Regulatory and Scientific Justification of Method Validation Acceptance Criteria

Steven R. Miller, Ph.D.

Presented to AAPS Southern California Pharmaceutical Discussion Group

January 20, 2005

Presentation available for download from www.drstevenmiller.com

Disclaimer

The contents of this presentation do not necessarily represent the policies of any U.S. or international regulatory agency or any company, past or present, by which I am employed (including, but not limited to my current employer, Watson Laboratories).

What is Method Validation?
Regulatory Definitions

FDA (Analytical Procedures and Methods Validation, August 2000):

> ***Method validation is the process of demonstrating that analytical procedures are suitable for their intended use.***

§211.194(a)(2):

> ***… The statement shall indicate the location of data that establish that the methods used in the testing of the sample meet proper standards of accuracy and reliability as applied to the product tested.***

Definitions continued ~

USP <1225>:

Validation of an analytical method is the process by which it is established, by laboratory studies, that the performance characteristics of the method meet the requirements for the intended analytical applications.

Definitions continued ~

ICH (Q2A Validation of Analytical Procedures, March, 1995):

> ***The objective of validation of an analytical procedure is to demonstrate that it is suitable for its intended purpose.***

What is Method Validation?

Intention

The intended use of analytical methods is to assess product quality and validation is the process of generating experimental data that provides evidence that the performance of an analytical method is adequate for reliably assessing product quality. Method validation acceptance criteria necessarily reflect what we believe is "adequate performance."

Sources of Information

- The USP - All of it, not just general chapter <1225>.
- The FDA's Draft Guidance on Analytical Procedures and Methods Validation (August 2000).
- The FDA's Guidance on Validation of Chromatographic Methods (November, 1994).
- The FDA's Guidance on Investigating Out of Specification (OOS) Test Results for Pharmaceutical Production (September, 1998).
- ICH Q2A and B Validation of Analytical Procedures: Text and Methodology.
- Historical information on origin of some USP tests.
- Literature from the period when system suitability started to enter USP methods.

Adequate Method Performance

- Method performance expectations as they relate to product potency (and indirectly the method's analyte concentration).
  - System suitability criteria for USP methods are typically the same, regardless of product potency.
  - No validation regulatory guidance makes mention of consideration of potency in regard to validation acceptance criteria

Adequate Method Performance

- Bioequivalence acceptance criteria are typically the same, regardless of product potency. Therefore, performance of bio-relevant dissolution methods should be the same.
- Dissolution, assay, and CU criteria are typically the same, regardless of product potency.
- The reporting limits for impurities are typically the same (except for very large doses).

Adequate Method Performance

- USP monographs are considered official regulatory methods and therefore their performance characteristics must be considered “adequate.”
- Clues to “adequate performance” can also be found in <467>, <711>, and <724>.
- Regulatory Guidances do not provide clues to “adequate performance”, but they indirectly imply that regulatory methods are a good place to look.

Adequate Method Performance

- Validation acceptance criteria described in this presentation will be independent of product potency and analysis technique.
- Regulatory methods (USP monographs and general chapters) will be our starting point for deriving validation acceptance criteria.
- Validation acceptance criteria will reflect the minimum performance needed for the method to be considered adequate for assessing product quality.

Statistical Approach and Assumptions

- Test result data are normally distributed.
- Standard deviations required for adequate method performance will be defined based on the review of available regulatory sources.
- Based on these “known” distributions, derived from the required standard deviations, probabilities of passing various method validation acceptance criteria will be calculated.

Statistical Approach and Assumptions

- For this presentation, t-tests and F-tests do not apply
  - t-tests are normally applied when the mean and/or variance is not known (is not defined).
  - F-tests are normally applied when the variance is not known (is not defined).
- When the distributions are defined, event probabilities (passing or not passing acceptance criteria) can be calculated directly from those distributions.

Validation Experimental Design

- The purpose of this presentation is not to advocate any particular experimental design for method validation.
- Validation acceptance criteria will be derived from regulatory sources and statistical considerations.
- Statistical considerations are intimately tied to the design of validation experiments.
- Where needed, the assumptions regarding experimental design and how they relate to the acceptance criteria are presented.
- If alternate designs are preferred, this statistical approach should be applied to derive alternate acceptance criteria.

Quantitative Validation Elements

- Precision
  - Repeatability
  - Intermediate Precision
  - Ruggedness (Reproducibility)
- Accuracy
  - Including the derivative elements of filter interference and solution stability
- Linearity
- Limit of Detection/Limit of Quantitation

Precision Definitions

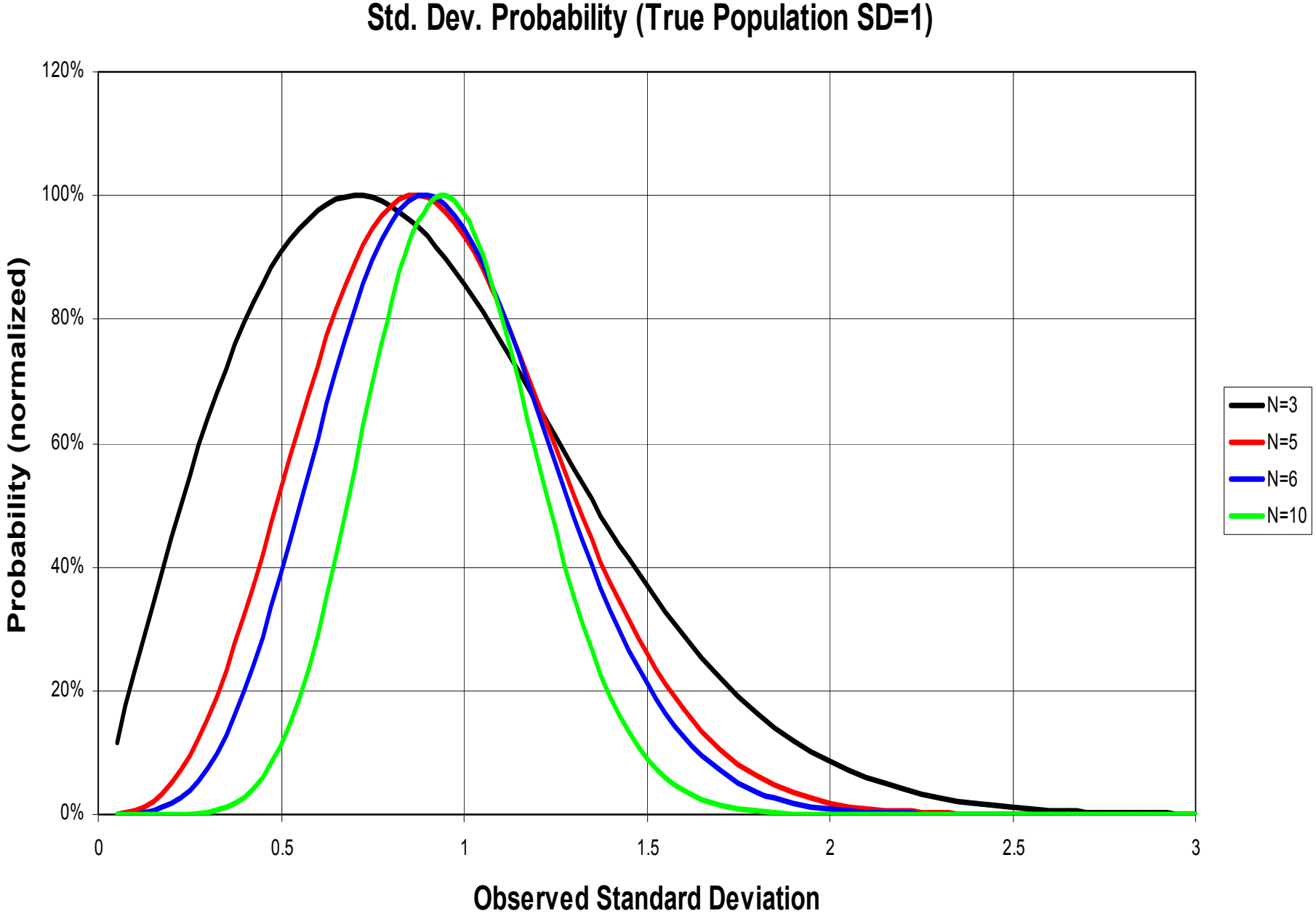
- Repeatability - Variation of repeated measurements of the same sample preparation.
- Intermediate Precision - Variation of repeated determinations (assays-multiple sample preparations) of the same sample.
- Ruggedness (Reproducibility) - Variation of repeated determinations (assays-multiple sample preparations) of the same, or different, sample by multiple instruments/analysts/ standardizations and multiple labs (sometimes).
- These definitions tie in well to the derivations and designs that follow.

