mean counts from the new method with the mean counts from the existing method
- For each suspension, the new method should provide a mean recovery count not less than 70% of the mean recovery count provided by the existing method
- Alternatively, perform a statistical comparison



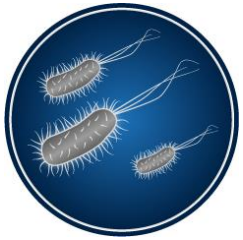
# Accuracy – PDA TR33

- Equivalency or Non-Inferiority tests
- t-test
  - When variances are equal
  - Use a Welch's correction when variances are not equal
  - If counts do not follow a normal distribution:
  - Transform data into a Gaussian distribution (e.g., $\log_{10}$)
  - If counts still do not follow a normal distribution, apply a non-parametric test (e.g., Mann-Whitney or Wilcoxon test)
- Use statistical software to determine if the data is normally distributed



# Accuracy – PDA TR33

- Analysis of Variance (ANOVA)
  - Used when more than two groups of data are compared
    - One dilution, one organism, 5 replicates, 3 analysts
  - May transform data to a Gaussian distribution (e.g., $\log_{10}$)
  - If transformed counts do not follow a normal distribution, apply a non-parametric test (e.g., Kruskal-Wallis one-way analysis of variance test)
- Tukey's test for data sets that are significantly different

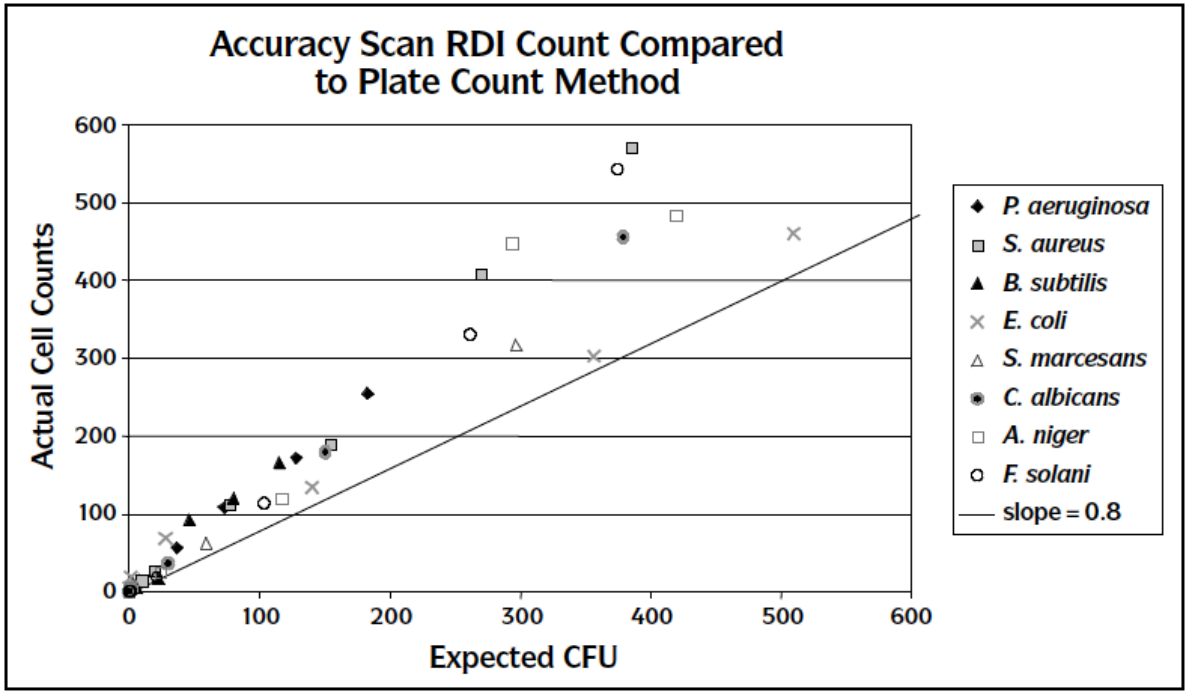


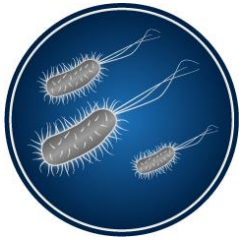
# Accuracy – PDA TR33

- Non-growth-based methods may recover a higher number of organisms than the existing method
- In this case, do not establish an upper level during Accuracy studies (i.e., use a Non-Inferiority test)
- The end-user may establish an acceptable upper limit to ensure that recovery counts are acceptable and not due to sample, method or instrumentation interference or background noise

# Accuracy Examples

- 2005. Validation of the ScanRDI® for Purified Water Testing. McCormick, P.J.; Norton, S. E.; Costanzo, S.P. In *Encyclopedia of Rapid Microbiological Methods*, Volume 2. Edited by Michael J. Miller. PDA and Davis Healthcare International Publishing.

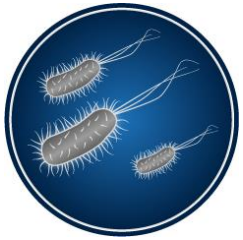




# Accuracy Examples

- 2005. ATP Bioluminescence Using Millipore's MilliFlex® Rapid System. Ohresser, S. In *Encyclopedia of Rapid Microbiological Methods*, Volume 2. Edited by Michael J. Miller. PDA and Davis Healthcare International Publishing.

| Dilution | 1:6 | 2:6 | 3:6 | 4:6 | 5:6 | 6:6 |
|---|---|---|---|---|---|---|
| Obtained CFUs **Standard** | 12, 8 | 23, 4 | 36 | 50, 4 | 66, 6 | 75, 4 |
| Obtained CFUs **Rapid** | 10, 2 | 20 | 35, 6 | 45, 8 | 56 | 63, 3 |
| % Recovery | **80** | **85** | **99** | **91** | **84** | **84** |

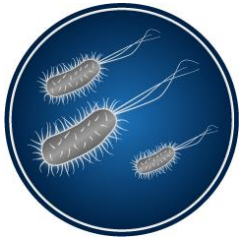


# Accuracy Examples

**RMM counts = 30, 50, 40, 33, 48 (mean = 40.2)**

**Plate count (CFU) = 50, 70, 38, 59, 52 (mean = 53.8)**

- The RMM recovers 74.7% of the plate count (> 70%)
- Perform a 2-sample Equivalency or Non-Inferiority test to confirm the methods are statistically the same



# Accuracy Examples
# 2-sample Equivalency Test

| Variable | N | Mean | StDev | SE Mean |
|---|---|---|---|---|
| RMM_4 | 5 | 40.200 | 8.8431 | 3.9547 |
| PC_4 | 5 | 53.800 | 11.798 | 5.2764 |

Pooled StDev = 10.42593

Difference: Mean(RMM_4) - Mean(PC_4)

| Difference | SE | 95% CI for Equivalence | Equivalence Interval |
|---|---|---|---|
| -13.600 | 6.5939 | (-25.862, 0) | (-16.14, 16.14) |

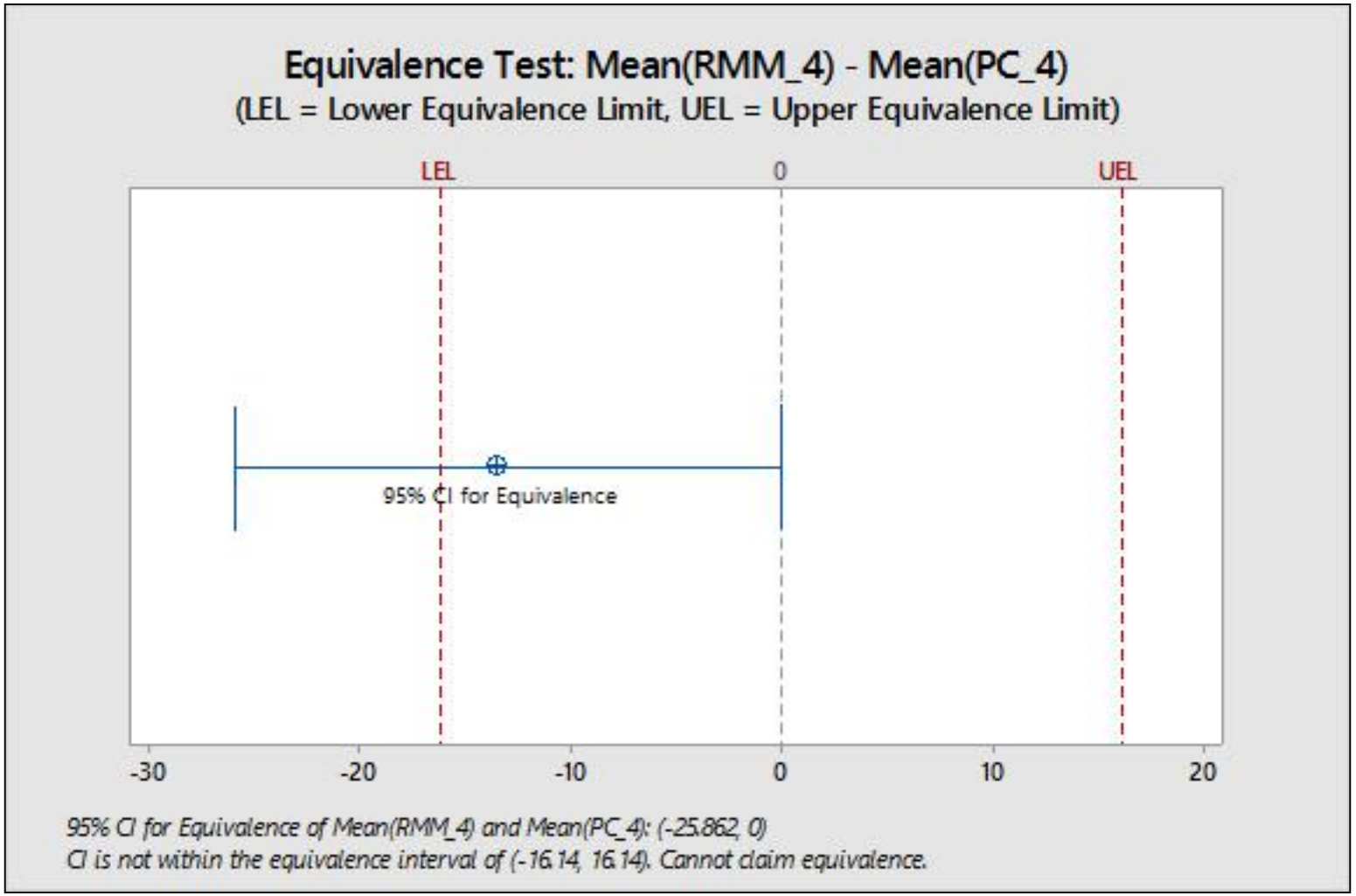
CI is not within the equivalence interval. Cannot claim equivalence.

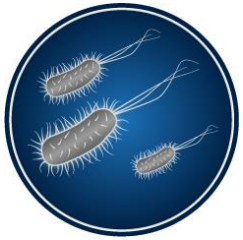
| | |
|---|---|
| Null hypothesis: | Difference ≤ -16.14 or Difference ≥ 16.14 |
| Alternative hypothesis: | -16.14 < Difference < 16.14 |
| α level: | 0.05 |

| Null Hypothesis | DF | T-Value | P-Value |
|---|---|---|---|
| Difference ≤ -16.14 | 8 | 0.38520 | 0.355 |
| Difference ≥ 16.14 | 8 | -4.5102 | 0.001 |

The greater of the two P-Values is 0.355. Cannot claim equivalence.

# Accuracy Example 3: Run a 2-sample Equivalency Test

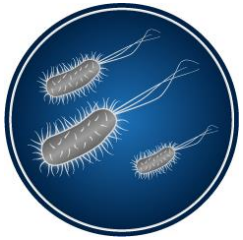




# Precision

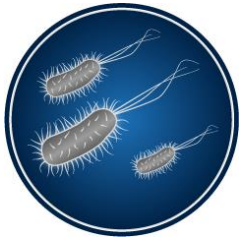
| Method | USP <1223> | Ph. Eur. 5.1.6 | PDA TR33 |
|---|---|---|---|
| Qualitative | | | |
| Quantitative | ✔ | ✔ | ✔ |

- USP also includes “Repeatability” to be performed for qualitative and quantitative methods
- This may simply be a subset of Precision
- However, USP does not provide guidance on testing repeatability for a qualitative method



# Precision – USP 1223

- The degree of agreement among individual test results when the procedure is applied repeatedly to multiple samplings of the same suspension of microorganisms and uses different suspensions across the range of the test (repeatability)
- *USP provides guidance on testing Precision only in the Equivalence section*
- Use at least six samples with no less than two bioburden levels near the specification limit relevant to the application (e.g., if a specification is NMT 100 CFU, use the same concentration for the test)

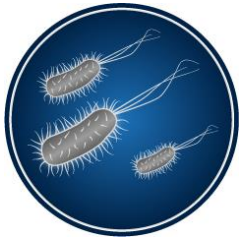


# Precision – USP 1223

- At each bioburden level, determine the sample variance ($S^2$) of the logarithms ($\log_{10}$) of the sample results. Then calculate the upper level (*UL*) precision using the following formula:

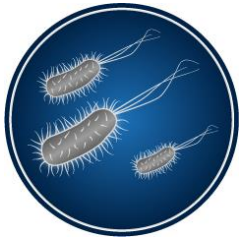
$$UL = 100 * \left[ anti\log\left( \sqrt{\frac{(n-1)S^2}{\chi^2_{.05,\, n-1}}} \right) - 1 \right]$$

  - $n$ = number of samples
  - $\chi^2_{.05,\, n\text{-}1}$ = lower 5% value of a Chi-square distribution with n-1 degrees of freedom



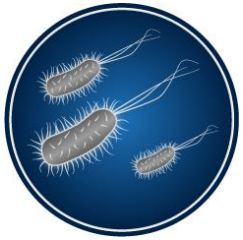
# Precision – USP 1223

- Precision is acceptable if *UL* ≥ **σ**, where **σ** is a predetermined maximal acceptable repeatability percent geometric coefficient of variation, %GCV
- However, USP does not specify what an acceptable %GCV is



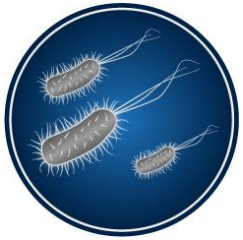
# Precision – Ph. Eur. 5.1.6

- The degree of agreement between individual test results when the method is applied to multiple homogeneous suspensions of microorganisms
- Precision should address repeatability and intermediate precision
- Repeatability (also known as within-run variability) is performed on the same sample (replicate) sample by a single analyst while using the same equipment in the same laboratory over a short period of time. This gives the minimum variability.



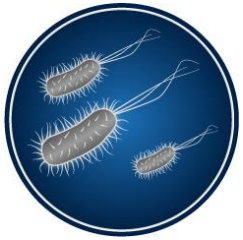
# Precision – Ph. Eur. 5.1.6

- Intermediate precision is performed using different sample preparations by different analysts, equipment, reagents and/or on different days but within the same laboratory. This gives the maximum variability.
- Precision is usually expressed as the standard deviation or relative standard deviation (also known as the coefficient of variation)



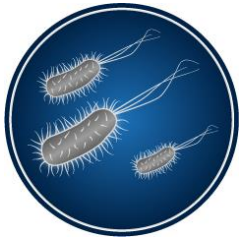
# Precision – Ph. Eur. 5.1.6

- At least 1 suspension in the middle of the test range is evaluated and the number of replicates is chosen such that the study can be completed in the same working session
- When assessing intermediate precision, other working sessions are carried out, using different reagents, operators and/or days, etc.
- The variance in the results observed in each of the working sessions is then calculated



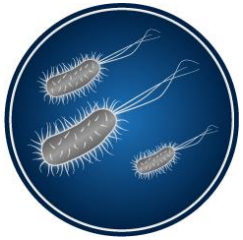
# Precision – Ph. Eur. 5.1.6

- If the variances are shown to be homogeneous, the variance of repeatability is determined
  - Note: use a software program to show homogeneity
- Next, the inter-group variance is calculated
- Intermediate precision is the sum of the variance of the repeatability and the inter-group variance
- Determine the coefficient of variation
- The alternative method must demonstrate precision comparable to that of the pharmacopoeial method



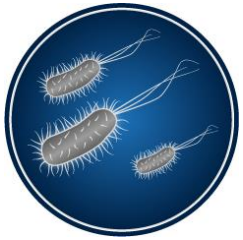
# Precision – PDA TR33

- The degree of agreement among individual test results when the procedure is applied repeatedly to multiple samplings of the same suspension of organisms and different suspensions across the range of the test
- Includes repeatability (within-run variability or intra-assay precision), intermediate precision and reproducibility (between laboratories; through collaborative studies)
- Intermediate precision and reproducibility is addressed under Ruggedness



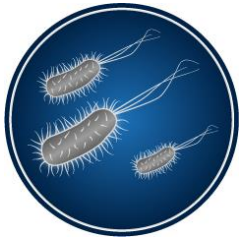
# Precision – PDA TR33

- Prepare a suspension of microorganisms in a suitable diluent at the upper end of the range of the test and serially dilute down to the lower end of the range
- Analyze at least two to five suspensions across the range of the test
- For each suspension, at least five to ten replicates should be assayed for recovered counts
- Perform for both the new and existing methods
- As counts approach the lower end of the range of the test (e.g., a single cell), variability in precision will increase; use appropriate suspensions and replicates



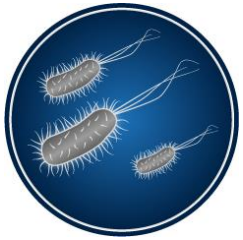
# Precision – PDA TR33

- The new method should not have a variability that is significantly larger than that of the existing method, except when a clear rationale or justification exists why such higher variability can be tolerated
- For a traditional plate count method, a coefficient of variation of less than 35% for a microbial number higher than 10 CFU is generally expected
- If the new method has a %CV ≤ 35, there is no need to compare the new method's CV with the existing method's CV



# Precision – PDA TR33

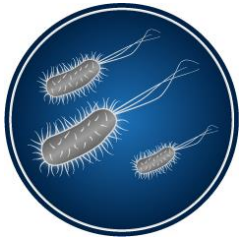
- However, if the new method has a %CV > 35, use a statistical test to demonstrate the new method is not significantly greater than the CV of the existing method
- Paired t-test
- Test for equal variance
  - F-test or Bartlett's test for normal distributed data
  - Levene's test for data not normally distributed
- McKay approximation (confidence intervals are compared)



# Precision Example

- 2005. Validation of the ScanRDI® for Purified Water Testing. McCormick, P.J.; Norton, S. E.; Costanzo, S.P. In *Encyclopedia of Rapid Microbiological Methods*, Volume 2. Edited by Michael J. Miller. PDA and Davis Healthcare International Publishing.