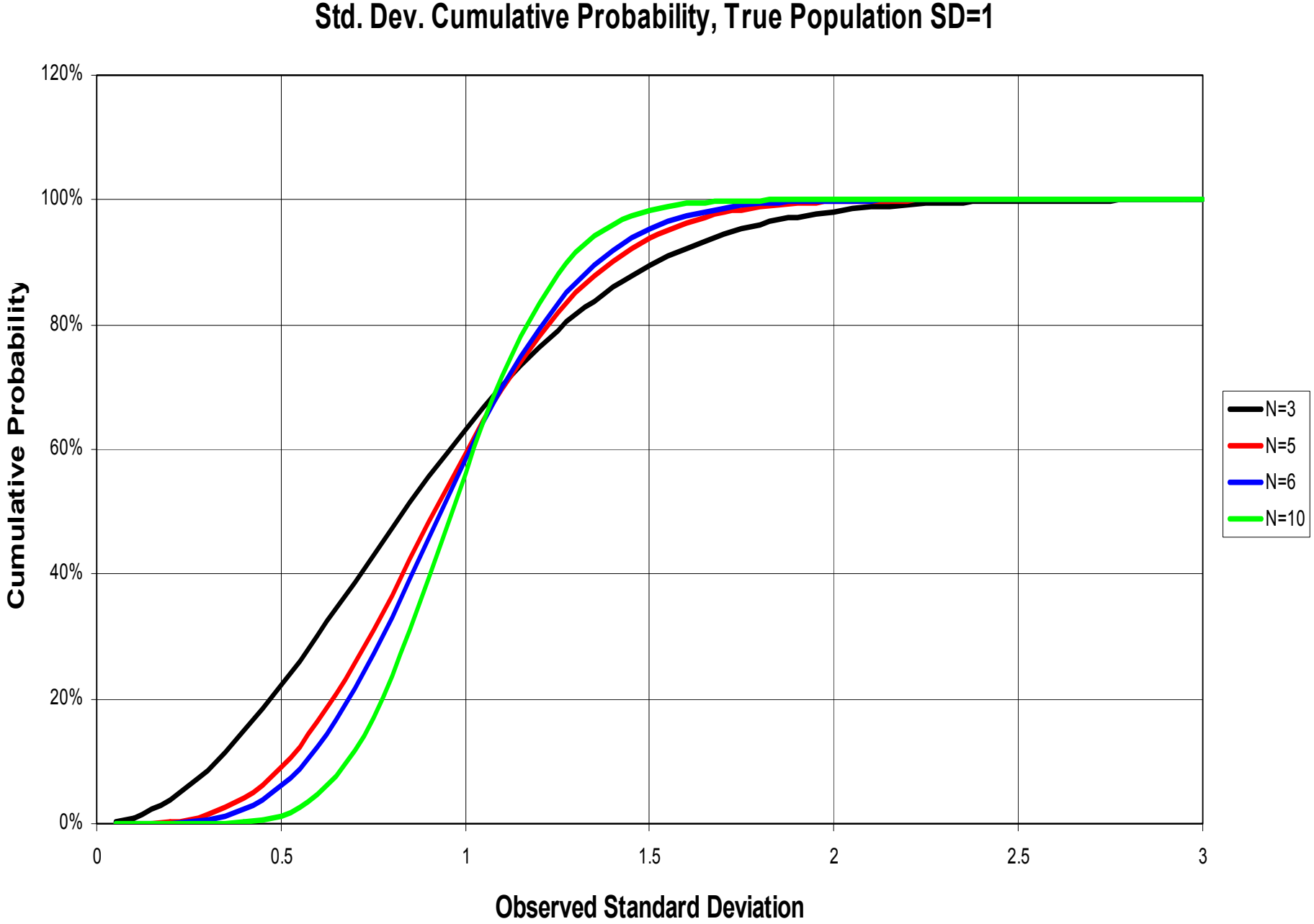
Method Categories (USP <1225>)

- Category I-Assays
- Category II-Determination of Impurities
- Category III-Performance tests (dissolution and drug release).
  - ICH makes no distinction between Category I and Category III. In reality, that can only be applied to the measurement portion of these tests. Other guidances clearly indicate that dissolution sample analysis and the dissolution testing can be treated separately.

Precision (Repeatability)- Determination

- Experimentally obtain multiple measurements from the same preparation and calculate the relative standard deviation (RSD).
- The experimentally determined RSD is an estimate of the true RSD of the population.
- Multiple RSD determinations from a population of data results in a distribution of RSD values.
- The number of measurements taken from the population for each RSD determination dictates the shape of the distribution.


Std. Dev. Probability (True Population SD=1)
120%
100%
80%
60%
40%
20%
0%
Probability (normalized)
0
0.5
1
1.5
2
2.5
3
Observed Standard Deviation
N=3
N=5
N=6
N=10


Std. Dev. Cumulative Probability, True Population SD=1
120%
100%
80%
60%
40%
20%
0%
Cumulative Probability
0
0.5
1
1.5
2
2.5
3
Observed Standard Deviation
N=3
N=5
N=6
N=10

Standard Deviations at Cumulative Probabilities

	N=3	N=5	N=6	N=10
90%	1.52	1.39	1.36	1.28
95%	1.73	1.54	1.49	1.37
99%	2.15	1.82	1.74	1.55

Factor for Required Population SD Based on Acceptance Criteria

	N=3	N=5	N=6	N=10
90%	0.658	0.719	0.735	0.781
95%	0.578	0.649	0.671	0.730
99%	0.465	0.549	0.575	0.645

Category I - Repeatability

- Definitive guidance exists in the form of USP system suitability criteria for this category.
  - Almost universally NMT 2.0%
  - N=5
- Therefore, use NMT 2.0%, N=5 to derive the "minimum acceptable" true repeatability for method validation of category I methods.
- Use tighter specifications where needed.
  - Just because you can do it doesn't mean you have to.