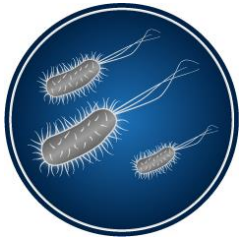
| Organism | Stock | | | Dilution | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | 1:10 | | | 1:100 | | |
| | Mean Cell Count | SD | %CV | Mean Cell Count | SD | %CV | Mean Cell Count | SD | %CV |
| *S. aureus* | 10,113.3 | 136.9 | 1.3 | 1,019.3 | 16.1 | 1.6 | 95 | 8.8 | 9.3 |
| *P. aeruginosa* | 8,700.5 | 131.3 | 1.5 | 525 | 132.5 | 13.6 | 19.5 | 8.5 | 44.1 |
| *B. subtilis* | 1,669 | 36.4 | 2.2 | 198 | 10.9 | 5.5 | 20.8 | 3.8 | 18.6 |
| *C. albicans* | 3,674.2 | 377.3 | 10.3 | 452.8 | 46.9 | 10.4 | 46.8 | 9.4 | 20.1 |
| *A. niger* | 2,892.7 | 112.5 | 3.9 | 249.3 | 31.2 | 12.5 | 14.7 | 2.6 | 18.6 |



# Precision Example

- 2005. ATP Bioluminescence Using Millipore's MilliFlex® Rapid System. Ohresser, S. In *Encyclopedia of Rapid Microbiological Methods*, Volume 2. Edited by Michael J. Miller. PDA and Davis Healthcare International Publishing.

| | | | | | |
|---|---|---|---|---|---|
| Milliflex Rapid System Counts (CFUs) | 61 | 26 | 112 | 118 | 68 |
| | 56 | 31 | 127 | 138 | 76 |
| | 63 | 30 | 117 | 143 | 70 |
| | 65 | 30 | 119 | 124 | 89 |
| | 55 | 31 | 119 | 145 | 71 |
| | 66 | 29 | 115 | 133 | 71 |
| | 59 | 28 | 118 | 112 | 79 |
| | 61 | 25 | 115 | 139 | 73 |
| | 45 | 30 | 117 | 144 | 83 |
| | 59 | 25 | 112 | 138 | ND |
| Mean | 59 | 28,5 | 117,1 | 133,4 | 75,6 |
| St Dev | 6,1 | 2,4 | 4,3 | 11,5 | 6,9 |
| **Rel St Dev (%)** | **10,3** | **8,3** | **3,7** | **8,6** | **9,2** |



# Precision Example

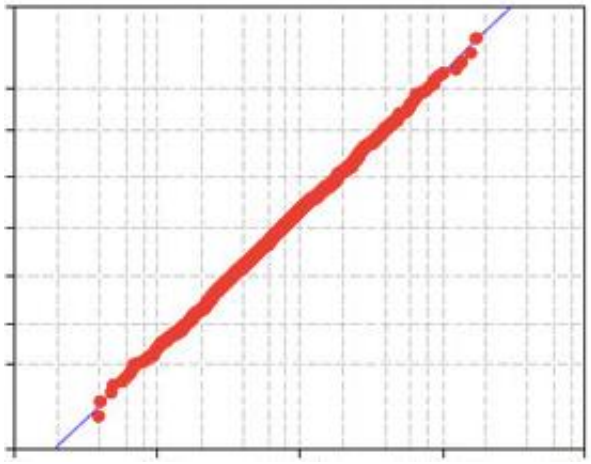
- RMM Data: 25, 35, 69, 21, 50, 80, 48, 61, 40, 88
- PC Data: 27, 38, 70, 27, 55, 76, 54, 74, 33, 80

| Variable | N | Mean | SE Mean | StDev | Variance | CoefVar | Minimum | Median | Maximum |
|---|---|---|---|---|---|---|---|---|---|
| RMM Precision | 10 | 51.70 | 7.14 | 22.59 | 510.23 | 43.69 | 21.00 | 49.00 | 88.00 |
| PC Precision | 10 | 53.40 | 6.64 | 21.00 | 440.93 | 39.32 | 27.00 | 54.50 | 80.00 |

- RMM %CV = 43.69; PC %CV = 39.32
- Is the RMM %CV statistically greater than the PC %CV?
- First, determine if the data is normally distributed
- Then, use an appropriate test for equal variances

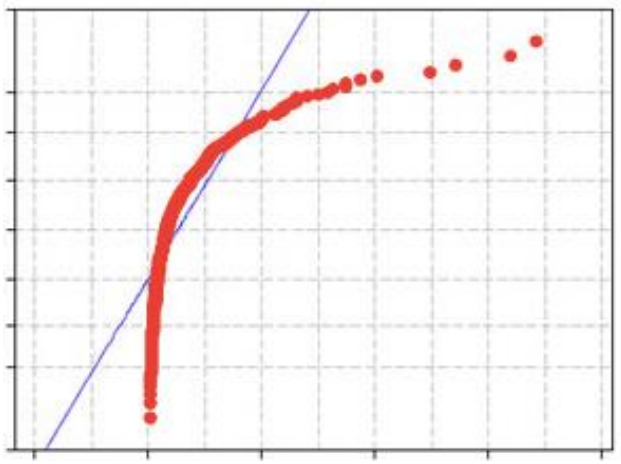
# Precision Example
# Test for Normal Distribution

- **$H_0$**: data follow a normal distribution ($p > 0.05$)
- **$H_1$**: data do not follow a normal distribution ($p \leq 0.05$)



$p = 0.665$

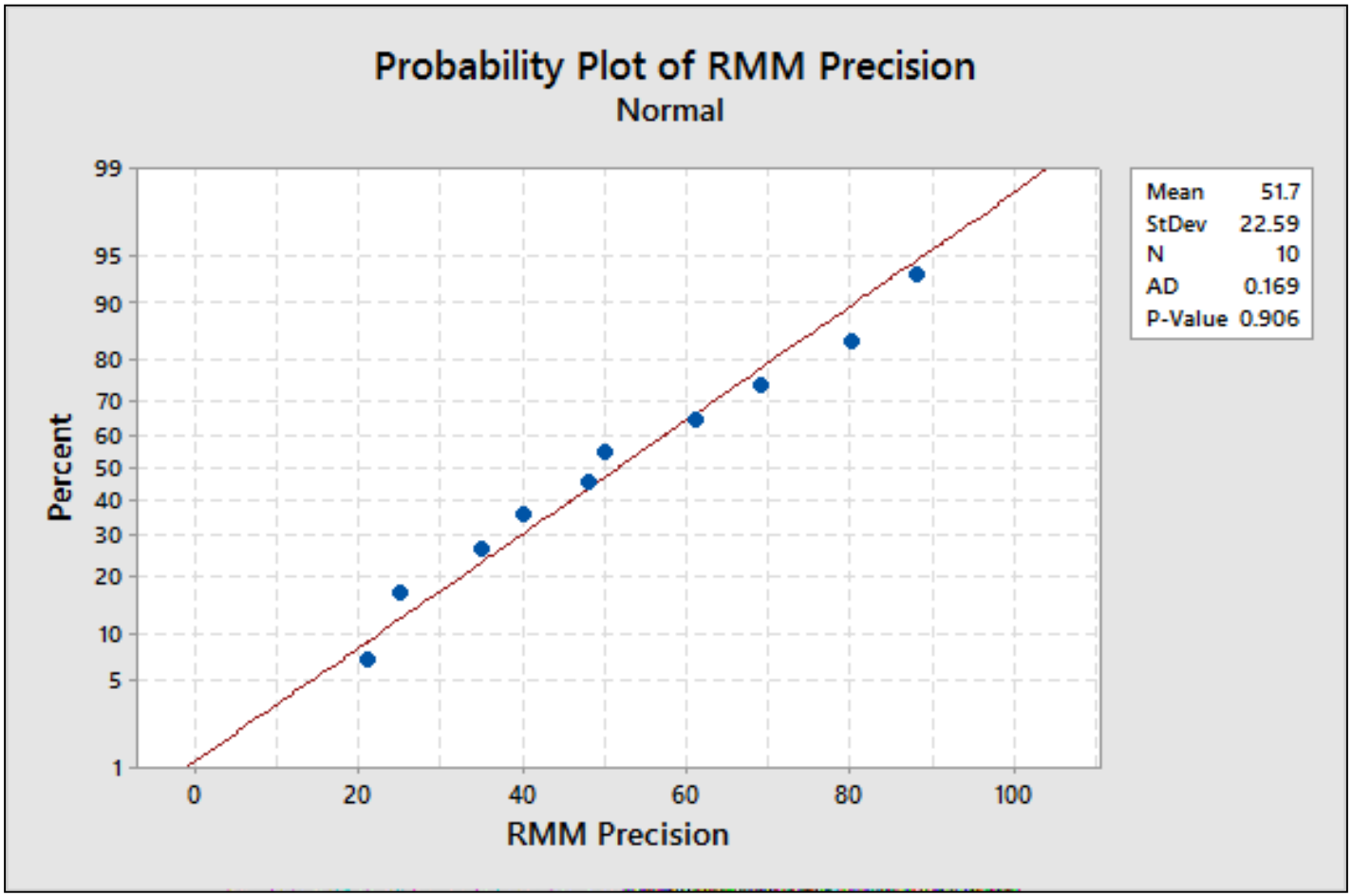
Normal Distribution Example



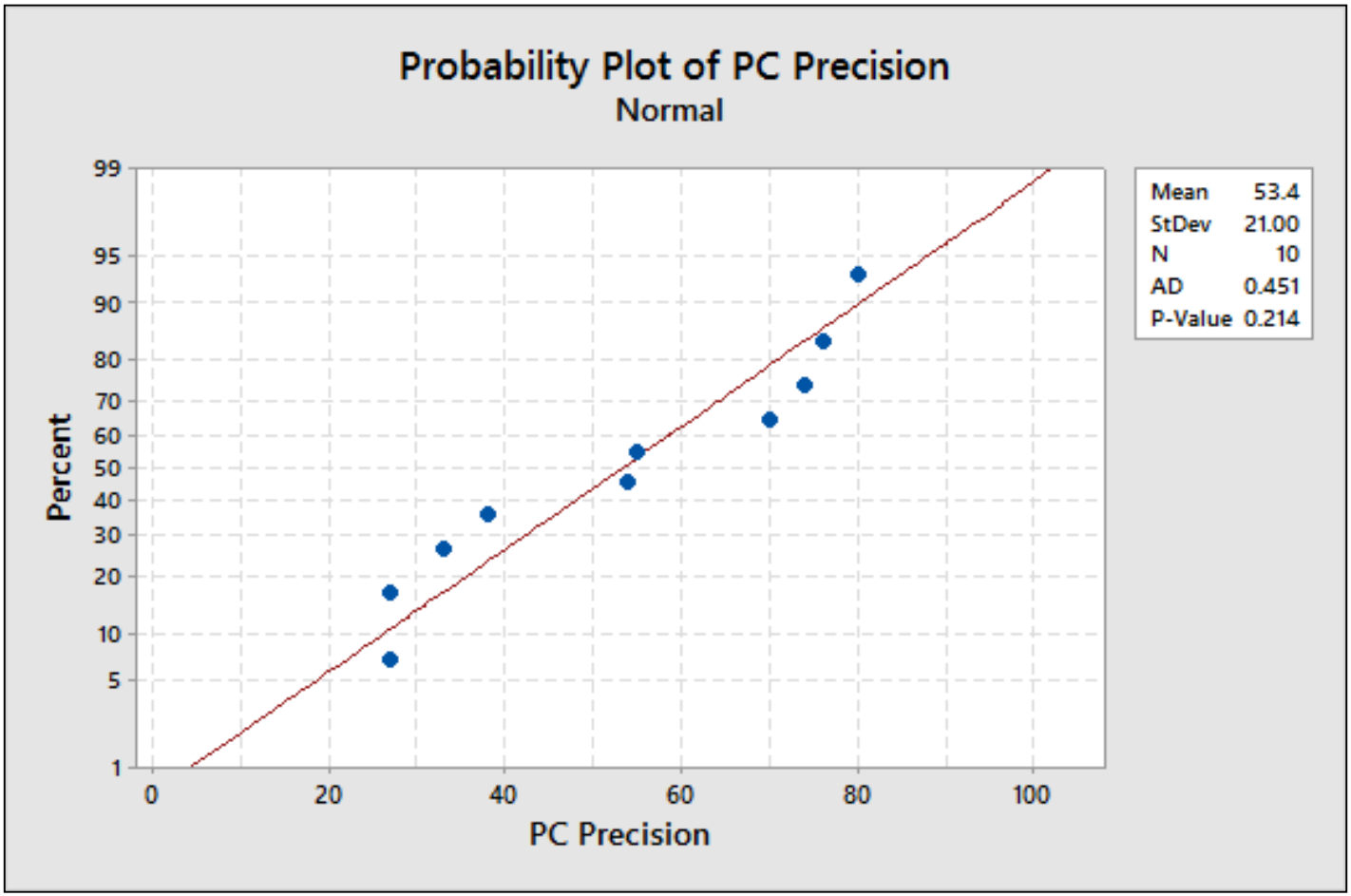
$p < 0.005$

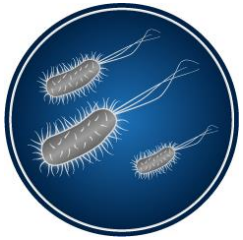
Non-Normal Distribution Example

# Precision Example
# Test for Normal Distribution



# Precision Example
# Test for Normal Distribution

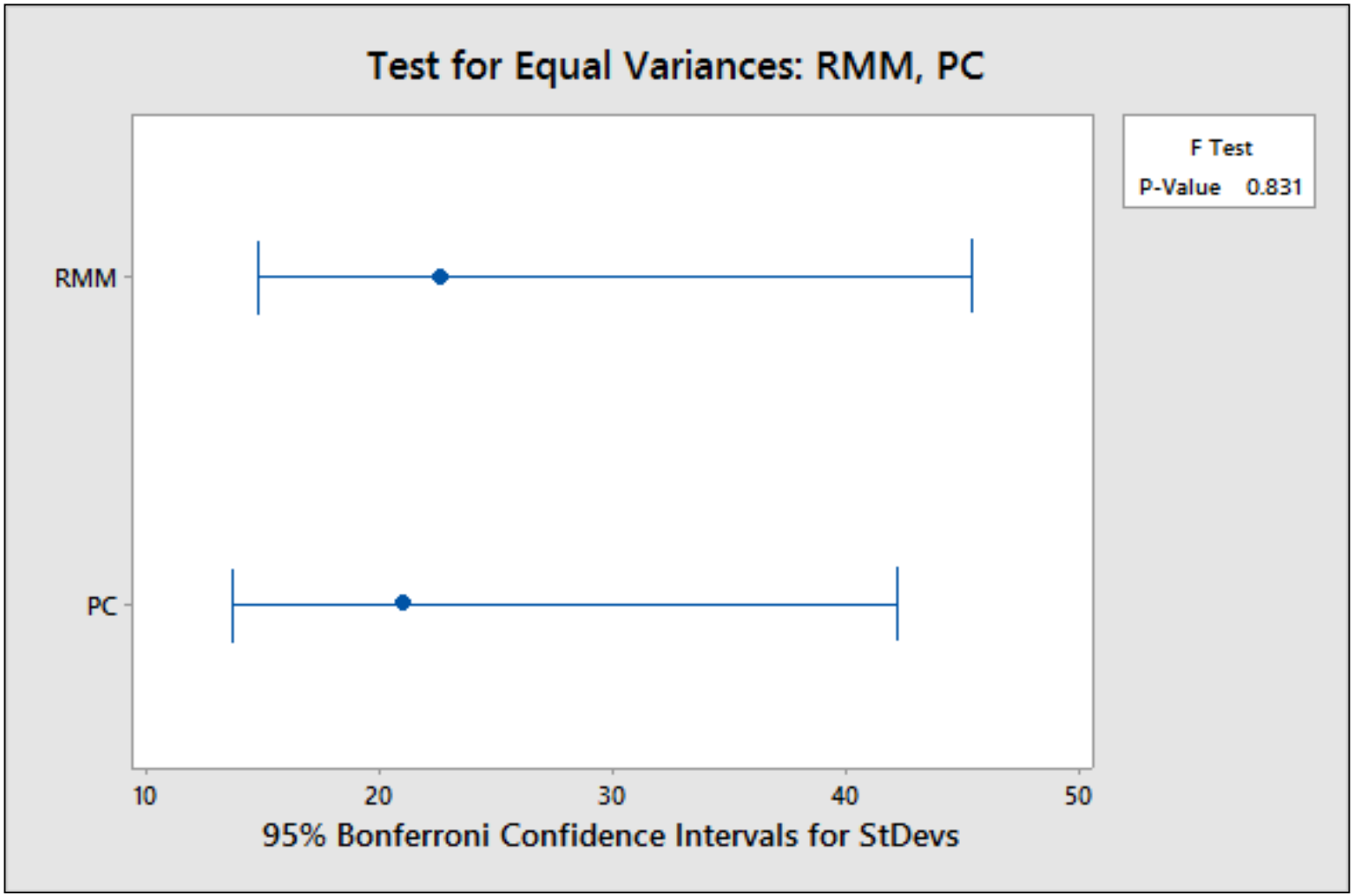


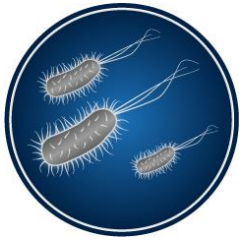


# Precision Example
# Test for Equal Variances (F-test)

```
Null hypothesis         All variances are equal
Alternative hypothesis  At least one variance is different
Significance level      α = 0.05

F method is used. This method is accurate for normal data only.


95% Bonferroni Confidence Intervals for Standard Deviations

Sample   N    StDev          CI
   RMM  10  22.5883  (14.7756, 45.4851)
    PC  10  20.9984  (13.7356, 42.2835)

Individual confidence level = 97.5%


Tests

            Test
Method  Statistic  P-Value
F            1.16    0.831
```

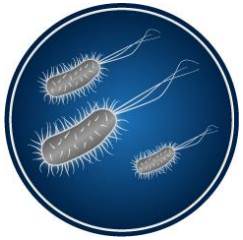
# Precision Example
# Test for Equal Variances (F-test)





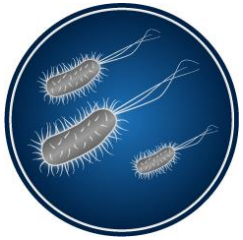
# Specificity

| Method | USP <1223> | Ph. Eur. 5.1.6 | PDA TR33 |
|---|---|---|---|
| Qualitative | ✔ | ✔ | ✔ |
| Quantitative | ✔ | ✔ | ✔ |



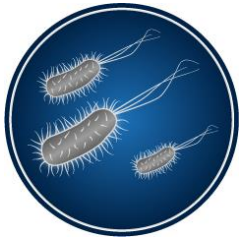
# Specificity – USP 1223

- Ability to detect a range of challenge microorganisms specific to the technology
- Microorganisms representing risk to patient or product, found in the manufacturing environment and product failures, that are appropriate for measuring the effectiveness of the alternative method
- Demonstrated by comparable recovery in both the compendial and alternate methods
- The microbial challenge is above the LOD or LOQ but at a level that provides a measure of efficacy of the methods



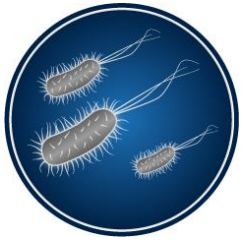
# Specificity – USP 1223

- Growth based methods: show recovery of low numbers (~ 100 CFU) of organisms in the alternative and compendial methods
- Non-growth based methods: use suitable negative and positive controls to demonstrate that extraneous matter (e.g., extracellular ATP, DNA, or inhibition and enhancement factors) does not interfere with the detection of challenge organisms
  - *NOTE: this may be difficult in the absence of suitable controls*
  - *May need to address this under method suitability using challenge organisms*



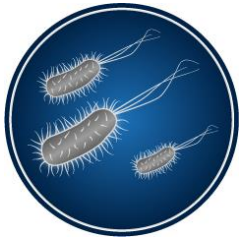
# Specificity – Ph. Eur. 5.1.6

- Ability to detect only the required microorganisms (i.e., does not generate false positive results)
- For a quantitative method, the ability to *quantify* only the required micro-organisms
- Demonstrated using a panel of appropriate microorganisms
- Where relevant for the purpose of the test, mixtures of organisms are used during validation



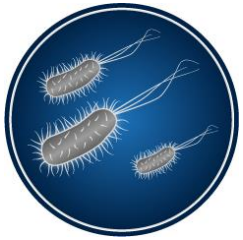
# Specificity – Ph. Eur. 5.1.6

- For growth based qualitative methods: demonstrate growth promotion properties of the media
- For non-growth based qualitative and quantitative methods: extraneous matter in the test system does not interfere with the test
- No guidance on how to perform extraneous matter testing
- The use of compendial test strains, in-house (environmental) isolates and stressed or slow-growing organisms is recommended



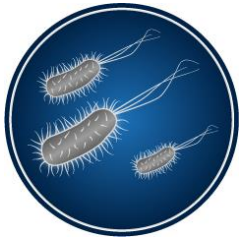
# Specificity – PDA TR33

- Ability to detect a range of microorganisms, which demonstrate that the method is fit for its intended use
- For qualitative methods that detect a target panel of specific microorganisms, such as those that employ nucleic acid amplification techniques (e.g., as described in Ph. Eur. 2.6.21 and USP <1125>), inclusivity and exclusivity should be demonstrated



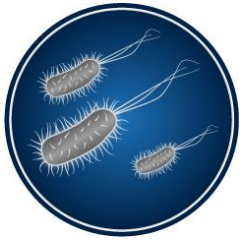
# Specificity – PDA TR33

- Inclusivity: target microorganisms that should be detected will provide a positive result
- Exclusivity: non-target, unrelated and closely related microorganisms that should not be detected will provide a negative result
- Select an appropriate number of replicates (e.g., at least 3) of target and non-target microorganisms



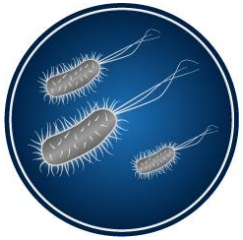
# Specificity – PDA TR33

- Use a representative panel of organisms
  - e.g., Gram-positive and Gram-negative bacteria, yeast, mold and/or bacterial and fungal spores
- The end-user should determine what types of organisms to use during the assessment of each of the validation criteria (e.g., Accuracy, Precision, LOD, LOQ)
- Sourced from culture collections (e.g., ATCC), environmental or facility isolates, in-process or sterility failure isolates, slow-growing, fastidious or anaerobic strains, and/or clinically relevant cultures



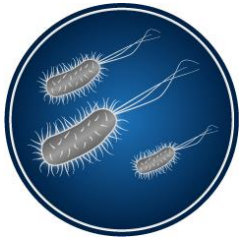
# Specificity – PDA TR33

- Consider the concentration of microorganisms used
  - Lower concentrations for inclusivity (e.g., at the LOD) and higher concentrations for exclusivity testing
- When applicable, relevant controls may need to be evaluated, such as nucleic acid standards
- However, some controls may not be sufficient to evaluate inclusivity in all cases, as they may not cover several aspects of method sample preparation, such as microorganism lysis, nucleic acid capture and purification



# Specificity – PDA TR33

- When appropriate, use mixed cultures or stressed organisms
- Mixed cultures: when the method should show the detection of more than one type of organism
- Stressed organisms: when the method should show detection of stressed cells, such as in a sterility test
  - Exposing cultures to environmental extremes, (e.g., UV, heat, cold, pH, tonicity), antimicrobials (disinfectants, preserved product) or sub-lethal sterilization conditions
  - The stress method should provide a reliable and reproducible challenge and may need to be qualified before use



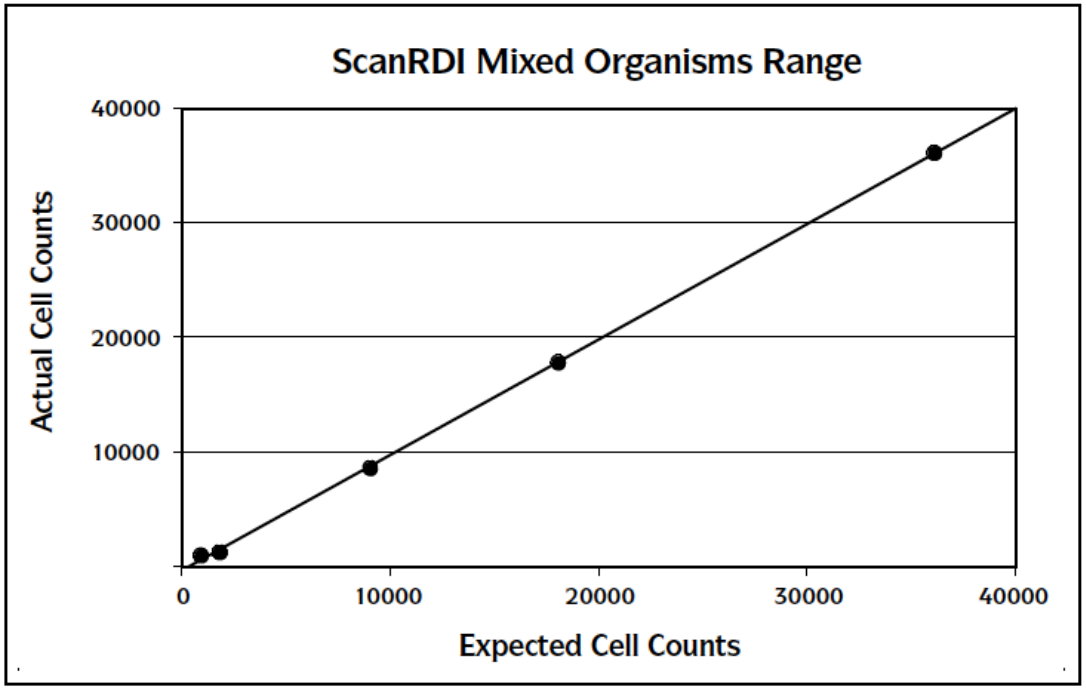
# Specificity – PDA TR33

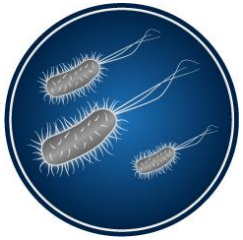
- All microorganisms should be successfully detected and/or enumerated, and meet the specific acceptance criteria of the validation test (e.g., Accuracy, Precision, etc.)
- Inclusivity testing: the new method detects the target microorganism(s) it is intended to
- Exclusivity: the new method does not produce a positive detection result for unrelated or closely related microorganisms

# Specificity Example

- 2005. Validation of the ScanRDI® for Purified Water Testing. McCormick, P.J.; Norton, S. E.; Costanzo, S.P. In *Encyclopedia of Rapid Microbiological Methods*, Volume 2. Edited by Michael J. Miller. PDA and Davis Healthcare International Publishing.

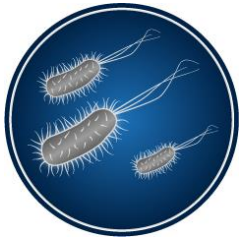
***P. aeruginosa*, *S. aureus* and *S.marscescens* (slope = 1.01, $r^2$ = 0.99).**





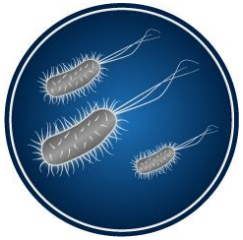
# Specificity Example

- Novartis case study
- 2010. Gray, J.C.; Staerk, A.; Berchtold, M.; Hecker, W.; Neuhaus, G.; Wirth, A. Growth-promoting Properties of Different Solid Nutrient Media Evaluated with Stressed and Unstressed Micro-organisms: Pre-study for the Validation of a Rapid Sterility Test. *PDA Journal of Pharmaceutical Science and Technology*. 64(3): 249-263.
- Used to select the optimal medium for a rapid sterility test



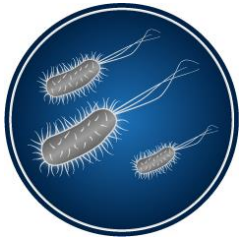
# Specificity Example

- Novartis used 22 microorganisms consisting of 7 ATCC strains and 15 production site-specific isolates from environmental monitoring samples, bioburden, and sterility test failures
- Organisms were also subjected a variety of microorganisms to environmental and chemical environments to demonstrate that stressed organisms will be detected in a rapid sterility test
- < 100 CFU; 5 reps per incubation parameter



# Specificity Example

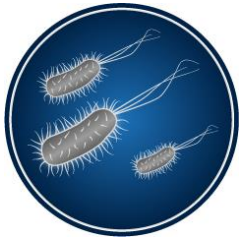
- Stressing conditions
  - UV light (240 - 250 W/cm$^2$)
  - Heat (50 - 70° C in a water bath)
  - Incubating the microorganisms in a dilution series of a parenteral drug product with microbicidal properties (Voltaren, Novartis) for 1–10 min each, taking aliquots every minute
  - Heating showed at least a 50% reduction in viable cell counts but also a reduced growth rate in the surviving organisms via optical density experiments over a subsequent 8-hour growth study
- Therefore, heat stressed and unstressed organisms were used to choose the most optimal medium and incubation parameters: γ-irradiated Schaedler Blood Agar for 5 days



# Limit of Detection (LOD)

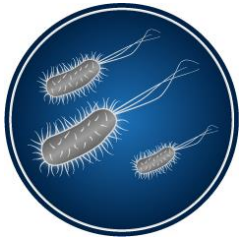
| Method | USP <1223> | Ph. Eur. 5.1.6 | PDA TR33 |
|---|---|---|---|
| **Qualitative** | ✔ | ✔ | ✔ |
| **Quantitative** | ✔ | * | ✔ |

***** May be needed in some cases.



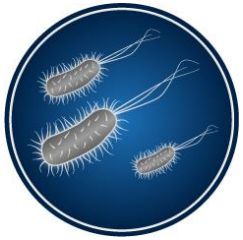
# LOD – USP 1223

- The lowest number of microorganisms in a defined volume of sample that can be detected, but not necessarily quantified, under the stated experimental conditions
- Performed with the QC organisms listed in USP 51, 61, 62, 63 and/or 71 as appropriate to the alternative method



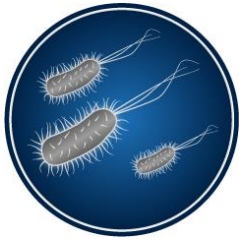
# LOD – USP 1223

- Dilute microorganisms in a suitable diluent to a concentration where the compendial test will show growth in 50% of the test samples
- Test the dilutions in the compendial and alternative methods using a sufficient number of replicates using a 0.05 alpha risk and a 0.20 beta risk
  - Alpha risk is your confidence level (risk of accepting the alternative when the null is true; *same as our prior discussions*)
  - Beta risk is the risk of failing to reject the null when the alternative is true
- Use a Chi-square or other appropriate statistical test to demonstrate equivalent recovery of microorganisms



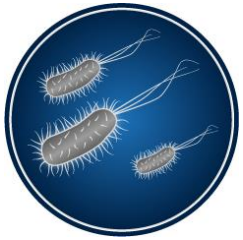
# LOD – USP 1223

- Can also use a Most Probably Number (MPN) test
- A series of 10-fold dilutions (e.g., from $10^{1}$ CFU to $10^{-2}$ CFU) or 5-fold dilutions (e.g., from 5 CFU to $10^{-1}$ CFU) are challenged in both methods
- Five replicates from each dilution are assayed and the MPN is determined from three dilutions in series that provide both positive and negative results (i.e., growth in the compendial method and an appropriate signal in the alternative method)
- Use a Chi-square or other appropriate statistical test to demonstrate equivalent recovery of microorganisms



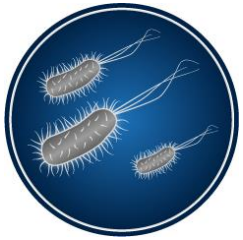
# LOD – Ph. Eur. 5.1.6

- The lowest number of microorganisms in a sample that can be detected under the stated analytical conditions
- The detection limit reflects the number of organisms in the original sample before any dilution or incubation steps
- The detection limit of the alternative method is a number not greater than that of the compendial method
- A sufficient number of replicates and independent determinations are recommended; however, the chapter does not provide specific guidance
- Can use Accuracy in place of LOD; *see prior discussion*



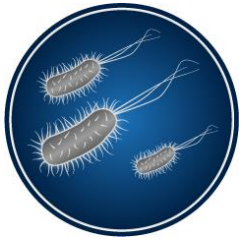
# LOD – PDA TR33

- The lowest concentration of microorganisms in a test sample that can be detected, but not necessarily quantified, under the stated experimental conditions
- LOD refers to the number of organisms present in the original sample, before any incubation step, not the number of organisms present at the point of assay
- The amount of sample tested, and the dilution of that sample, may determine the limit of detection



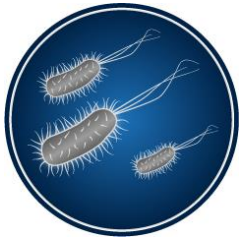
# LOD – PDA TR33

- As it is not possible to consistently obtain a reliable sample containing a very low level of microorganisms (e.g., a single viable cell), it is essential that the LOD is determined from an appropriate number of replicates
- The number of replicates may depend on the statistical method(s) used



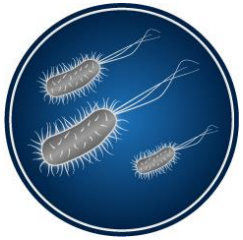
# LOD – PDA TR33

- Challenge the new method with an inoculum that represents the method's purported LOD
- For example, if the method can detect < 5 CFU, challenge with 1-5 CFU
- Alternatively, adjust the inoculation level until at least 50% of the samples show growth in the existing method
- Or, make dilutions into the fractional range (50, 5, 0.5 and 0.05 CFU) and use a MPN technique