Category I - True Repeatability

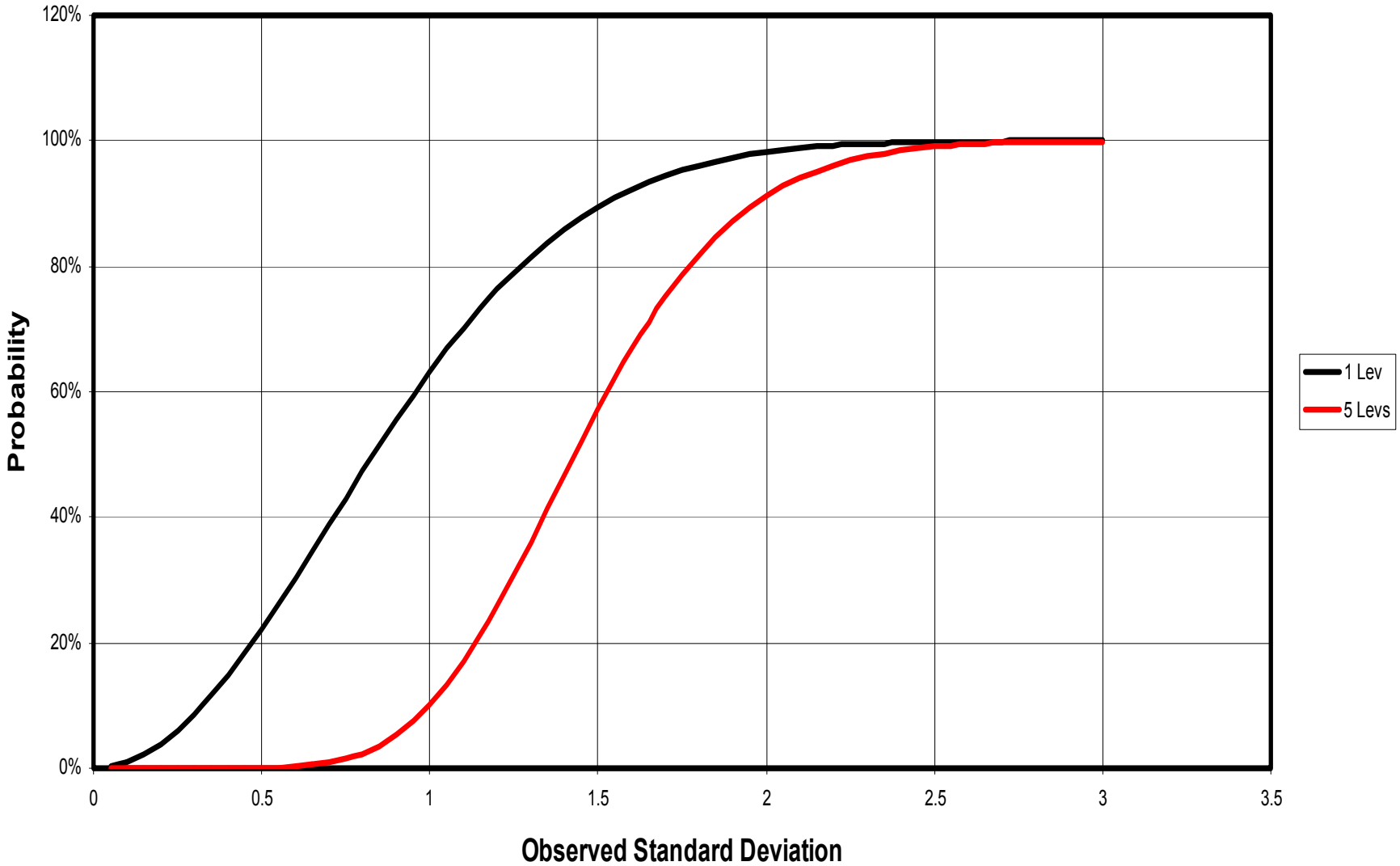
- Based on the acceptance criteria of NMT 2.0% and the cumulative probabilities, what is the true variation of the population?
  - Assume the acceptance criteria is designed for a 95% pass rate.
  - The true variation is 1.3% (2.0% x 0.649=1.3%)
- The true variation is needed to derive other validation acceptance criteria.

Category I - Repeatability Experimental Design

- An argument is often made that better performance is needed than the USP system suitability criteria in order to assure adequate performance in routine use (QC labs).
- Use N=3, RSD NMT 2.0%.
  - Due to the probability distribution of RSD values, this is a tighter specification than the N=5 applied by the USP for system suitability.
  - In this case, the true population RSD would have to be NMT 1.2% to pass 95% of the time.

Category I - Repeatability Experimental Design

- Apply a criteria of N=3, RSD NMT 2.0%, at all the levels used for linearity (typically 5).
  - The RSD criteria must be met at all 5 levels in order to pass.
  - In order to pass this criteria 95% of the time, the population RSD must be even smaller.
  - Based on the analysis of the cumulative probabilities, a true population RSD of 0.9% would be needed.


Cumulative prob-N=3 vs range
120%
100%
80%
60%
40%
20%
0%
Probability
0
0.5
1
1.5
2
2.5
3
3.5
Observed Standard Deviation
1 Lev
5 Levs

Category I - Repeatability Experimental Design

- The FDA's 1994 guidance on Validation of Chromatographic Methods states: "…a minimum of 10 injections with an RSD of ≤1% is recommended."
  - 1.4% rounds to 1%.
  - This acceptance criteria would correspond to a true population RSD of 1.0%

Category I - Repeatability Required Performance

- Based on the analysis of USP system suitability requirements, a method with a true repeatability of 1.3% possesses “adequate performance”.
- Other considerations:
  - QC labs often argue method validation should employ acceptance criteria tighter than the minimum requirements in order to ensure there is a “margin of safety” in the performance of the method.
  - The FDA’s 1994 Method Validation guidance implies a true repeatability of 1.0% is needed.
- Based on all considerations, design validation experiments to verify the true repeatability is about 1.0%, or less.

Sources of Variation in the Determination of Repeatability

- An RSD is the standard deviation "normalized" to the response
- Detection: standard deviation is relatively constant and RSD increases with decreasing concentration.
- Sample measurement and handling: standard deviation is proportional to concentration and RSD is relatively constant.

Sources of Variation in the Determination of Repeatability

- At high concentration, variation from sample measurement and handling predominates.
  - RSD is relatively constant over a reasonable range of concentrations.
  - SD is approximately proportional to concentration.
  - Behavior typically observed for Category I and some Category III methods.

Sources of Variation in the Determination of Repeatability

- As concentration decreases, the primary source of variation transitions from sample measurement and handling to detection.
  - SD transitions from concentration proportional to relatively constant
  - RSD transitions from relatively constant to concentration proportional
  - Behavior frequently observed for Category II methods and some Category III methods.

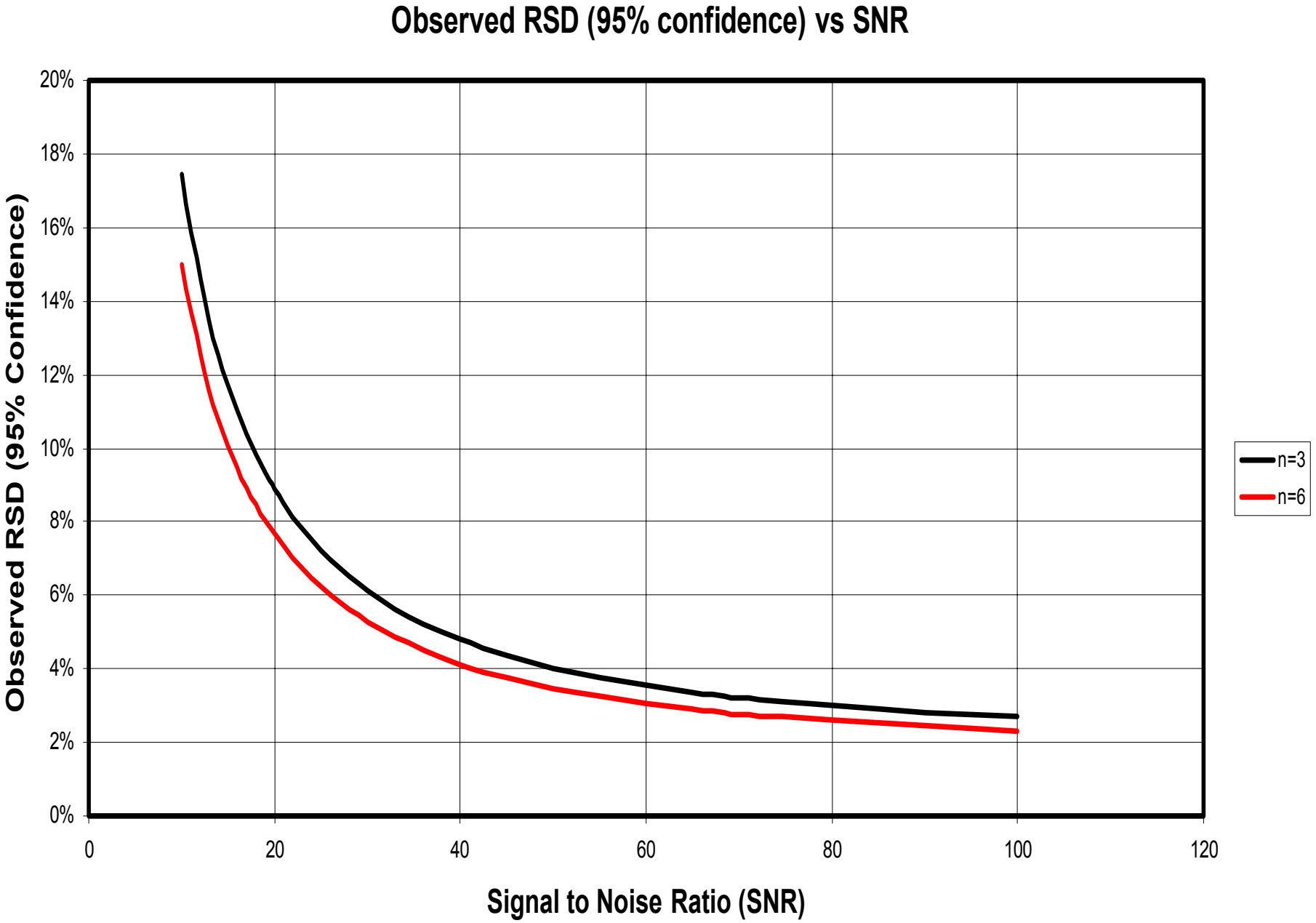
Category II - Repeatability

- The repeatability acceptance criteria for Category I methods assumed that the RSD was independent of concentration. This is not a valid assumption for Category II.
- In Category I methods, variation from measurement and handling predominated.
  - Therefore, the 1.0% RSD for Category I methods is due to measurement and handling, and should be about the same for Category II methods.
- As concentration drops, the contribution of detection variation increases. Overall RSD rises.