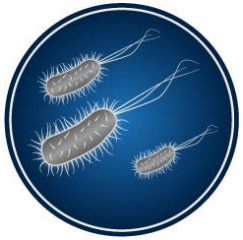
# LOD – PDA TR33

- Compare the rate of recovery (number of positive to negative results) using a suitable statistical test
  - Equivalence, Fisher's exact test, Chi-Square
- The LOD of the new method should not be significantly worse than that of the existing method, except when a clear rationale or justification exists why a higher detection limit can be tolerated



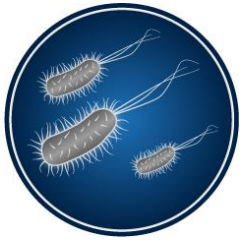
# PDA - A Note About Statistical Power

- Use Chi-Square for large sample sizes; Fisher's exact test for small sample sizes
- When using less than 5 CFU, the statistical power of these evaluations may be reduced
- If no test power calculations are performed, not less than 50 replicates per test should be included in the statistical evaluation (derived from a test power simulation with the Fisher's exact test, assuming a mean inoculum of 2 CFU)
- You can pool the results obtained for different test runs and microorganisms in order to achieve adequate test power
  - e.g., 6 organisms × 10 reps × 3 runs = 180 data points



# LOD Example: 1-5 CFU

- Perform a Non-Inferiority Test
  - The number of positive results in the RMM
  - The number of positive results in the compendial method
- Lower equivalency level (Δ) = - 0.2
  - alpha = 0.05 (95% confidence level)
- If the p-value is ≤ 0.05, we reject the null hypothesis and claim non-inferiority for the RMM



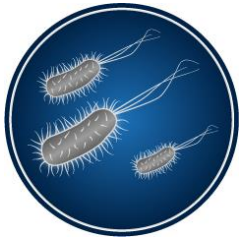
# LOD Example: 1-5 CFU

- Assume data for 6 organisms, 10 reps and 3 runs:

| Method | # Positives | # Negatives | Total |
|---|---|---|---|
| **RMM** | 171 | 9 | 180 |
| **Existing** | 162 | 18 | 180 |

```
Sample    X    N  Sample p
1       171  180  0.950000
2       162  180  0.900000

Difference = p (1) - p (2)
Estimate for difference:  0.05
95% lower bound for difference:  0.00453865
Test for difference = -0.2 (vs > -0.2):  Z = 9.05  P-Value = 0.000
```

- p-value ≤ 0.05; reject null; proved RMM is non-inferior



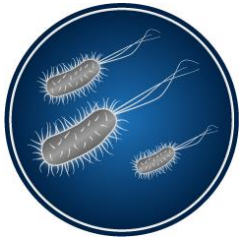
# LOD Example: 1-5 CFU

- Assume data for 6 organisms, 10 reps and 3 runs:

| Method | # Positives | # Negatives | Total |
|---|---|---|---|
| **RMM** | 145 | 35 | 180 |
| **Existing** | 172 | 8 | 180 |

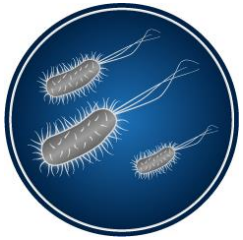
```
Sample    X    N  Sample p
1       145  180  0.805556
2       172  180  0.955556

Difference = p (1) - p (2)
Estimate for difference:  -0.15
95% lower bound for difference:  -0.204706
Test for difference = -0.2 (vs > -0.2):  Z = 1.50  P-Value = 0.066
```

- p-value > 0.05; fail to reject null; did not prove the RMM is non-inferior



# LOD Example: 1-5 CFU

- Can also use traditional hypothesis tests such as the Chi-Square and Fisher's exact test
- However, we lose the ability to designate an acceptable equivalency or non-inferiority margin
- We also require a sufficient number of samples, especially for the Chi-Square test
- Let's apply these tests to the previous data…
  - $H_0$: The two methods are statistically the same
  - $H_1$: The two methods are statistically different

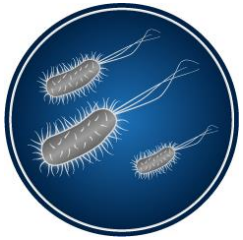


# LOD Example: 1-5 CFU

| Method | # Positives | # Negatives | Total |
|---|---|---|---|
| **RMM** | 171 | 9 | 180 |
| **Existing** | 162 | 18 | 180 |

```
          RMM LOD   USP LOD   All

1           171       162     333
2             9        18      27
All         180       180     360

Pearson Chi-Square = 3.243, DF = 1, P-Value = 0.072
Likelihood Ratio Chi-Square = 3.301, DF = 1, P-Value = 0.069

Fisher's exact test: P-Value =  0.107966
```

- p-value > 0.05; fail to reject null; did not prove the RMM is different

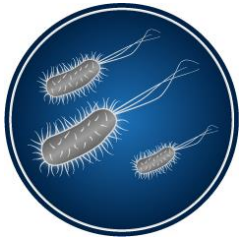


# LOD Example: 1-5 CFU

| Method | # Positives | # Negatives | Total |
|---|---|---|---|
| **RMM** | 145 | 35 | 180 |
| **Existing** | 172 | 8 | 180 |

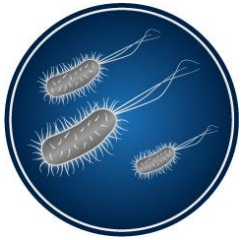
```
     RMM LOD_1  USP LOD_1   All

1        145       172      317
2         35         8       43
All      180       180      360

Pearson Chi-Square = 19.253, DF = 1, P-Value = 0.000
Likelihood Ratio Chi-Square = 20.595, DF = 1, P-Value = 0.000

Fisher's exact test: P-Value =  0.0000136
```

- p-value ≤ 0.05; reject null; proved the RMM is different



# LOD Example: Most Probable Number (MPN)

- Prepare 10-fold, serial dilutions to provide a final inoculum concentration equal to 50 CFU, 5 CFU, 0.5 CFU and 0.05 CFU
- Ten (10) replicates per dilution per organism; 3 runs
  - e.g., 3 organisms × 10 reps × 3 runs = 90 data points per dilution
- Calculate MPN values and upper and lower confidence intervals using FDA MPN Table 3 for 10 tubes
- http://www.fda.gov/food/foodscienceresearch/laboratorymethods/ucm109656.htm

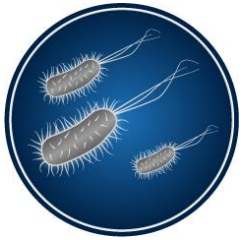


# LOD Example: MPN

**Table 3**
**For 10 tubes at each of 0.1, 0.01, and 0.001 g inocula, the MPNs and 95 percent confidence intervals.**

| Pos. tubes | | | | Conf. lim. | | Pos. tubes | | | | Conf. lim. | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.1 | 0.01 | 0.001 | MPN/g | Low | High | 0.1 | 0.01 | 0.001 | MPN/g | Low | High |
| 0 | 0 | 0 | <.90 | – | 3.1 | 8 | 2 | 0 | 17 | 7.7 | 34 |
| 0 | 0 | 1 | 0.9 | 0.04 | 3.1 | 8 | 2 | 1 | 19 | 9 | 34 |
| 0 | 0 | 2 | 1.8 | 0.33 | 5.1 | 8 | 2 | 2 | 21 | 10 | 39 |
| 0 | 1 | 0 | 0.9 | 0.04 | 3.6 | 8 | 2 | 3 | 23 | 11 | 44 |
| 0 | 1 | 1 | 1.8 | 0.33 | 5.1 | 8 | 3 | 0 | 19 | 9 | 34 |

- FDA only uses 3 dilutions to calculate the MPN
- Therefore, we must choose the best 3 dilutions from the 4 dilutions we made for the test



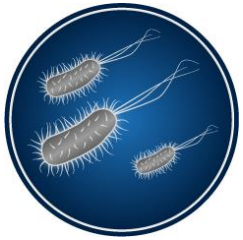
# LOD Example: MPN

- Assume the following results:

| Dilution (CFU) | 50 | 5 | 0.5 | 0.05 |
|---|---|---|---|---|
| **Result (# positive/10)** | 10 | 8 | 5 | 1 |

- Find the highest dilution with all positive tubes: 50
- If this dilution is not within the remaining 3 highest dilutions, select the next two higher dilutions: 5 and 0.5
- Include the next highest dilution with positive tubes: 0.05

| Dilution (CFU) | 50 | 5 | 0.5 | 0.05 | Values To Use |
|---|---|---|---|---|---|
| **Result (# positive/10)** | 10 | 8 | 5 | 1 | 8-5-1 |

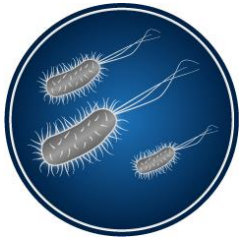


# LOD Example: MPN

- Next, determine the MPN value and its upper and lower confidence level from Table 3

| Pos. tubes | | | | Conf. lim. | |
|---|---|---|---|---|---|
| 0.1 | 0.01 | 0.001 | MPN/g | Low | High |
| 8 | 5 | 1 | 27 | 12 | 50 |

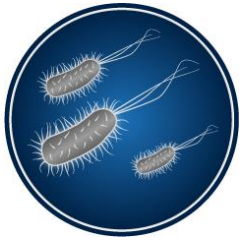
- NOTE: the FDA tables apply to 0.1, 0.01, and 0.001 dilutions
- When different dilutions are used, multiply (or divide) the MPN and confidence limits to make the sample dilutions match the table dilutions



# LOD Example: MPN

| Pos. tubes | | | | Conf. lim. | |
|---|---|---|---|---|---|
| 0.1 | 0.01 | 0.001 | MPN/g | Low | High |
| 8 | 5 | 1 | 27 | 12 | 50 |

- Since we used 5, 0.5 and 0.05 dilutions, divide the MPN and confidence levels by a factor of 10
- The final MPN becomes 2.7 and the final confidence levels become 1.2 and 5.0

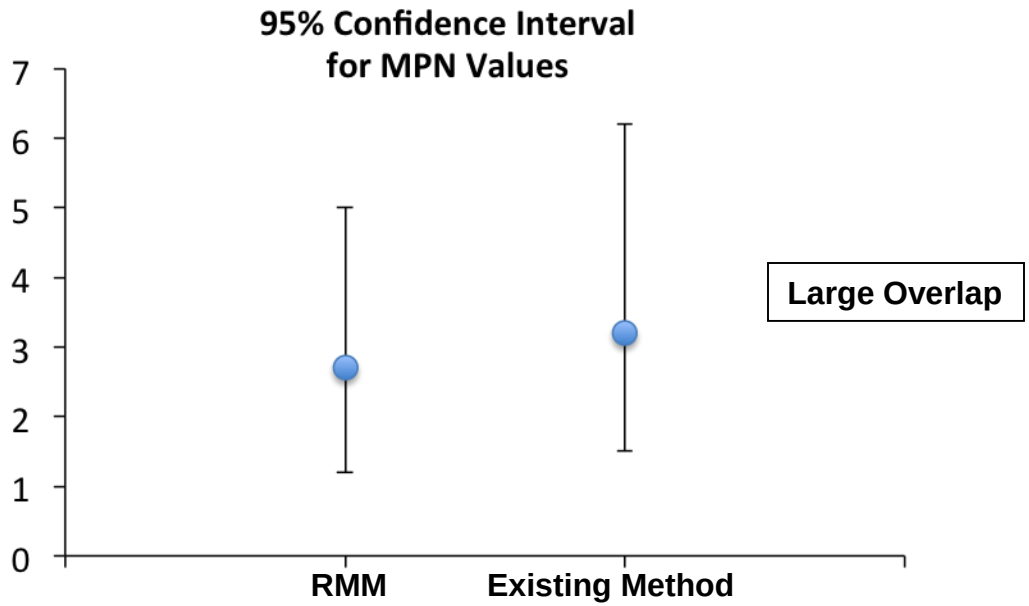


# LOD Example: MPN

- Demonstrate the confidence intervals for the MPN values for both methods overlap
  - If there is no overlap, the methods are not equivalent
  - If the overlap is large, most likely the methods are equivalent; however, it is wise to perform a statistical analysis to confirm this
  - If the overlap is small, always perform a statistical analysis
- 2-sample Non-Inferiority test on MPN values
- t-test on MPN values

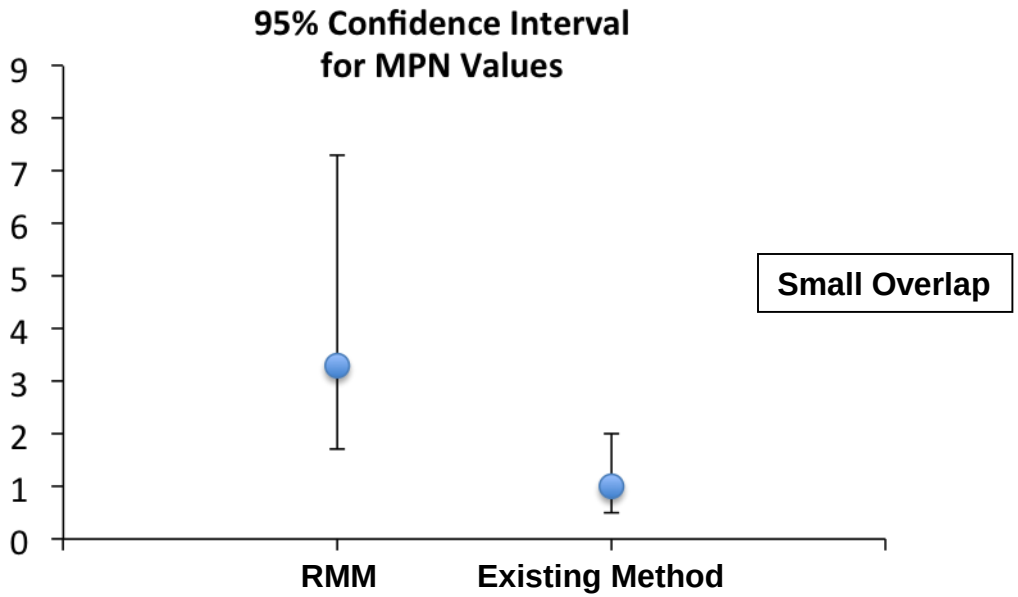
# LOD Example: MPN

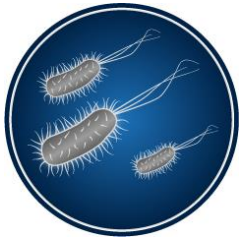
| Method | MPN Value | 95% Confidence Levels | |
|---|---|---|---|
| | | Low | High |
| **RMM** | 2.7 | 1.2 | 5.0 |
| **Existing Method** | 3.2 | 1.5 | 6.2 |



# LOD Example: MPN

| Method | MPN Value | 95% Confidence Levels | |
|---|---|---|---|
| | | Low | High |
| **RMM** | 3.3 | 1.7 | 7.3 |
| **Existing Method** | 1.0 | 0.5 | 2.0 |





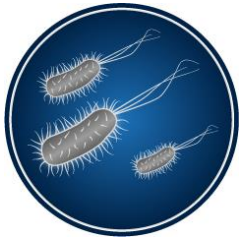
# LOD Example: MPN

- Assume the following MPN data representing 3 different microorganisms and 2 independent runs per organism:

**RMM MPN's = 4.5, 4.6, 3.7, 4.2, 3.3, 4.3 (mean = 4.1)**

**Existing method MPN's = 3.3, 4.5, 2.9, 4.5, 2.5, 2.9 (mean = 3.4)**

- If the 95% confidence interval for the difference is > - 0.2 and $p \leq 0.05$ for the left tail, we can claim non-inferiority for the RMM



# LOD Example: MPN
# 2-sample Non-Inferiority Test

| Variable | N | Mean | StDev | SE Mean |
|---|---|---|---|---|
| RMM MPN | 6 | 4.1000 | 0.50200 | 0.20494 |
| USP MPN | 6 | 3.4333 | 0.86410 | 0.35277 |

Pooled StDev = 0.70664

Difference: Mean(RMM MPN) - Mean(USP MPN)

| Difference | SE | 95% Lower Bound | Lower Limit |
|---|---|---|---|
| 0.66667 | 0.40798 | -0.072774 | -0.68667 |

Lower bound is greater than -0.68667. Can claim Mean(RMM MPN) - Mean(USP MPN) > -0.68667.

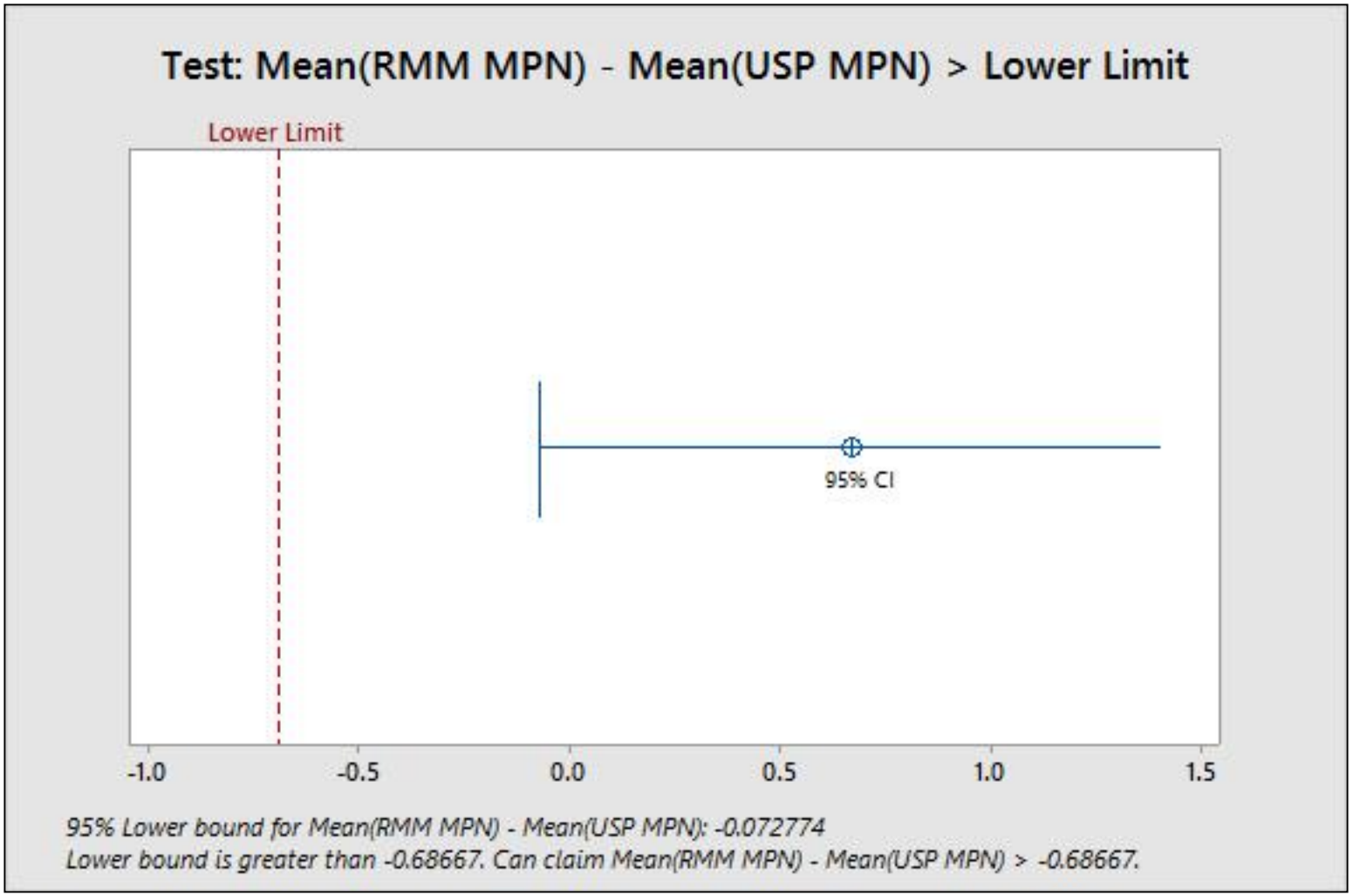
Null hypothesis: Mean(RMM MPN) - Mean(USP MPN) ≤ -0.68667
Alternative hypothesis: Mean(RMM MPN) - Mean(USP MPN) > -0.68667
α level: 0.05

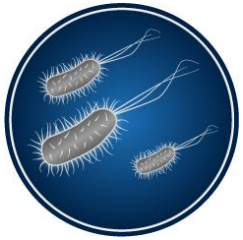
| DF | T-Value | P-Value |
|---|---|---|
| 10 | 3.3172 | 0.004 |

P-Value ≤ 0.05. Can claim Mean(RMM MPN) - Mean(USP MPN) > -0.68667.

- Proves the RMM LOD is statistically non-inferior

# LOD Example: MPN
# 2-sample Non-Inferiority Test

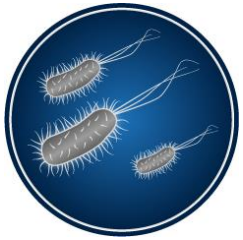




# LOD Example: MPN
# Traditional 2-sample t-test

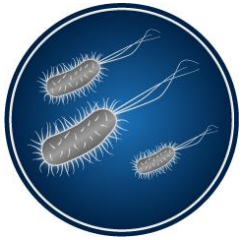
```
          N   Mean  StDev  SE Mean
RMM MPN   6  4.100  0.502     0.20
USP MPN   6  3.433  0.864     0.35

Difference = μ (RMM MPN) - μ (USP MPN)
Estimate for difference:  0.667
95% CI for difference:  (-0.242, 1.576)
T-Test of difference = 0 (vs ≠): T-Value = 1.63 P-Value = 0.133 DF = 10
Both use Pooled StDev = 0.7066
```

- p-value > 0.05; fail to reject null hypothesis
- The method LODs are not statistically different



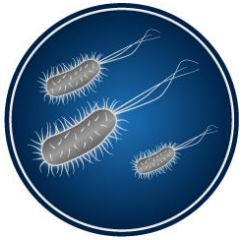
# Novartis Example of LOD Data for a Qualitative Test

- 2010. Gray, J.C.; Staerk, A.; Berchtold, M.; Mercier, M.; Neuhaus, G.; Wirth, A. Introduction of a Rapid Microbiological Method as an Alternative to the Pharmacopoeial Method for the Sterility Test. American Pharmaceutical Review. 13(6): 88-94.
- 22 stressed microbial strains
  - 7 ATCC strains and 15 isolates from a Novartis Pharma manufacturing site
  - Yeasts/molds, Gram positive sporulating bacteria, Gram negative rods, Gram positive cocci and Gram positive rods (both aerobic and anaerobic microorganisms)
- 1-5 CFU with ten replicates for each strain with both the traditional and rapid sterility tests. 2 independent test runs per organism.
- Statistical analysis was performed using the Chi-square and Fisher's exact tests and showed no significant statistical difference between the two test methods to detect viable microorganisms at a 95% confidence level (*note these were traditional hypothesis tests*)



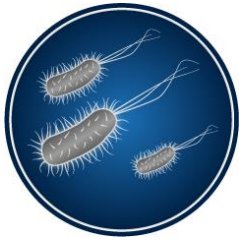
# Novartis Example of LOD Data for a Qualitative Test

- Their second approach used an MPN method
  - Stressed environmental isolates *Kocuria rhizophila, Acinetobacter lwoffii, Bacillus clausii, Penicillium spec.* and *Propionibacterium acnes* in ten replicates
  - Diluted suspension to extinction: 50 CFU, 5 CFU, 0.5 CFU and 0.05 CFU (10 replicates). 2 independent test runs per organism.
  - Used MPN tables to calculate the MPN value and 95% upper and lower confidence intervals
  - 95% confidence levels overlapped; performed a t-test on the MPN data
  - No significant statistical difference between the methods except for *P. acnes*, which was not always detected in the traditional sterility test
- Conclusions: The RMM was numerically superior and statistically non-inferior to the traditional sterility test method (*note these were traditional hypothesis tests)*



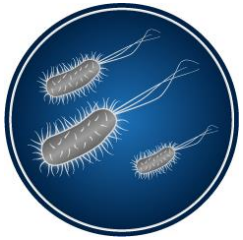
# Limit of Quantitation (LOQ)

| Method | USP <1223> | Ph. Eur. 5.1.6 | PDA TR33 |
|---|---|---|---|
| Qualitative | | | |
| Quantitative | ✔ | ✔ | ✔ |



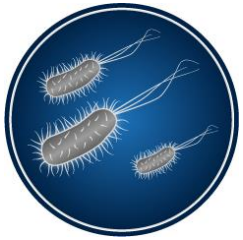
# LOQ – USP 1223

- The lowest number of microorganisms in a test sample that can be enumerated with acceptable accuracy and precision under defined experimental conditions
- USP does not provide specific guidance on how to conduct LOQ testing
- However, the results from Accuracy and Precision testing may confirm a method's LOQ as long as the concentration of organisms used are consistent with the purported LOQ by the method supplier



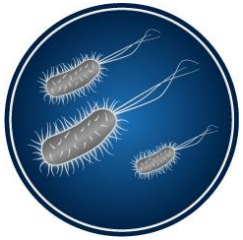
# LOQ – Ph. Eur. 5.1.6

- The lowest number of CFUs in a sample which can be quantitatively determined with suitable Precision and Accuracy
- The results of Accuracy and Linearity can also be used to demonstrate the quantitation limit; the lowest concentration in the linear range is considered the quantitation limit
- Use a suitable number of replicates
- The quantitation limit of the alternative method must not be greater than the quantitation limit of the compendial method



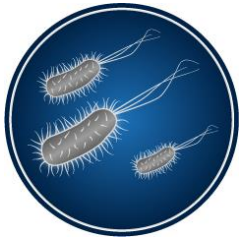
# LOQ – PDA TR33

- The lowest number of microorganisms in a test sample that can be enumerated with acceptable Accuracy and Precision under the stated experimental conditions
- As it is not possible to consistently obtain a reliable sample containing a very low level of microorganisms (e.g., a single viable cell), use an appropriate number of replicates (at least five to ten)
- Use different concentrations of organisms in a suitable diluent; including at or near the desired LOQ
- Choose the appropriate types of organisms (Specificity)



# LOQ – PDA TR33

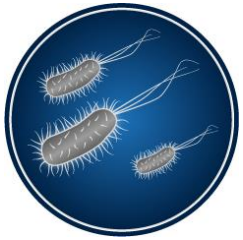
- The LOQ for the new method should be at least as sensitive as the existing method, except when a clear rationale or justification exists why a higher quantification limit can be tolerated
- If the RMM purports to have a LOQ of 1 cell, then 1 cell should be counted in test replicates when performing Accuracy and Precision studies
- As sterile samples may result when conducting studies at these low levels, mean count results of several replicates should be obtained



# LOQ Example

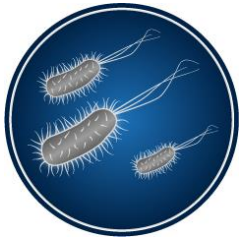
- 2005. Validation of the ScanRDI® for Purified Water Testing. McCormick, P.J.; Norton, S. E.; Costanzo, S.P. In *Encyclopedia of Rapid Microbiological Methods*, Volume 2. Edited by Michael J. Miller. PDA and Davis Healthcare International Publishing.

| Organism | Parameter | Sample Dilution | | | | |
|---|---|---|---|---|---|---|
| | | 1:2 | 1:4 | 1:10 | 1:20 | 1:40 |
| ***P. aeruginosa*** | Mean | 132.2 | 26.8 | 1.6 | 0.4 | 0.2 |
| | SD | 19.3 | 4.5 | 0.5 | 0.5 | 0.44 |
| | %CV | 14.6 | 16.9 | 34 | 136.9 | 223.6 |
| ***S. aureus*** | Mean | 94.4 | 25.4 | 2 | 0.8 | 2.6 |
| | SD | 12.5 | 7.8 | 1.22 | 0.83 | 2.8 |
| | %CV | 13.2 | 30.7 | 61.2 | 104.6 | 107.4 |
| ***B. subtilis*** | Mean | 116.8 | 31.2 | 3.0 | 0.6 | 0.4 |
| | SD | 19.1 | 3.3 | 2.12 | 0.89 | 0.89 |
| | %CV | 16.3 | 10.5 | 70.7 | 149 | 223 |
| ***C. albicans*** | Mean | 67.4 | 15 | 2.2 | 1.4 | 0 |
| | SD | 2.2 | 4.3 | 1.1 | 0.9 | 0 |
| | %CV | 3.2 | 28.7 | 49.8 | 63.9 | 0 |
| ***A. niger*** | Mean | 162.8 | 41.1 | 3.8 | 0 | 0 |
| | SD | 11.6 | 7.6 | 2.2 | 0 | 0 |
| | %CV | 3.4 | 18.4 | 60.0 | 0 | 0 |



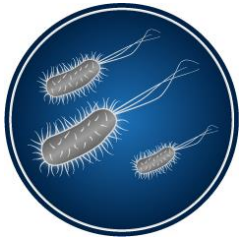
# Linearity

| Method | USP <1223> | Ph. Eur. 5.1.6 | PDA TR33 |
|---|---|---|---|
| Qualitative | | | |
| Quantitative | ✔ | ✔ | ✔ |



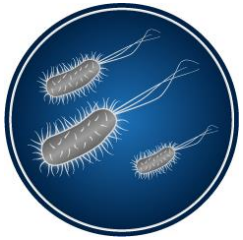
# Linearity – USP 1223

- The ability to produce results that are proportional to the concentration of microorganisms present in the sample within a given range
- *USP provides guidance on Linearity only in the Equivalence section*
- Prepare at least two samples at each of four different bioburden levels covering the range from near LOQ to one log above the specification limit defined in the compendial assay



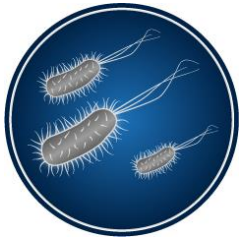
# Linearity – USP 1223

- Plot the $\log_{10}$ values of recovered counts for the alternative method (y-axis) and compendial test (x-axis)
- Correlation is acceptable if at least 0.95 (or $r^2$ value is at least 0.9025)
- In the case of a nonlinear relationship, use the Spearman (nonparametric) correlation instead of the Pearson correlation



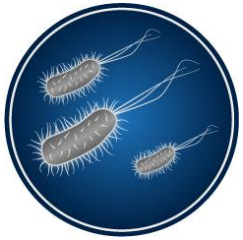
# Linearity – Ph. Eur. 5.1.6

- Ability (within a given range) to produce results that are proportional to the concentration of microorganisms present in the sample
- The concentration range should be reasonable for the purpose of the test (e.g., $10^0$ - $10^6$ CFU/mL)
- Analyze several replicates from different concentrations of microorganisms during the same working session
- If the variances for the results at each concentration is homogeneous, calculate a regression line
- The slope must be significant and the test for deviation from linearity is non-significant (refer to chapter 5.3, *Statistical analysis of results of biological assays and tests*)



# Linearity – PDA TR33

- Ability to elicit results that are proportional to the concentration of microorganisms present in the sample within a given range, where Accuracy and Precision are demonstrated
- Test at least five replicates from at least five different concentrations of microorganisms in a suitable diluent and across the range of the assay
- The mean of the replicates from each concentration are used when demonstrating Linearity
- The end-user should determine the most appropriate types of microorganisms to use

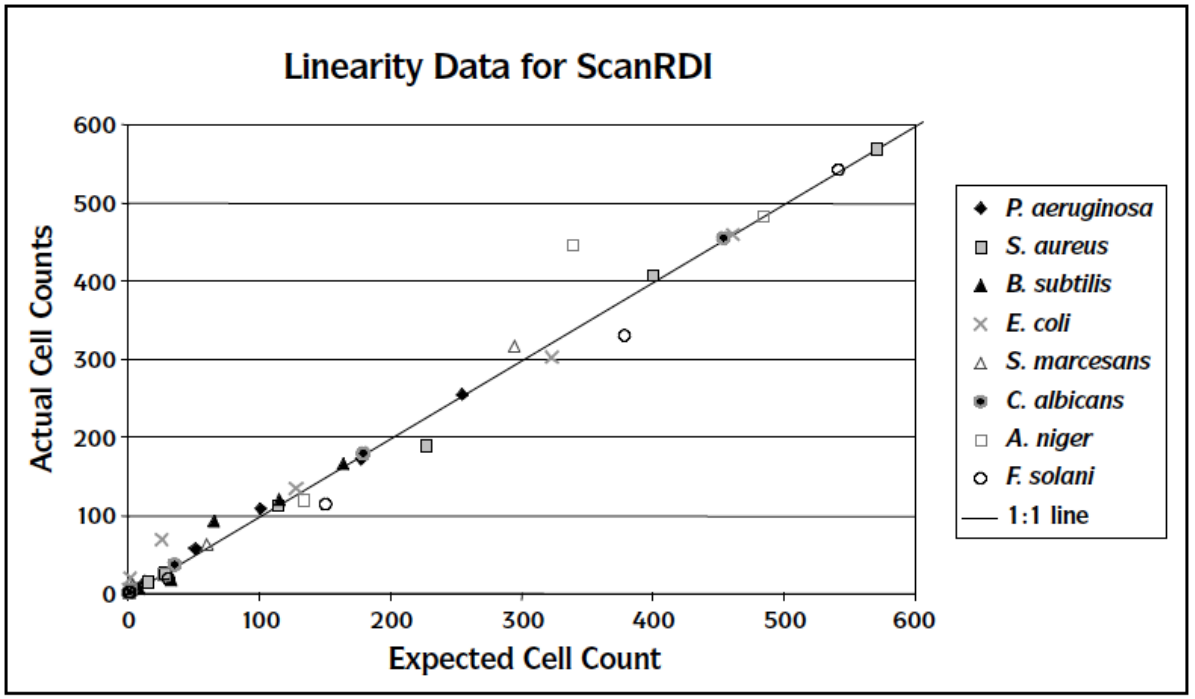


# Linearity – PDA TR33

- Use linear regression analysis
- The correlation coefficient, $r^2$, should be 0.9 or better and the slope of the line is not diverging more than 20% from 1.0
- An exception to the slope criteria may be appropriate if the new method consistently recovers higher numbers than the existing method (e.g., comparing non-growth-based methods to a growth-based reference)

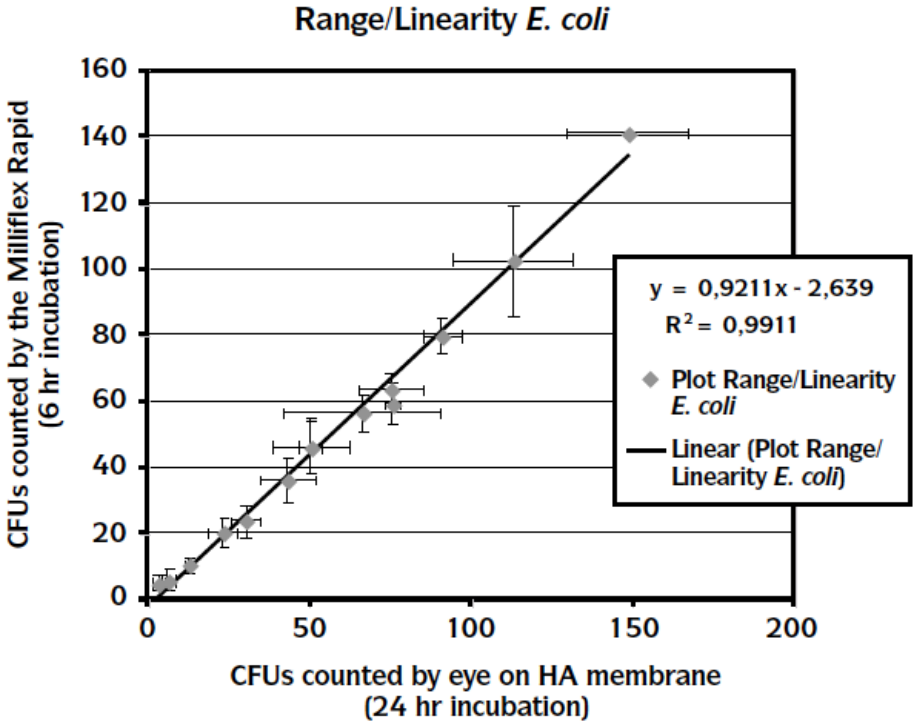
# Linearity Example

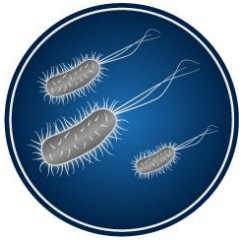
- 2005. Validation of the ScanRDI® for Purified Water Testing. McCormick, P.J.; Norton, S. E.; Costanzo, S.P. In *Encyclopedia of Rapid Microbiological Methods*, Volume 2. Edited by Michael J. Miller. PDA and Davis Healthcare International Publishing.