Category II - Repeatability

$$RSD = \sqrt{RSD_{Det}^2 + RSD_{Meas}^2}$$

$$RSD = \sqrt{\left(\frac{1}{SNR}\right)^2 + 1.0\%^2}$$


Observed RSD (95% confidence) vs SNR
Observed RSD (95% Confidence)
20%
18%
16%
14%
12%
10%
8%
6%
4%
2%
0%
0
20
40
60
80
100
120
Signal to Noise Ratio (SNR)
n=3
n=6

Category II - Repeatability

- USP guidance:
  - Quantitation Limit (LOQ): "…the lowest amount of analyte in a sample that can be determined with acceptable precision and accuracy…"
  - Signal to noise ratio of 10:1
- A signal to noise ratio of 10:1 corresponds to a true population RSD of 10%.
- FDA guidance, USP, and ICH define LOQ the same way.

Category II - Repeatability

- In effect, all three sources of information (FDA, USP, and ICH) provide consistent guidance on maximum RSD that is considered acceptable.
- Based on Probability distributions for N=6 and a true population RSD of 10%, an acceptance criteria of NMT 15% would be acceptable.

Category II - Repeatability Experimental Design

- An acceptance criteria of NMT 15% for repeatability is consistent with what all the regulatory information dictates.
- The population RSD is level dependent because we are operating near the LOQ.
  - The acceptance criteria is more easily met at higher concentrations.
  - Theoretically, tiered acceptance criteria can be applied, but it adds needless complexity.

Category II - Repeatability Experimental Design

- Use N=3, RSD NMT 15%.
  - Due to the probability distribution of RSD values, this is a tighter specification than the N=6 applied by the USP for system suitability. A true population RSD of 8.7% is needed to pass this criteria.
  - This further tightens the acceptance criteria.
  - These tighter specifications are partially offset by the concentration dependence of the RSD.
  - Can't calculate specific probabilities without considering the range of concentrations studied. Probably 7%-8% depending on range.

Category II - Repeatability True Population RSD

- 15% x 0.578=8.7%
- Assume same concentration independent variability component as category I (1.0%)
- Detection contribution=

  $\text{Sqrt}(8.7\%^2-1.0\%^2)=8.6\%$
- These values are needed for setting acceptance criteria for linearity (Monte Carlo simulation)

Category III - Repeatability

- USP <1225> considers these test differently than ICH
- ICH Makes no distinction between Category I and III.
  - However, the FDA's guidance makes it clear that the dissolution portion and the analytical technique can be considered separately.
- Apply the same repeatability acceptance criteria as Category I? Yes and No.

Category III - Repeatability

- If we have designed our methods properly, Category III methods are "bio-relevant."
- For a given drug, the change in therapeutic effect is approximately proportional to the change in drug delivered.
- Acceptably small variation in the drug delivered corresponds to acceptably small changes in therapeutic effect.

Category III – Repeatability

- Category I methods are allowed a true repeatability of 1.0% and an observed repeatability of 2.0% (described previously).
- This defines an “acceptably small” variation at 100% potency. Per ICH, this level of variation would apply to Category III methods at 100% drug release.
- This level of absolute variability would be acceptable at all levels in our bio-relevant Category III method.

Category III – Repeatability Experimental Design

- N=3, 5 levels of linearity data apply criteria to all levels.
  - Determine the response at the 100% level.
  - Calculate 2.0% of the 100% level and apply this value as the standard deviation acceptance criteria.
  - Determine the standard deviation at all levels
  - The observed standard deviation must be less than, or equal, to the acceptance criteria.

Intermediate Precision

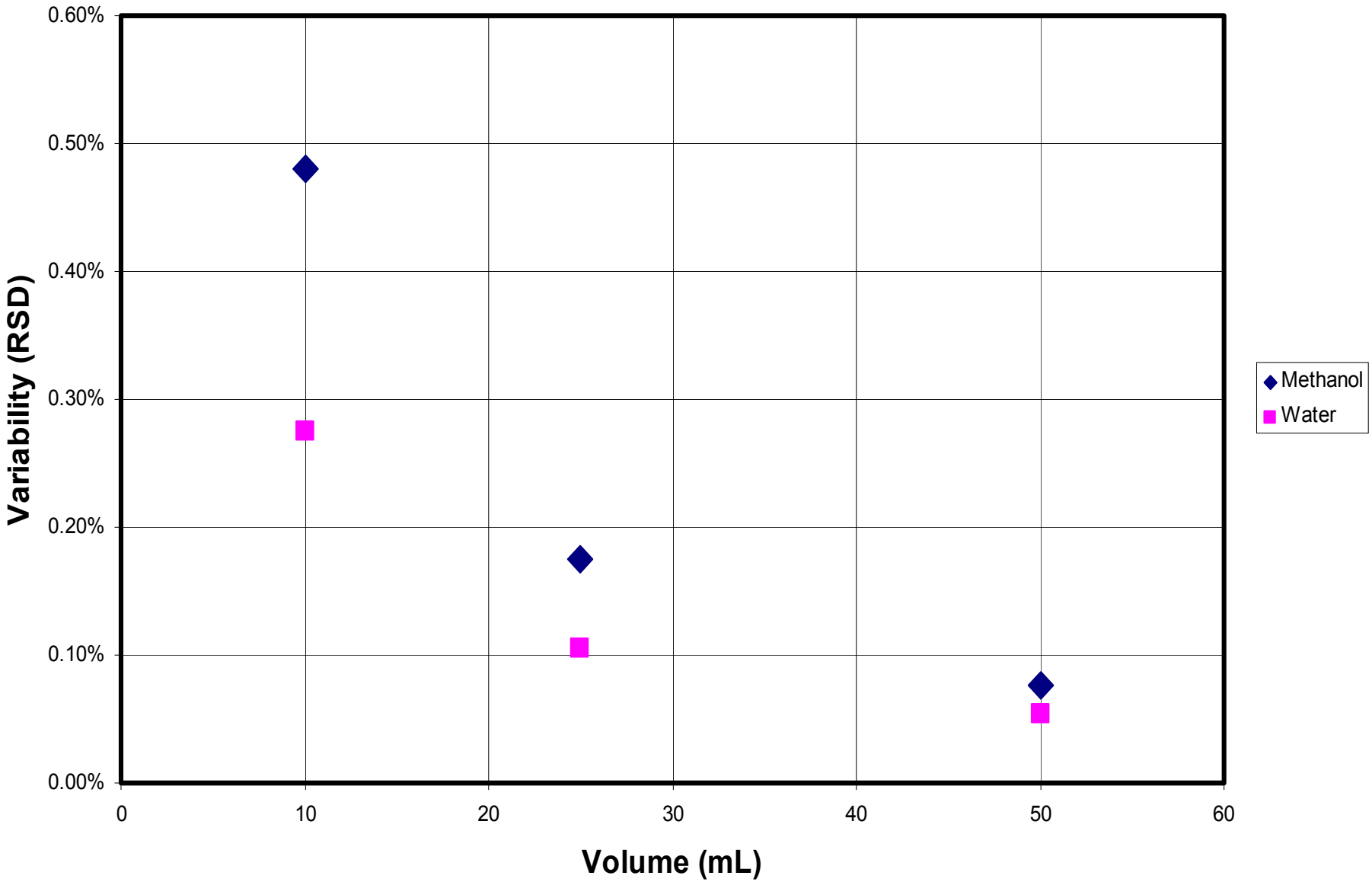
- Variability contributions from other aspects of the assay (besides repeatability)
  - Weighing of standards and samples
  - Dilutions
  - Sample variability

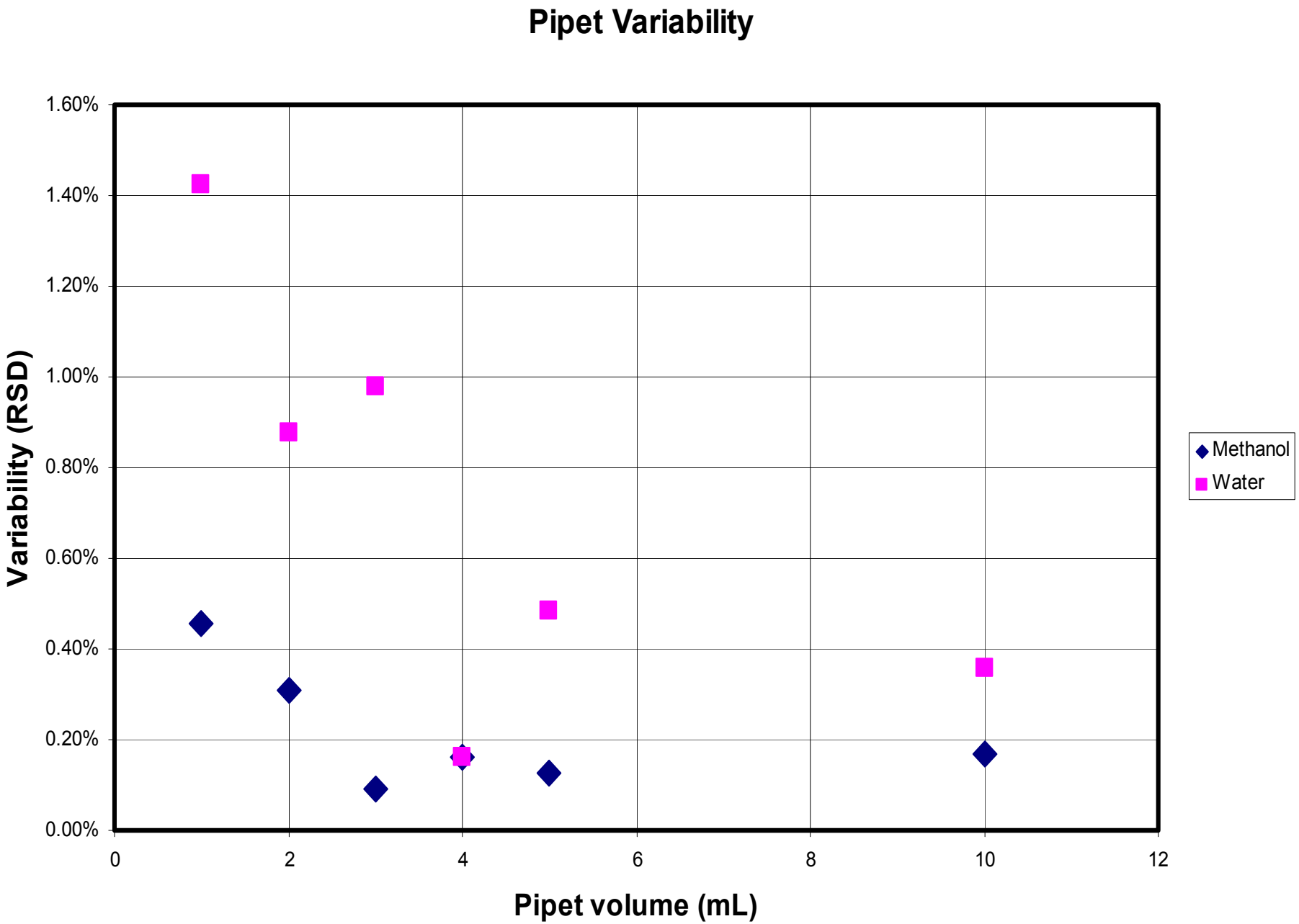
Intermediate Precision Weighing Variability

- Typically 0.1% to 0.2% for Categories I and III, depending on the amount weighed. Typically 0.5% to 1.0% for the standard and 0.1% to 0.2% for the sample for Category II.
- Two contributions, one for the standard, one for the sample.
- Very minor contributor to overall variability.
- Consideration of standard weighing variability may be omitted for intermediate precision (typically only one standard preparation is used).

Intermediate Precision Dilution Variability

- Significant contributions from glassware.
  - Typically 3 dilutions considering both the sample and the standard.
  - Sometimes less dilutions for Category I.
  - Significantly less dilutions for Category I raw material assays.
- Dissolution vessel fill volume variability for category III. The USP allows ±1%.


Flask Variability
Variability (RSD)
0.60%
0.50%
0.40%
0.30%
0.20%
0.10%
0.00%
0
10
20
30
40
50
60
Volume (mL)
Methanol
Water


Pipet Variability
Variability (RSD)
1.60%
1.40%
1.20%
1.00%
0.80%
0.60%
0.40%
0.20%
0.00%
0
2
4
6
8
10
12
Pipet volume (mL)
Methanol
Water

Intermediate Precision Sample Variability

- Raw materials - no variability, they are considered homogeneous
- Products - dosage unit variability is acceptable
  - Stage 1 CU criteria is RSD NMT 6.0%.
  - Part of the 6.0% is the method's intermediate precision.
  - Must determine intermediate precision of Category I assays of single dosage units, determine what portion of the 6.0% is due to the product, and then determine what the intermediate precision of the product assay must be.
- Product variability contributes variability to the three method categories in different ways.

Intermediate Precision Category I, CU determination

- Product variability not considered in the sample preparation.
- This derivation is to partition sources of variance.
- Sources of variation:
  - Repeatability - 1.0%
  - Standard weighing - 0.1%
  - Dilutions - 0.6% (two stock solutions, one dilution of each)
  - Total analytical variability - 1.2%

Intermediate Precision Category I, CU determination

- USP specifications - CU RSD NMT 6.0%
  - Product designed to meet this 95% of the time.
  - N=10, true RSD=4.4% (6.0% x 0.73), or less
  - Maximum allowable true product variability- 4.2% (square root($4.4\%^2$-$1.2\%^2$))
- Product variability contribution to assay variability
  - N=10, 1.3% (4.2% / square root of 10)
  - N=20, 0.9% (4.2% / square root of 20)

Intermediate Precision Category I

Source of Variability	Raw Material	Product
Dilutions	0.14%	0.6%
Sample	None	0.9%
Sample weighing	0.1%	0.1%
Standard weighing	0.1%	0.1%
Repeatability	1.3%	1.3%
Total	1.3%	1.7%

Intermediate Precision Category I

- Apply an experimental design of N=6, single determination of each preparation.
- Multiply the population variation (all sources) by 1.49 to determine the acceptance criteria that can be met 95% of the time.
- Intermediate precision acceptance criteria
  - Raw material assays: NMT 1.9%≈2.0%
  - Product assays: NMT 2.5%

Intermediate Precision Category II

Source of Variability	Variation
Dilutions	0.7%
Sample	0.9%
Sample weighing	0.1%
Standard weighing	0.5%
Repeatability	10%
Total	10.1%

Intermediate Precision Category II

- Apply an experimental design of N=6.
- Will need to spike to the “decision level.”
- Multiply the population variation (all sources) by 1.49 to determine the acceptance criteria that can be met 95% of the time.
- Intermediate precision criteria=15%.

Intermediate Precision Category III

Source of Variability	Variation
Dilutions	0.6%
Product variability	4.2%
Standard weighing	0.1%
Repeatability*	1.3%
Vessel fill variability	1%
Performance variability	1%
Total	4.7%

Intermediate Precision Category III

- Analyze 6 dosage units in the dissolution instrument.
- * The repeatability will need to be converted to standard deviation at 100% and scaled to the percent released at the decision level (same principle was applied to repeatability).
- Multiply the population variation (all sources) by 1.49 to determine the acceptance criteria that can be met 95% of the time.
- At Q of 75%, this would be 7.2%, 100% dissolved, this would be 7.0%.

Ruggedness

- Should be about the same as intermediate precision.
- The only theoretical addition to variability is multiple calibrations, which are a minor contributor to total variability
- Labs, analysts, and instruments can add variability
  - The regulatory definition of ruggedness implies that these factors must not contribute significantly
  - Don’t consider these factors as contributors to variability (and acceptance criteria) without suitable assessment of the impact to a methods ability to assess product quality.