# Linearity Example

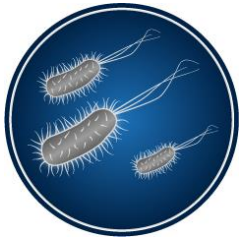
- 2005. ATP Bioluminescence Using Millipore's MilliFlex® Rapid System. Ohresser, S. In *Encyclopedia of Rapid Microbiological Methods*, Volume 2. Edited by Michael J. Miller. PDA and Davis Healthcare International Publishing.





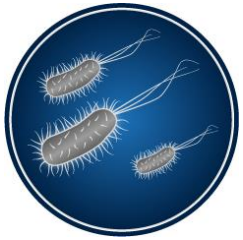
# Range

| Method | USP <1223> | Ph. Eur. 5.1.6 | PDA TR33 |
|---|---|---|---|
| Qualitative | | | |
| Quantitative | ✔ | ✔ | ✔ |



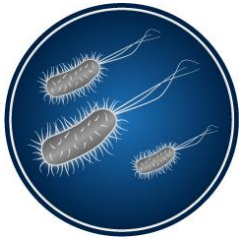
# Range – USP 1223

- The interval between the upper and lower levels of microorganisms that have been demonstrated to be determined with specified Accuracy, Precision, and Linearity
- Therefore, range is determined by testing these three validation criteria



# Range – Ph. Eur. 5.1.6

- The interval between the upper and lower levels of microorganisms as determined from the related studies of Precision, Accuracy and Linearity using the specified method
- It is dependent on the intended application

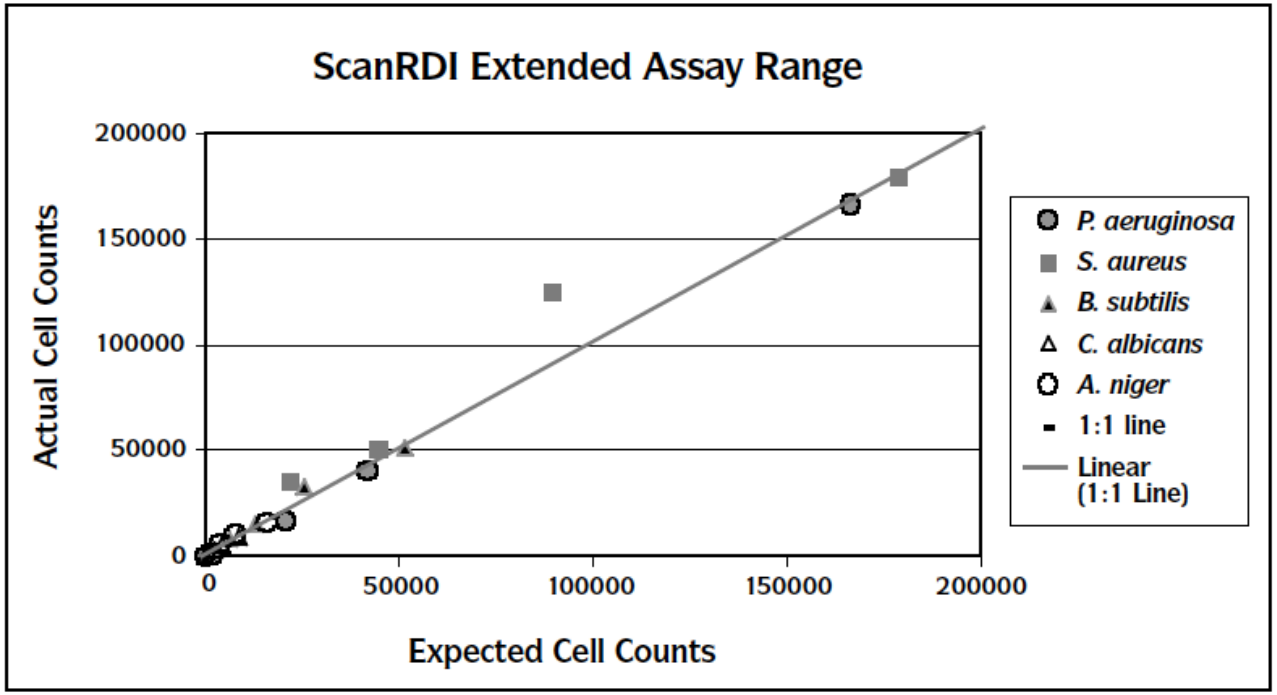


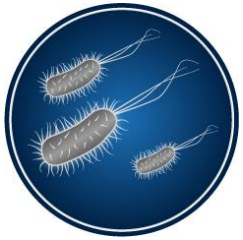
# Range – PDA TR33

- The interval between the upper and lower levels of microorganisms that have been demonstrated to be determined with Accuracy, Precision and Linearity
- The new method provides acceptable Accuracy, Precision and Linearity when applied to samples containing microorganisms in a suitable diluent at the upper and lower concentrations of the range, as well as within the range

# Range Example

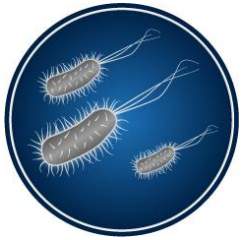
- 2005. Validation of the ScanRDI® for Purified Water Testing. McCormick, P.J.; Norton, S. E.; Costanzo, S.P. In *Encyclopedia of Rapid Microbiological Methods*, Volume 2. Edited by Michael J. Miller. PDA and Davis Healthcare International Publishing.





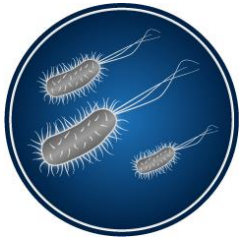
# Ruggedness

| Method | USP <1223> | Ph. Eur. 5.1.6 | PDA TR33 |
|---|---|---|---|
| **Qualitative** | ✔ | | ✔ |
| **Quantitative** | ✔ | (Intermediate Precision) | ✔ |



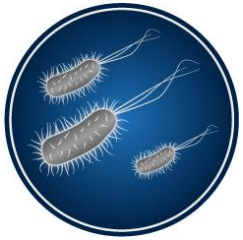
# Ruggedness – USP 1223

- The degree of precision of test results obtained by the analysis of the same samples under a variety of typical test conditions such as different analysts, instruments, testing days, and reagent lots
- The method to demonstrate ruggedness may follow supplier recommendations, or it could be based solely on data supplied by test method manufacturer



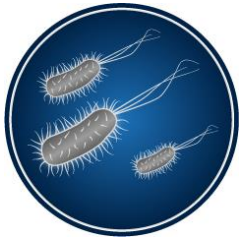
# Ruggedness – Ph. Eur. 5.1.6

- The term “Ruggedness” is not used
- Follow the recommendations for intermediate precision



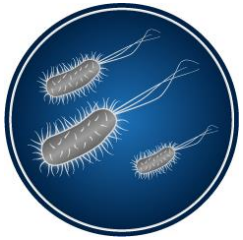
# Ruggedness – PDA TR33

- The degree of intermediate precision or reproducibility of test results obtained by assessing the same samples under a variety of normal test conditions, such as different analysts, different instruments, different lots of reagents or on different days
- **Intermediate precision** is performed within the same laboratory
- **Reproducibility** is performed between laboratories
- Ruggedness can also be considered the intrinsic resistance to the influences exerted by operational and environmental variables on the results of the method



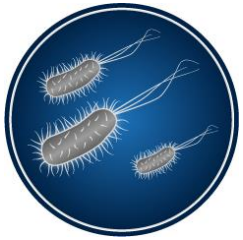
# Ruggedness – PDA TR33

- Ruggedness is best suited to be determined by the supplier of the test method who has easy access to multiple instruments and batches of components
- Data provided by the supplier are admissible to prove validation of ruggedness
- It is the responsibility of the end-user to review the supplier's data and identify gaps with respect to any modifications of the method for in-house use



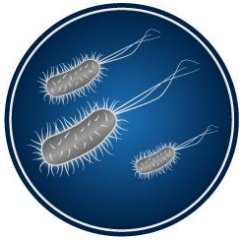
# Ruggedness Example

- Prepare a suspension of microorganisms and evaluate at least five to ten replicates against each relevant test condition
- Acceptance criteria, statistical analyses and data evaluation approaches for validation criteria (e.g., Accuracy, Precision, Specificity, LOD, LOQ, Linearity) may be applied



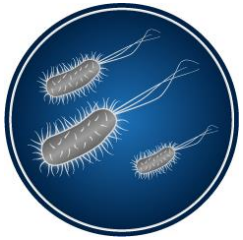
# Robustness

| Method | USP <1223> | Ph. Eur. 5.1.6 | PDA TR33 |
|---|---|---|---|
| Qualitative | ✔ | ✔ | ✔ |
| Quantitative | ✔ | ✔ | ✔ |



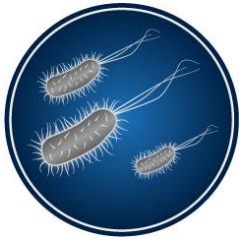
# Robustness – USP 1223

- A method's capacity to remain unaffected by small but deliberate variations in method parameters, such as, reagent volume, time or temperature of incubation providing an indication of its reliability during normal usage
- It is a necessary component of validation of the alternate method so that the user understands the limits of the operating parameters of the method
- The user may rely on data supplied by test method supplier



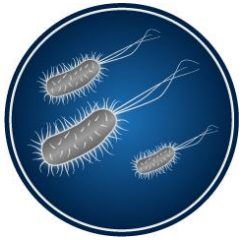
# Robustness – Ph. Eur. 5.1.6

- A measure of a method's capacity to remain unaffected by small but deliberate variations in method parameters (e.g. incubation period or incubation temperature range)
- It is best suited to determination by the supplier
- Nevertheless, if the user modifies critical parameters, the effects on robustness must be evaluated
- Robustness is judged by its ability to detect or accurately enumerate test microorganisms after deliberate variations to the method parameters



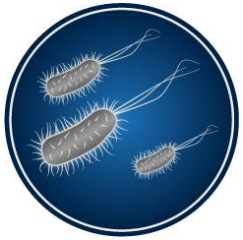
# Robustness – PDA TR33

- A measure of a method's capacity to remain unaffected by small but deliberate variations in method parameters and provides an indication of its reliability during normal usage
- Data provided by the supplier are admissible to prove validation of robustness
- It is the responsibility of the end-user to review the supplier's data and identify gaps with respect to any modifications of the method for in-house use



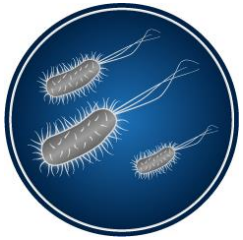
# Robustness – PDA TR33

- If the new method is shown to fail at providing robust results for particular test conditions, this should be accepted as a limitation(s) of the method
- If the new method is shown to be particularly sensitive towards a certain type of procedure, system parameter or manipulation, the results should be used to define adequate precautions or limitations when the method is used routinely



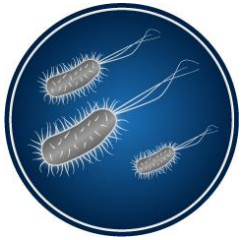
# Robustness Example - PDA

- The same procedure recommended for Ruggedness testing may be used while changes on identified critical method and system parameters are introduced
- For each test condition a range should be demonstrated, within which the new method operates in a robust manner (e.g., no significant difference)
- Reagent concentrations, instrument operational limits, and incubation parameters (e.g., time and temperature for methods requiring microbial growth) may be assessed



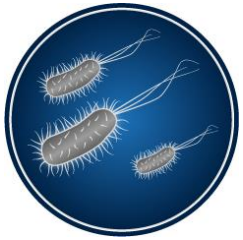
# Equivalence

| Method | USP <1223> | Ph. Eur. 5.1.6 | PDA TR33 |
|---|---|---|---|
| Qualitative and Quantitative | Without product | Without product<br><br>Alternatively, and in some cases additionally, with product | With product |



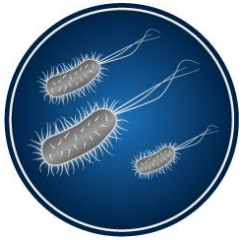
# Equivalence – USP 1223

- When the test results from two procedures are sufficiently close for the intended use of the procedures
- Demonstration of equivalence requires a pre-specified measure of how similar the test results need to be
- Equivalency is intended to show, using standardized microorganism challenges, the alternative method is equivalent or non-inferior to the compendial method



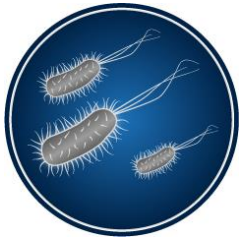
# Equivalence – USP 1223

- USP states Equivalence is demonstrated in the **absence** of product
- However, one sentence in the chapter contradicts this:
- “After an alternative method has been shown to be equivalent to the compendial test **with one product**, it is not necessary to repeat the equivalency parameters for every new product; it is merely necessary to verify the method suitability for each additional product.”
- Additional clarity from USP is warranted, especially since PDA TR33 and Ph. Eur. 5.1.6 (in some instances) states to test Equivalence with product



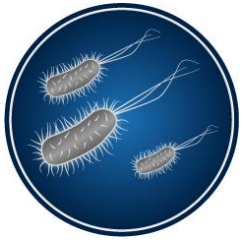
# Equivalence – USP 1223

- Support for equivalence may come from peer-reviewed papers or regulatory submissions (e.g., a vendor submitted Drug Master File to the FDA, or prior submission from a company on a technology)
- However, this may not be sufficient for the manner in which the method will be used; the end-user may need to determine if additional Equivalence testing is required



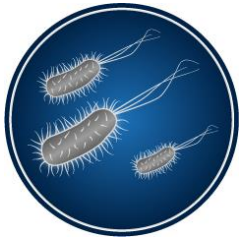
# Equivalence – USP 1223

| Option | Demonstration | Comparison to Compendial Method | Based on Numerical Results or Conclusion | Number of Characteristics |
|---|---|---|---|---|
| Acceptable Procedures | Acceptable | NO | Results | Multiple |
| Performance Equivalence | Equivalent | YES | Results | Multiple |
| Results Equivalence | Equivalent | YES | Results | Single |
| Decision Equivalence | Equivalent | YES | Conclusions | Single |



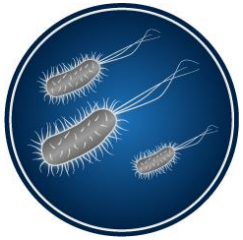
# Equivalence – USP 1223

- **Option 1 – Acceptable Procedure**
- No comparison between an alternative and a compendial method
- A reference material with known properties is used to demonstrate performance characteristics or acceptance criteria are met
  - Standard inoculum of a specific microorganism, highly purified nucleic acid material, ATP or another appropriate signal specific to the method
- It may be required to measure the signal in the presence of the test sample using validation criteria that are appropriate for the technology, although there is no explanation under what conditions this would be required



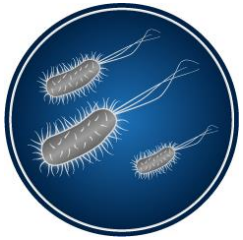
# Equivalence – USP 1223

- **Option 2 – Performance Equivalence**
- Comparison of multiple validation criteria between an alternative and a compendial method
- Use validation criteria that are relevant to the alternative method (e.g., Accuracy, Precision, LOD, LOQ, etc.)
- Although an alternative method may not meet certain validation parameters it may still be acceptable for use because of other advantages (e.g., time to result)
- This option appears to be the closest to PDA TR33 and Ph. Eur. 5.1.6