Precision Summary

	Repeat	Int. Prec.	Rugged.
Cat. I, Raw Material	2.0%	2.0%	2.0%
Cat. I, Product	2.0%	2.5%	2.5%
Category II	15%	15%	15%
Category III	2.0%	7.0%*	7.0%*

*Decision level and product variability specific

Historical Perspectives on Precision

- During the 1970's, dissolution testing and modern system suitability were under development.
- During this time period, large amounts of testing and study was carried out to characterize the performance of dissolution testing and standardize the apparatus.
- In the late 1970's and early 1980's, during the development of dissolution calibrators, PMA-QCS collaborative studies compiled enough data to conclude that a multi-laboratory precision (ruggedness) of 6% to 8% was acceptable performance.

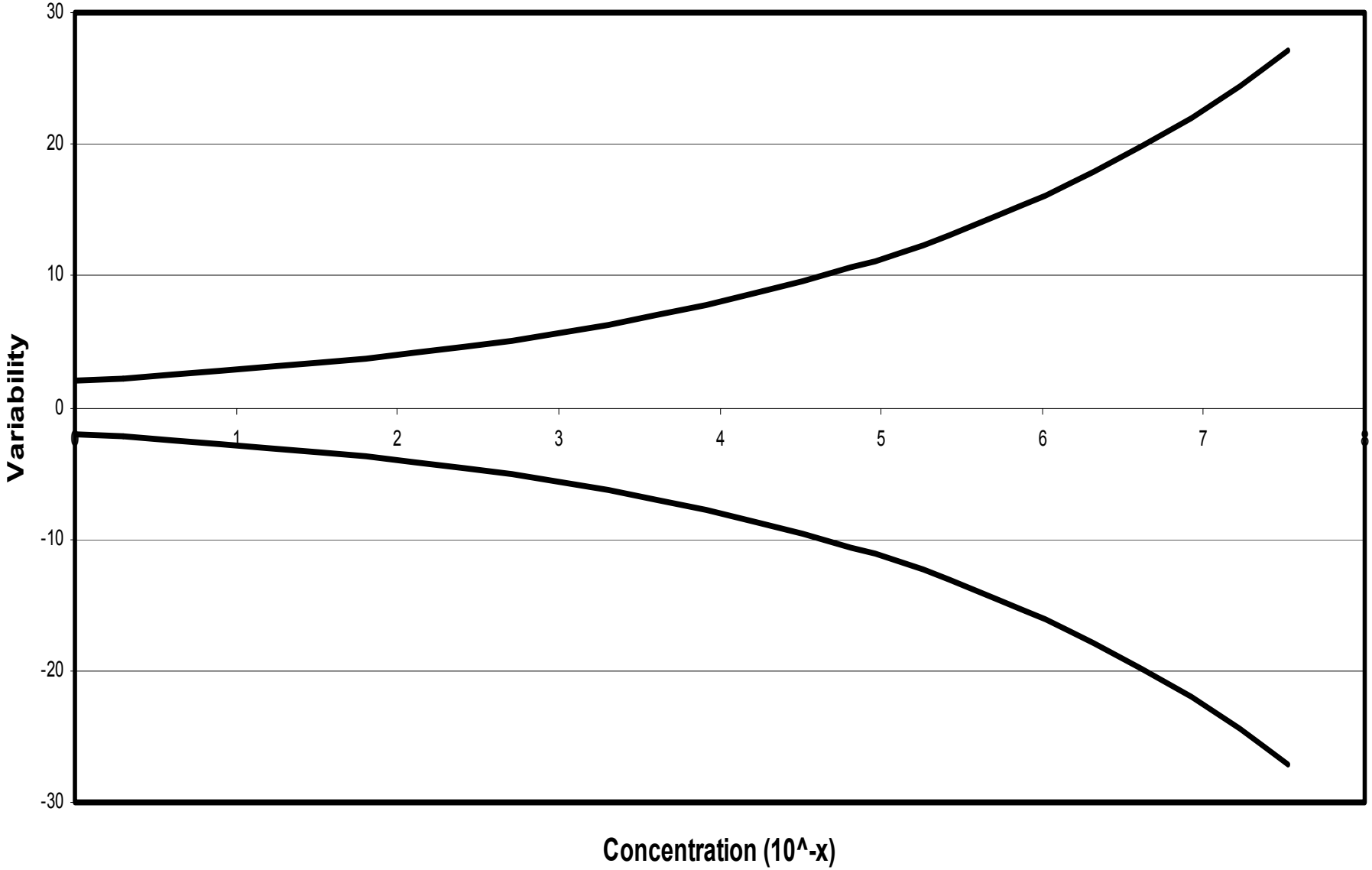
Historical Perspectives on Precision

- In 1982, William Horwitz of the FDA published a review of inter-laboratory variability.
  - Examined about 150 multi-laboratory studies
  - Concluded the ruggedness was related to the sample concentration (w/w in the sample)
- The data could be empirically described by: $RSD=2^{(1-0.5\log C)}$
- Repeatability is typically half this value
- This relationship is commonly referred to as the "Horwitz curve."

Historical Perspectives on Precision

- William Horwitz and Richard Albert (both of the FDA) published on this subject again in 1997
  - By 1997, almost 10,000 inter-laboratory studies had been examined.
  - All the data supported the Horwitz curve and confirmed its validity in analytical chemistry.
  - This publication also included a derivation, from first principles, of the mathematical equation underlying the Horwitz curve.
- The Horwitz curve provides guidance on what level of performance is acceptable.

Horwitz Curve



Horwitz Curve and Pharmaceuticals

	C	Rugg.	Repeat.
Raw material	1	2	1
High dose product	0.9	2	1
Low dose product	0.0001	8	4
Typical modern product	0.05	3	1.5
Impurity in high dose	0.0009	6	3
Impurity in low dose	1E-07	23	12
Impurity modern product	0.00005	9	5

Accuracy

- FDA's 1994 guidance on the Validation of Chromatographic Methods.
  - "The measure of how close the experimental value is to the true value".
  - "Full recovery should be obtained for the compound(s) of interest".
- Result = true value ± bias
- Determination of accuracy is confounded by variability:
  - Accuracy acceptance criteria include considerations of both bias and analytical variability.
  - Result = true value ± bias + variability

Accuracy

- FDA's 1994 guidance on the Validation of Chromatographic Methods.
  - Recommends: "Recovery data, at least in triplicate, at each level (80, 100, and 120% of label claim) is recommended. The mean is an estimate of accuracy and the RSD is an estimate of sample analysis precision".
  - USP <1225> recommends something similar, but is less specific.

Accuracy: Experimental Considerations

- Prepare 3 preparations at 80, 100, and 120% levels, and analyze them
- The average value at each level is the accuracy (recovery). Apply a statistically sound acceptance criteria to this average.
- Don't use the RSD of the individual assays. These are contrived samples and are not the best representation of the method's intermediate precision. Use the experimental design described in the precision section.

Accuracy: Experimental Considerations

- Unlike precision studies, the purpose of an accuracy study is to generate "reportable" results for the assessment of accuracy.
- The concepts of what is a reportable result in the FDA's 1998 guidance on Investigating Out of Specification Test Results should be considered in both your experimental design and acceptance criteria.

Accuracy: Experimental Considerations

- The preparations used for accuracy are not necessarily homogeneous. Therefore, each preparation assay is necessarily a result and an acceptance criteria for each one should also be applied (in the absence of applying an acceptance criteria the individual injections), or an acceptance criteria for variability (RSD) should be applied.

Accuracy: Experimental Considerations

- Each solution preparation for analysis is homogeneous. Individual measurements (injections) may be pooled (averaged).
- However, you must have documented control mechanisms (in SOPs or protocols) to prevent excessive variability of these measurements.
  - Example-there must be controls that prevent individual injection recovery results of 90% and 110% from being reported as an assay value of 100%, without other action.
  - In the absence of these controls, acceptance criteria for each measurement (injection) would be required.

Accuracy: Acceptance Criteria

- Category I - Raw Material Assays
  - For specific assays where analysis and calibration are the same (chromatography), no bias is allowed.
  - For titrametric assays, most compendial assay specifications are unsymetric. Presumably this is an allowance for bias and appears to typically be 0.5%. In addition, ICH Q2B indicates that when non-specific assays are used (titration is given as an example), they must be used in combination with other supporting analytical procedures (impurity determination). This implies a certain amount of bias is acceptable.

Accuracy: Acceptance Criteria

- Category I - Raw Material Assay (Continued)
  - Three assay preparations, duplicate injections of each (to give an assay result).
  - Individual value (injection) variability=1.4% $\text{Sqrt}(1.3\%^2+0.6\%^2)$
  - Assay result variability=1.1% $\text{Sqrt}((1.3\%/\sqrt{2})^2 + 0.6\%^2)$
  - Average assay (recovery) variability=0.6% $1.1\%/\sqrt{3}$
  - Multiply by 2 to get acceptance criteria

Accuracy: Acceptance Criteria

- Category I - Raw Material Assay (Continued)
  - Assay result from individual injection
    True value ± 2.8% (tighten to 2.5%)
  - Assay result from each preparation
    True value ± 2.2% (tighten to 2.0%)
  - Average assay result
    True value ± 1.2% (tighten to 1.0%)

Accuracy: Acceptance Criteria

- **Category I and III product assays**
  - **Treat the same as Category I, RM assay, and allow for a ± 1% bias**
  - **Assay result from individual injection True value ± 3.8% (tighten to 3.5%)**
  - **Assay result from each preparation True value ± 3.2% (tighten to 3.0%)**
  - **Average assay result True value ± 2.2% (tighten to 2.0%)**

Accuracy: Acceptance Criteria

- Category II
  - Similar statistical treatment to category I, repeatability of 10%, dilution/preparation variation 1%, and a bias of ± 5%
  - Assay result from individual injection
    True value ± 25%
  - Assay result from each preparation
    True value ± 20%
  - Average assay result
    True value ± 13%

Solution Stability

- Use the same acceptance criteria as accuracy.
- Consider regressing the solution injection peak areas versus age.
  - This eliminates day to day calibration variability.
  - Recalculate the acceptance criteria without the preparation and handling variability. They won't change very much because detection variability is a major contributor to the acceptance criteria.
  - Special suitability solutions designed to demonstrate detector stability may be needed (proof that any observed change is not detector drift).

Filter Qualification

- The first filter would necessarily be part of the other validation studies and no separate study would be needed.
  - If a filter was used for accuracy, precision, linearity, LOD, and LOQ, it is obviously suitable for the method if the acceptance criteria were met.
- For other filters, apply the accuracy acceptance criteria and experimental design.
  - For category II methods, consider repeating linearity (and use the same acceptance criteria) if adsorption might be a concern.

Limit of Detection (LOD) and Limit of Quantitation (LOQ)

- Available guidances are very clear about how to determine the LOQ and LOD.
- Not applicable for categories I and III.
- Surprisingly, none of them make any mention of what is an actual acceptable LOQ level (concentration level).
  - LOQ should be the ICH/FDA reporting limit for impurities, or lower.
  - I recommend 50% of the reporting limit, where possible.
  - Actual level of LOD is not relevant for quantitative methods. Determine it and report for information only.
  - Use sensitivity test solutions as part of system suitability in your methods. LOD and LOQ are instrument specific.

Linearity

- USP <1225>
  - Its ability to elicit test results that are directly, or by a well-defined mathematical transformation, proportional to the concentration of analyte in samples within a given range
- Required because the vast majority of methods use single level calibration.
  - Therefore, there must be an algebraic relationship between the standard and sample concentrations and their respective responses
  - Beer's Law is assumed to apply. Linearity provides proof.

Linearity

- Ideal behavior: y=mx+b
  - Response=Sensitivity x Concentration + Bias
- Real behavior:
  - $y \pm \sigma_y = m \times (C \pm \sigma_C) + f(x)$
  - $y \pm (\sigma_y + m \times \sigma_C) = m \times C + f(x)$
  - σ_y = Detection noise
  - σ_C = Sample preparation and handling variation
  - $f(x)$ = bias and non-linear behavior

Linearity

- σ_y, σ_C, and *f*(x) contribute to imperfect correlation
  - F(x) is a systematic contributor to correlation. Correlation will be less than one, but will not vary from experiment to experiment due to this factor.
  - σ_y and σ_C are variability contributors to correlation. Correlation will be less than one and will vary about a true value from experiment to experiment.
- For method validation purposes, *f*(x) is considered to be zero.

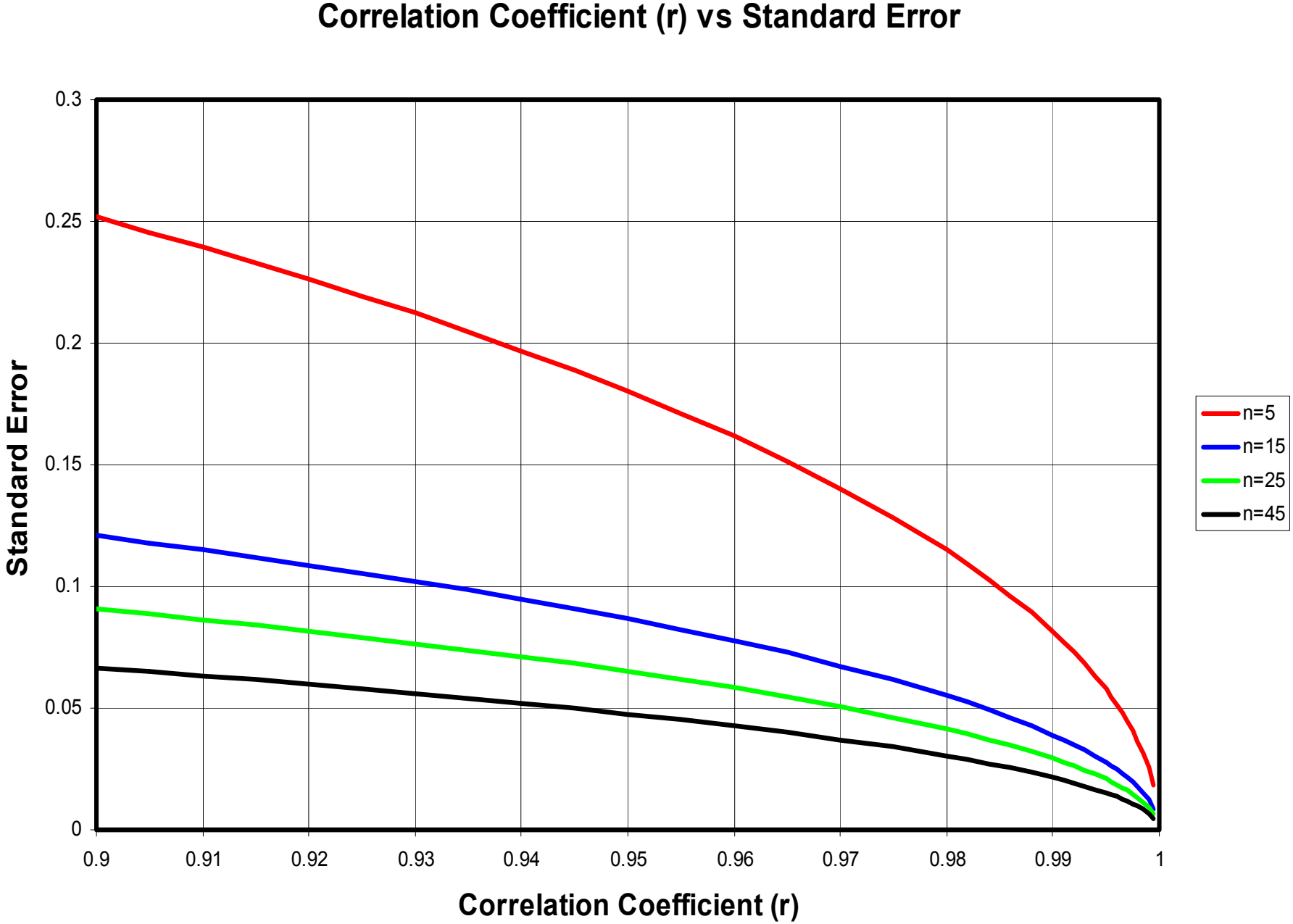
Linearity Determination

- Experimentally obtain measurements from multiple concentration levels (typically 5) and calculate the correlation coefficient (r).
- The experimentally determined correlation coefficeint is an estimate of the true r of the population.
- Multiple correlation coefficient determinations from a population of data results in a distribution of r values.
- The statistical significance of r is related to the value of r.
- A variety of factors influence the distribution of the correlation coefficient.