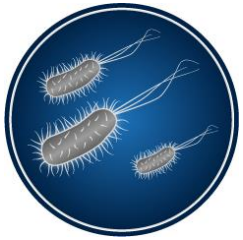
# Equivalence – USP 1223

- **Option 3 – Results Equivalence**
- Demonstrates an alternative and compendial method give an equivalent **numerical** result; a tolerance interval is established when comparing the two methods (*maybe the Δ?*)
- The alternative method is shown to be numerically superior or non-inferior (*NOTE: regulators do not require an alternative method to be superior*)
- Because some non-growth-based methods may produce significantly higher cell counts than a growth-based method that reports outcomes in CFU, a calibration curve showing a correlation between the two methods in the product specification range can be used



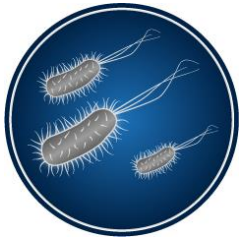
# USP Option 3 Example

- Some alternative methods may provide a different signal than the CFU and this signal may be numerically different in magnitude and units
- Therefore, USP recommends demonstrating equivalence between the two methods using precision (repeatability) and correlation (linearity)
- Follow the same strategies for Precision and Linearity as previously discussed



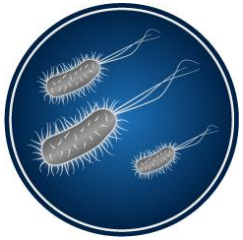
# USP Option 3 Example

- If linearity is not demonstrated, use Option 4 as an alternative
- The user establishes a qualitative acceptance criterion for the new method that would match the quantitative specification in the compendial test
- e.g., if the required is NMT $10^2$ CFU (maximum acceptable count is 200 CFU), the laboratory will need to determine an acceptance criterion for the alternative procedure that will match that value from the perspective of making a decision regarding microbial quality
- Validate as a qualitative test (*could be similar to a dilute-to-spec method*)



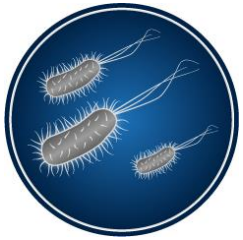
# Equivalence – USP 1223

- **Option 4 – Decision Equivalence**
- Demonstrates an alternative and compendial method give an equivalent **qualitative** result, such as a pass/fail outcome
- The incidence of positive to negative results for an alternative method should be no worse (i.e., the method is non-inferior) to the results obtained with the compendial method
- Spiking the test samples with low levels of microorganisms may be considered



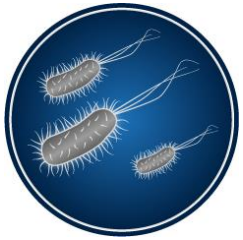
# USP Option 4 Example

- Organisms represent what has been recovered from the product, present a risk to patients and/or are recommended in the USP. For each organism, perform the following:
- Characterize the *sensitivity* of the method with 1 CFU
- Establish the *acceptability* of the method with 100-200 CFU (detection should occur at least 75% of the time)
- Demonstrate non-inferiority ($\Delta = 0.20$) with 10-50 CFU (detection should occur 50-75% of the time)
- In the non-inferiority experiment, use at least 75 replicates for each assessment, providing an 80% test power, or 100 replicates for a 90% test power



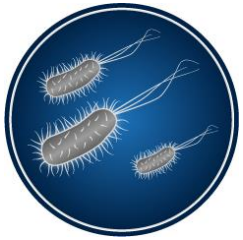
# USP Option 4 Example

- Alternatively, use a MPN approach
- MPN results obtained by both methods are converted to $\log_{10}$ values and the sample mean and the sample variance of the log values are determined
  - Run a test for normal distribution to see if you really need to convert data into log values
- Demonstrate non-inferiority with the resulting data



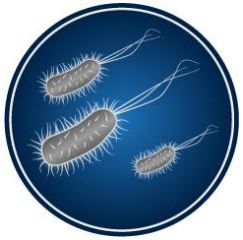
# Equivalence – Ph. Eur. 5.1.6

- Equivalence is demonstrated by performing the validation parameters for a qualitative or quantitative method
- Use low levels (e.g., less than 5 CFU) of relevant strains of microorganisms and a relevant number of replicates
- Alternatively, and in some cases in addition to performing testing using a panel of microorganisms, demonstrate equivalence via parallel testing of a predefined number of test samples or for a period of time (this is justified based on a risk assessment)
- The chapter does not provide any additional guidance



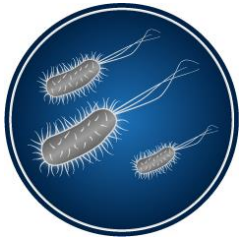
# Equivalence – Ph. Eur. 5.1.6

- If an alternative quantitative method result is expressed as a CFU, a statistical analysis shall demonstrate the method will comply with the standards of the relevant monographs (i.e., meet the same acceptance criteria in terms of CFU per weight or volume)
- However, if the alternative method result cannot be expressed as a CFU, then *suitable parameters* shall be used, followed by a statistical analysis to demonstrate the alternative method will comply with the standards of the relevant monographs
- The chapter does not provide any additional guidance



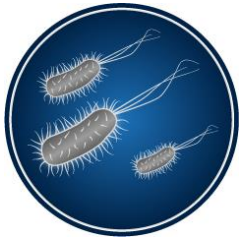
# Equivalence – PDA TR33

- A measure of how similar the new method test results are when compared with the existing method
- TR33 recommends performing equivalency in the presence of product or actual test samples
- Prior to testing, the test material must have been assessed for the potential to cause background noise, interference, false positive or false negative results (method suitability)



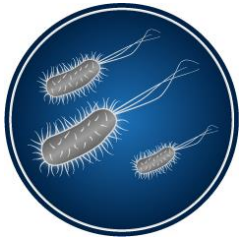
# Equivalence – PDA TR33

- When possible, test samples should be identified that are expected to contain microorganisms
- When test samples are sterile, it will be necessary to challenge the samples with microorganisms
- Statistically, you can’t compare zeros to zeros



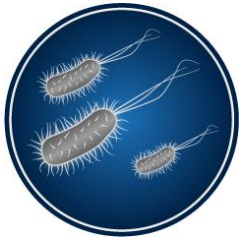
# Equivalence – PDA TR33

- Strategies for inoculating test samples can be the same as what was used for validation criteria testing with microorganisms in a suitable diluent
- The end-user should determine the numbers, types and physiological state of the challenge microorganisms
- Regulators expect very low levels of stressed organisms, including slow growers and facility isolates, to be utilized when validating an alternative sterility test
- For qualitative nucleic acid amplification detection methods, samples may need to be inoculated with actual target microorganisms or with nucleic acid standards



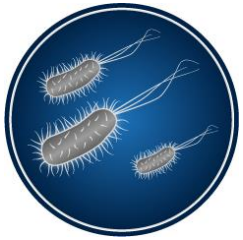
# Equivalence – PDA TR33

- The new method is run in parallel with the existing method for a specified period of time or number of product batches or test samples
- The end-user should determine the most appropriate strategy for the duration and extent of these studies, which may be influenced by the critical nature of the test method, the material being analyzed, the statistical methods used when interpreting the resulting data, regulatory expectations and/or other quality requirements



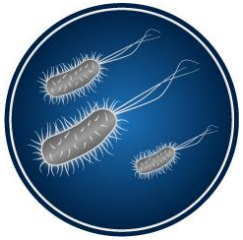
# Equivalence Example

- At least 3 independent tests using at least 3 different lots/batches of the test sample
  - 3 separate manufacturing lots of an in-process sample evaluated over a three-day period (one lot per day)
  - 10 separate use points on a purified water loop tested once per week over a one-month period
  - 6 separate tanks containing mammalian cell culture assessed every day for a period of one week
  - 5 different environmental monitoring sampling sites evaluated at least 3 times during the course of an 8-hour filling run and for an entire year to evaluate seasonal changes



# Equivalence Example

- Sterility test: compare the new method with the compendial test using an LOD assay, using very low levels of microorganisms inoculated into sterile product, and statistically comparing the rates of positive to negative results using a Non-Inferiority Test
- Bioburden test: compare the new method with an existing bioburden assay using Accuracy and Precision studies and an Equivalency or Non-Inferiority Test



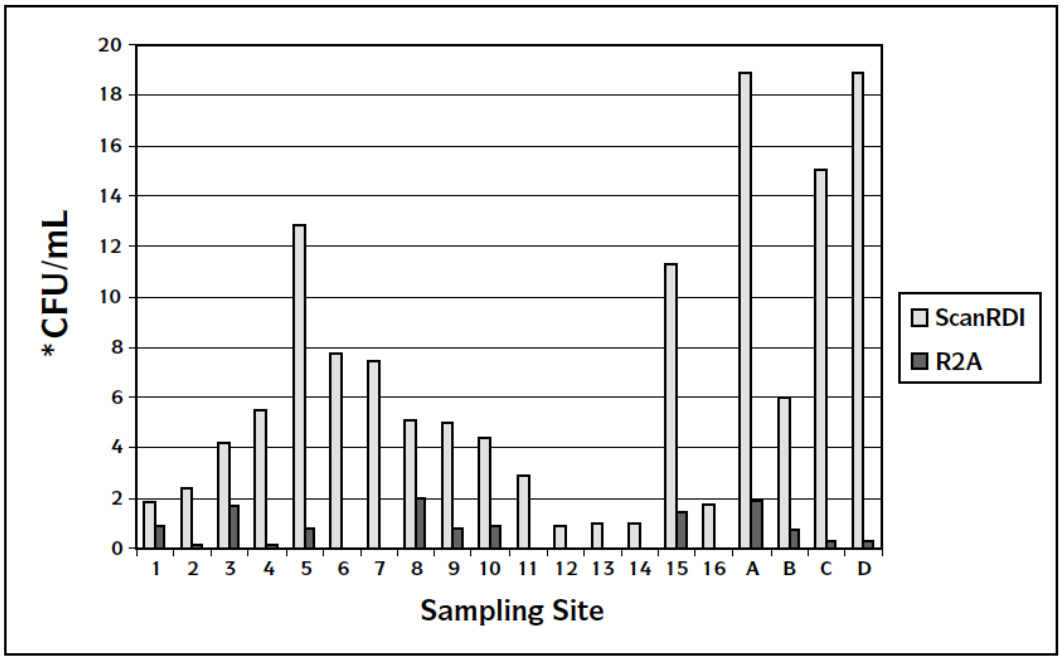
# Equivalence – PDA TR33

- The new method must be shown to be at least statistically equivalent or statistically non-inferior (i.e., it is not worse), to the existing method
- The new method may also be shown to be statistically superior, but this is not required
  - e.g., higher recovery, a greater amount of microbial detection or a lower LOD to the existing method

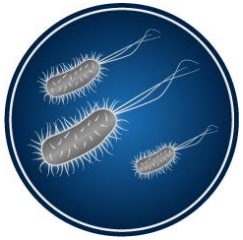
# Equivalence Example

- 2005. Validation of the ScanRDI® for Purified Water Testing. McCormick, P.J.; Norton, S. E.; Costanzo, S.P. In *Encyclopedia of Rapid Microbiological Methods*, Volume 2. Edited by Michael J. Miller. PDA and Davis Healthcare International Publishing.

**Comparison of Mean ScanRDI Actual Counts and Mean R2A Plate Counts Over a Period of Three Months from Twenty Sampling Sites (1-16 Manufacturing Sites, A-D Laboratory Sites).**

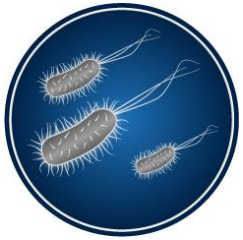



*Indicates the actual cell counts for ScanRDI and CFU/mL for viable count method.



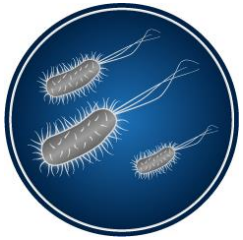
# Equivalence Example - Novartis

- 2010. Gray, J.C.; Staerk, A.; Berchtold, M.; Mercier, M.; Neuhaus, G.; Wirth, A. Introduction of a Rapid Microbiological Method as an Alternative to the Pharmacopoeial Method for the Sterility Test. American Pharmaceutical Review. 13(6): 88-94.



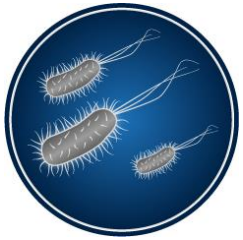
# Equivalence Example - Novartis

- Six drug products; 90 Rapid vs. 90 compendial sterility tests (3 strains x 3 runs x 10 reps = 90 tests)
- Challenged each product with 1-5 CFU using three different strains of stressed microorganisms
- A statistical analysis was performed with the Chi-Square and Fisher's exact test
  - NOTE: equivalency or non-inferiority tests may be more appropriate
- Only method suitability was performed on subsequent products



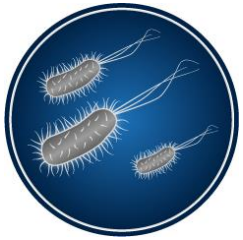
# USP and the CFU

- “Attempts to use statistics to compare the CFU results to signals arising from biochemical, physiological, or genetic methods of analysis may have limited value because the different methods used cannot be expected to yield signals that could be compared statistically in terms of mean values and variability.”
- USP concludes the CFU cannot be used as acceptance criteria and it is the user’s responsibility to propose values (supported when necessary by scientific literature) that they can demonstrate are appropriate for the alternative method
- However, the chapter contradicts this position



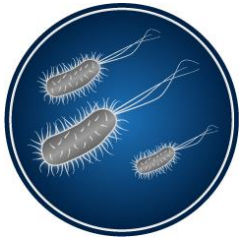
# USP and the CFU

- Option 3 (Results Equivalence)
- “Reports on the use of alternative non-growth-based methods have shown that they may produce significantly higher cell count estimates than a growth method that reports outcomes in CFU. In this case, the analyst could **use a calibration curve showing a correlation** between the two methods in the product specification range.”



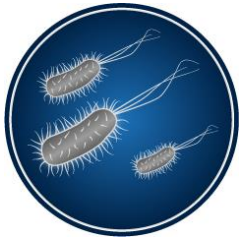
# Additional USP Discussion Points

- The presence of a greater number of cells based on an alternative method with a signal other than CFU has not correlated with more user risk or a higher likelihood of pathogens being present when there is an established safety record. **Do you agree?**
- USP states if an alternative method is not as sensitive as the compendial method but has other advantages, such as a reduced time to result, the alternative method may still be used as long as it "allows for a quality decision on the product that is non-inferior to the compendial method." **Is this appropriate for a critical assay such as the sterility test?**



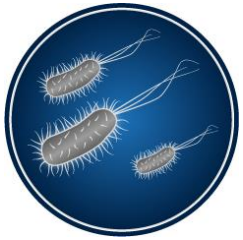
# Method Suitability

- Demonstrates the actual test sample or product does not interfere with a method's ability to detect or recover microorganisms
- Interference may include the generation of background noise that results in a false positive response or an inhibition of chemical reactions that are required to detect microorganisms, thereby generating a false negative result
- USP, Ph. Eur. and TR33 provide guidance on how to demonstrate method suitability



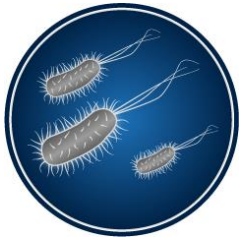
# Method Suitability – USP 1223

- Demonstration of lack of enhancement or inhibition by the product on the signal generated by the method
- Suitability to the specific test; e.g., compendial requirements for the quantity of material to be tested
- Use the number of units, quantities prescribed and sample preparation appropriate for the product and the required test sensitivity to determine the absence of a product effect that would obscure the signal of the alternative method



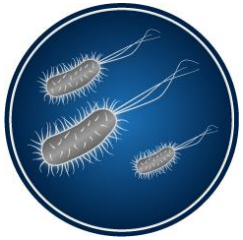
# Method Suitability – USP 1223

- Demonstrated using three independent tests
- Only Accuracy and Precision are required for quantitative methods
- Recovery of challenge organisms as indicated in <62>, <71> and <1227> is sufficient for qualitative methods
  - *NOTE: USP states to perform method suitability at the required test sensitivity, which may be in conflict with the test chapters in which you use < 100 CFU*
  - *Unclear about the actual organisms to use*



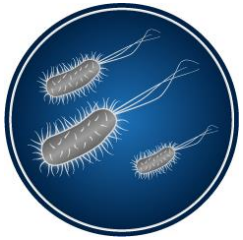
# Method Suitability – USP 1223

- "After an alternative method has been shown to be equivalent to the compendial test **with one product**, it is not necessary to repeat the equivalency parameters for every new product; it is merely necessary to verify the method suitability for each additional product."
- *NOTE: this infers you will use at least one product when demonstrating equivalency*



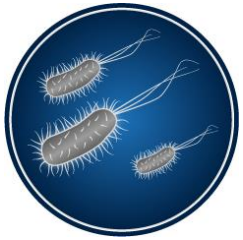
# Method Suitability – Ph. Eur. 5.1.6

- The alternative method must be applied according to the specified procedure and with the samples to be analyzed
- The test sample must not interfere with the system's detection capacity or microbial recovery Address the ability of the test to detect microorganisms in the presence of the sample matrix
- Verify if the sample matrix interferes with the alternative system (e.g. background signal or inhibiting chemical reactions
- Acceptance criteria for the method in routine use will need to be defined as a function of the application and the validation data



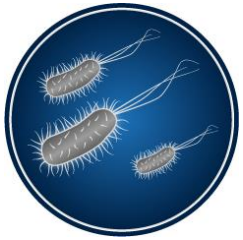
# Method Suitability – PDA TR33

- Each test material should be evaluated for the potential to produce interfering or abnormal results
  - **False positives**:  a positive result when no viable microorganisms are present
  - **False negatives**: a negative result when microorganisms are present
- Can also evaluate the impact of cellular debris, dead organisms or mammalian cell cultures



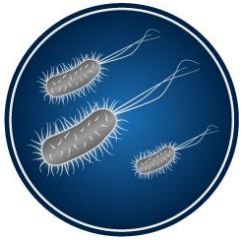
# PDA TR33 - False Positive Testing

- Determines if the test sample contains components that produces background noise or interfering signals, resulting in a false positive result
- The sample should contain no viable microorganisms
  - Treat the sample such that it will not contain viable microorganisms and sample properties are not altered
- The size (e.g., volume, weight) of the test sample should be the same as what will be used during routine analysis
- The evaluation should be performed using an appropriate number of replicates and sample batches



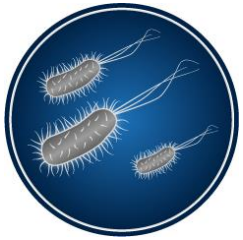
# PDA TR33 - False Positive Testing

- False positives should be reviewed to determine if the result is due to contamination of a supposedly sterile test sample
  - Caution: some contaminants may not be detectable by classical methods (viable but non-culturable, VBNC)
- True false positives should be resolved before the new method is used to routinely to test a particular product or sample matrix
  - May be resolved with dilutions, rinsing or other strategies



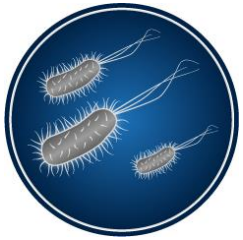
# PDA TR33 - False Positive Testing

- If false positives cannot be resolved, the test sample may be incompatible with the new method
- However, some end-users may find a low false positive rate is still acceptable if a follow-up confirmatory test is utilized
- Also, the presence of normal background noise should be fully understood, and in the event this interference is unavoidable, the end-user should determine if the background noise is an acceptable baseline for detecting microorganisms in the test sample under routine use



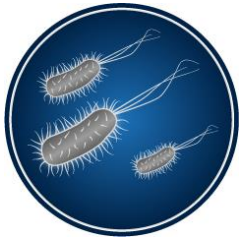
# PDA TR33 - False Negative Testing

- Determines if the test sample contains material that may quench, mask or otherwise prevent the detection or enumeration of microorganisms when they are present, thereby producing a false negative result
- The test sample should be inoculated with a known level and appropriate type of microorganism
- A positive control should be prepared which is inoculated the same way but does not contain the test sample
  - e.g., sterile buffer



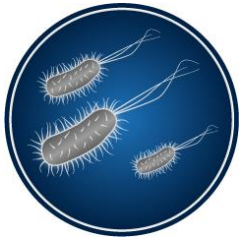
# PDA TR33 - False Negative Testing

- The size of the test sample should be the same as what will be used during routine analysis
- Use an appropriate number of replicates and test sample batches
- When applicable, markers specific for the technology being evaluated may be added to the test sample to demonstrate that the marker will be detected under the conditions of the test (e.g., ATP)



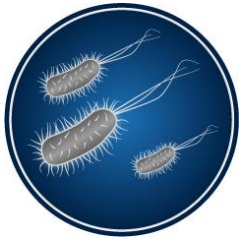
# PDA TR33 - False Negative Testing

- True false negatives should be resolved before the new method is used to routinely to test that particular product or sample matrix
  - May also be resolved with dilutions, rinsing or other strategies
  - Could be a neutralization issue (preservatives, etc.)
- If false negatives cannot be resolved, then the specific test sample may be incompatible with the new method



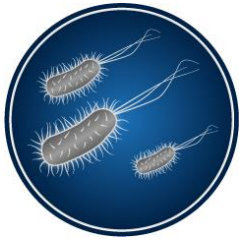
# Method Suitability Example
# Novartis Rapid Sterility Test

- To test for false positives, background signals or interference, they filtered sterile product and showed no recovery in the RMM
- To test for false negatives, they filtered each product and added 10-100 CFU of compendial ATCC organisms, 1-2 in-house isolates and stressed *P. acnes* in the final rinse
- The filter was placed on RSTM and incubated 3 days (bacteria) or 5 days (stressed *P. acnes*, in-house isolates and fungi)
- CFUs were visually counted unless there were no CFU's observed after incubation. In this case, fluorescent counts were obtained with the RMM.
- Compared counts with positive control (Fluid A; no product). Showed at least 70% recovery. Repeated 3x for each organism.



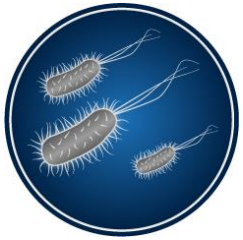
# Method Suitability Example - Statistics

- False negative testing on a new quantitative method can compare recovered counts with a control (no product) using a 70% recovery rule and a 2-sample equivalency or non-inferiority test (similar to Accuracy)
  - A traditional t-test may be used to show there are no statistical differences
  - ANOVA may be used when more than 2 groups are being compared



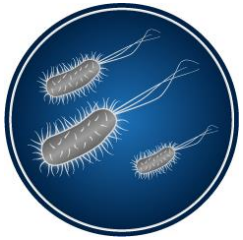
# Identification Methods

| Validation Criteria | USP <1223> | Ph. Eur. 5.1.6 | PDA TR33 |
|---|---|---|---|
| Accuracy | ✔ | ✔ | ✔ |
| Precision | ✔ | | ✔ |
| Specificity | | ✔ | |
| Robustness | | ✔ | |



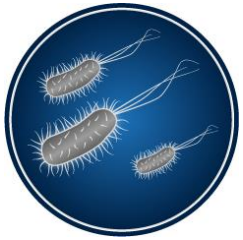
# Identification Methods - Accuracy

- No acceptance criteria in Ph. Eur. 5.1.6; use well characterized organisms such as type strains
- PDA TR33 follows guidance in USP 1113
- USP 1113 states the user should establish suitable acceptance criteria taking into account method capability
- But USP infers > 90% agreement can be achieved for microorganisms that **are appropriate** for the ID system
- Accuracy % = (# of correct IDs / Total # of IDs) x 100



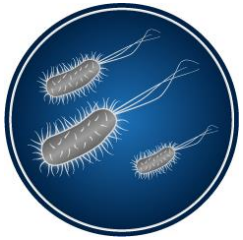
# Identification Methods - Precision

- PDA TR33 follows guidance in USP 1113
- USP 1113 states the user should establish suitable acceptance criteria taking into account method capability
- Precision % = (# of correct IDs in agreement / Total # of IDs) x 100
- **NOTE:** from personal experience, you may achieve > 95% repeatability and > 90% intermediate precision



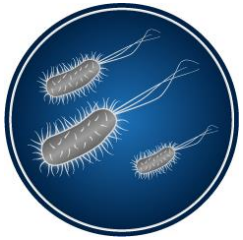
# Identification Methods - Specificity

- Only in Ph. Eur. 5.1.6
- Ability to discriminate microorganisms actually present from interfering factors that cause false identification results
- Chemical substances, mixtures of microorganisms, presence of mixtures of DNA from more than one organism



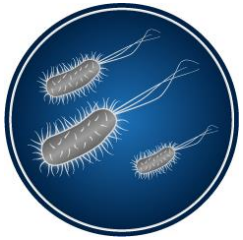
# Identification Methods - Robustness

- Only in Ph. Eur. 5.1.6
- Ability to correctly identify the test microorganisms after deliberate variations to method parameters (e.g. incubation period or incubation temperature range)
- Best suited to determination by the supplier
- If the user modifies critical parameters, the effects on robustness have to be evaluated



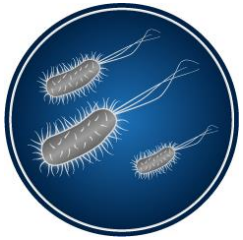
# Identification Methods Considerations

- May consider pooling data from multiple runs (different organisms)
- The microbial identification system may not be able to identify an isolate because the organism is not included in the database, the system parameters are not sufficiently comprehensive to identify the organism, the isolate may be nonreactive in the system, or the species may not have been taxonomically described



# Japanese Pharmacopoeia 17

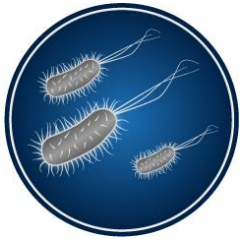
- Rapid Microbial Methods chapter is similar to USP
- “When using a type strain, the result of validation should be equivalent to or better than that of the conventional method. However, because the detection principles of new methods are usually different from that of conventional methods, the correlation between them is not always required.”
- “For detection of environmental bacteria, it is important that the physiological state of the type strain should be maintained as close as possible to that of environmental bacteria, in order to obtain reliable results.”



# Japanese Pharmacopoeia 17

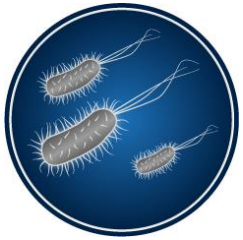
- “Although, it is important in principle that a new method should have an equal or greater capability than the conventional method, a new method may be used after verifying their validity, even in the absence of equivalence to conventional methods.”

- Additional clarification to fully understand the chapter is warranted



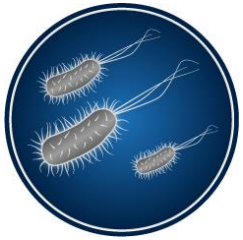
# Alternative and Rapid Endotoxin Methods

- Methods are expected to meet the appropriate compendial requirements for the validation of endotoxin methods
- Validation of the instrumentation usually involves verification that standard or reference solutions of endotoxin yield the specified standard curves, analytical responses and/or LOD or LOQ
- Methods that do not meet existing compendial requirements should be validated as an alternative method
  - Recombinant Factor C; phage protein binding methods



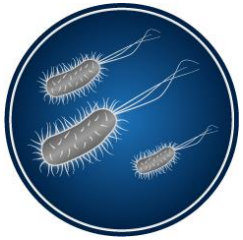
# Alternative and Rapid Methods for Mycoplasma Detection

- May use the same validation strategies as specified in TR33
- Recommend to review PDA TR50, *Alternative Methods for Mycoplasma Testing*
- Ph. Eur. Chapter 2.6.7, *Mycoplasmas*



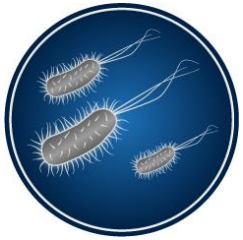
# Unique Methods Requiring Additional or Modified Validation Strategies

- Some alternative or rapid detection systems may not be compatible with liquid microbial suspensions
- Technologies that utilize airborne, aerosol, or other non-liquid-based samples may fall into this category
- Expanding or adapting the validation strategies may be appropriate, as long as the testing methods are scientifically justified
- Because the end-user may not posses the expertise or specialized equipment to conduct such studies, the supplier may need to play a greater role during validation



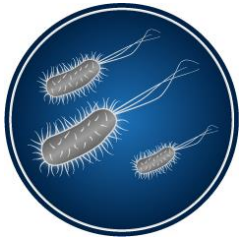
# Guidance on Changing Acceptance Criteria

- Some alternative or rapid methods will not report results in terms of a CFU
- May include number of viable cells, a spectral analysis, relative light units, fluorescent units, etc.
- Alternative or rapid methods may also be more sensitive than classical methods (e.g., recovery higher numbers of stressed organisms)
- The trending of data may be lost at some point in time in order to bridge the gap between “old” and “new” data analysis, and this is understood by regulators
  - 2006. Hussong and Mello, *American Pharmaceutical Review*. 9(1): 62-69.



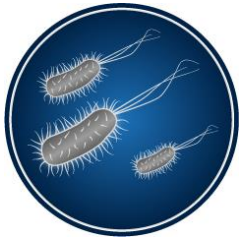
# Guidance on Changing Acceptance Criteria

- In the event that a new method is qualified to provide greater sensitivity than the method intended to be replaced, an understanding of the impact to existing acceptance levels, in-process or product specifications, and compendia and regulatory expectations is required
- A statistical comparison between the data observed between the two methods should be performed, when appropriate, and this information may be used to justify any recommendations for changes to current acceptance levels and/or specifications
- In these cases, discussions with relevant regulatory agencies is highly recommended



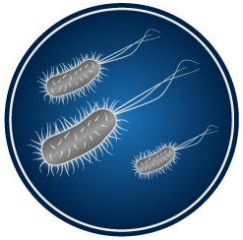
# Automated Methods

- Some alternative or rapid technologies may be considered as automated traditional or compendial microbiological test methods, especially when the results are in CFU
- These technologies may be qualified for their intended use without the need for demonstrating certain method validation requirements
- At least Accuracy and Precision assessments should be performed, in addition to Method Suitability and Equivalence/Comparability studies
- A risk assessment should be performed to support a reduced validation strategy



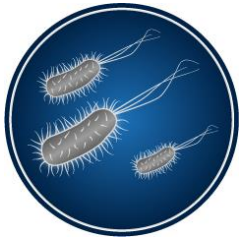
# Implementation and Tech Transfer

- Once the validation plan has been executed and approved, the RMM may be implemented for routine use
- It is *usually* not necessary to repeat the same qualification test plan (in its entirety) for identical technologies that will be installed in the same location or at secondary facilities at a different geographic location (e.g., manufacturing sites)
- In this case, a copy of the original qualification package can be provided to the secondary location, and a reduced test plan developed for the installation and qualification of the identical technology at that site



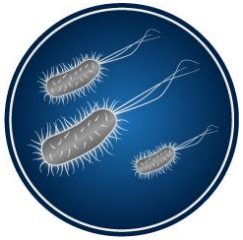
# Implementation and Tech Transfer

- When an identical technology is installed at the secondary location, a standard equipment IQ and OQ should be performed following the original qualification package
- The original hardware/software security configuration testing may not need to be repeated unless there will be systems in use at a secondary site that were not evaluated in the initial test plan
  - Different data handling or archiving platform
- Other environmental conditions may need to be addressed (e.g., altitude, humidity or temperature)



# Implementation and Tech Transfer

- A reduced microbiological challenge from the original qualification should be performed to demonstrate that the system is operating as intended
  - A few reference organisms, identical to what was used during the original qualification, to confirm basic functionality and that key requirements are met (e.g., Accuracy and Precision)
- If the reference strains used during the original method validation are not representative of the isolates recovered at the secondary site, the local qualification plan should include these facility isolates

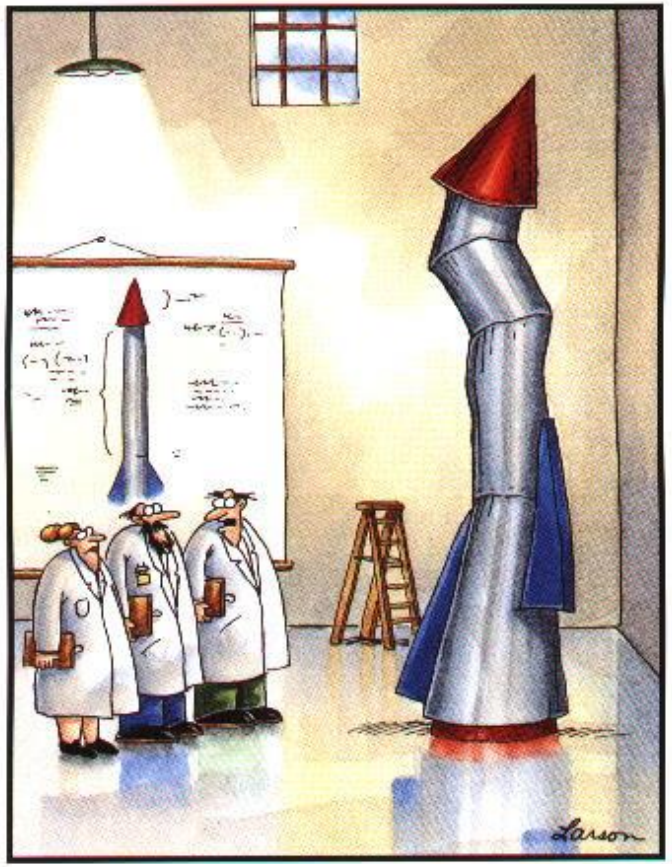


# Implementation and Tech Transfer

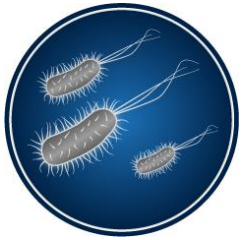
- Any product or test samples that were not included in the original validation must be evaluated for Equivalency and Method Suitability
- However, product or test samples that were included in the original validation should be evaluated for Equivalency, as the microbial load (number and type of microorganisms) may be different
- Similarly as with the original test site, all training, SOPs, and maintenance and calibration programs should be developed prior to performing the qualification

# Validation Summary

- You are now validation experts!
- Use a sound scientific approach
- Discuss with the regulatory authorities
- There are numerous validation case studies that can be reviewed for additional information




**“It’s time we face reality, my friends …**
**“We’re not exactly rocket scientists.”**



# Publications of Interest

- 2015. Miller, M.J. A fresh look at USP <1223> *Validation of Alternative Microbiological Methods* and how the revised chapter compares with PDA TR33 and the proposed revision to Ph. Eur. 5.1.6. American Pharmaceutical Review. 18(5): 22-35.
- 2017. Miller, M.J. A Comprehensive Review of the Revised European Pharmacopoeia Chapter 5.1.6. Part 1. *European Pharmaceutical Review*. 22(5): 20-24.
- 2017. Miller, M.J. A Comprehensive Review of the Revised European Pharmacopoeia Chapter 5.1.6. Part 2. *European Pharmaceutical Review*. 22(6): 27-30.