Correlation Coefficient

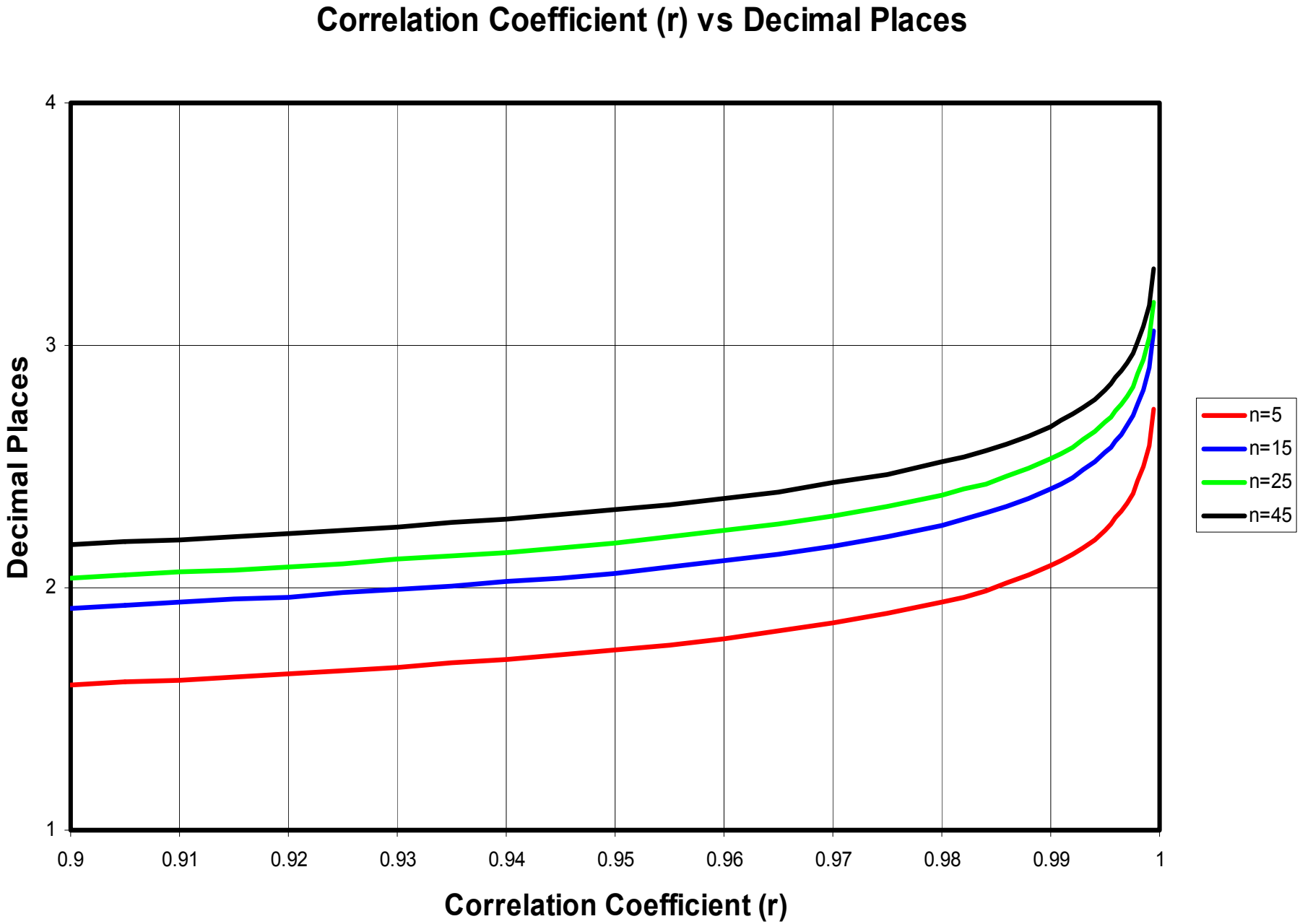
- The correlation coefficient is “unitless”.
- The standard error of the correlation coefficient is related to r and is independent of the underlying data set.

$$s_r = \sqrt{\frac{1 - r^2}{n - 2}}$$


Correlation Coefficient (r) vs Standard Error
Standard Error
0.3
0.25
0.2
0.15
0.1
0.05
0
0.9
0.91
0.92
0.93
0.94
0.95
0.96
0.97
0.98
0.99
1
Correlation Coefficient (r)
n=5
n=15
n=25
n=45

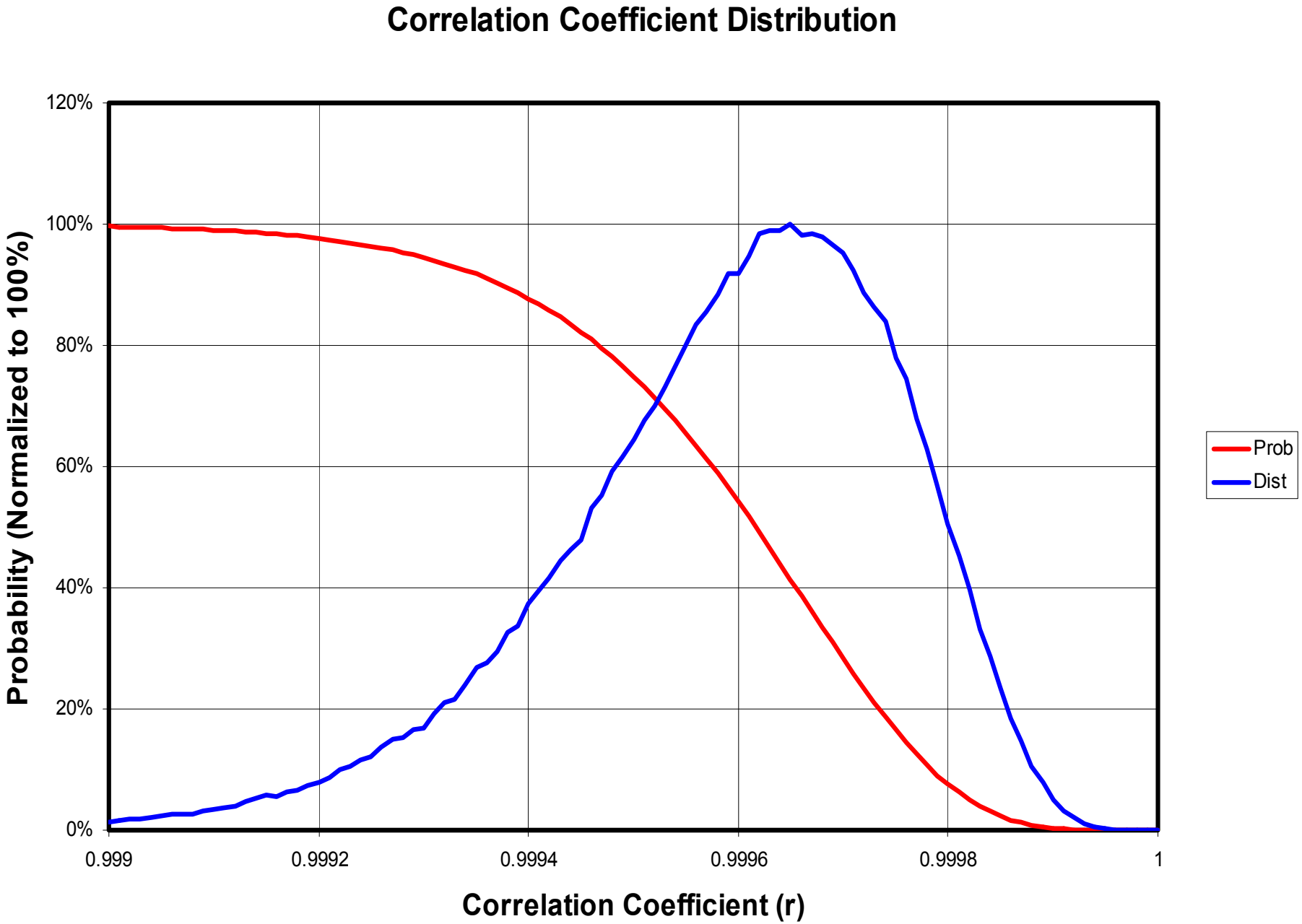
Correlation Coefficient

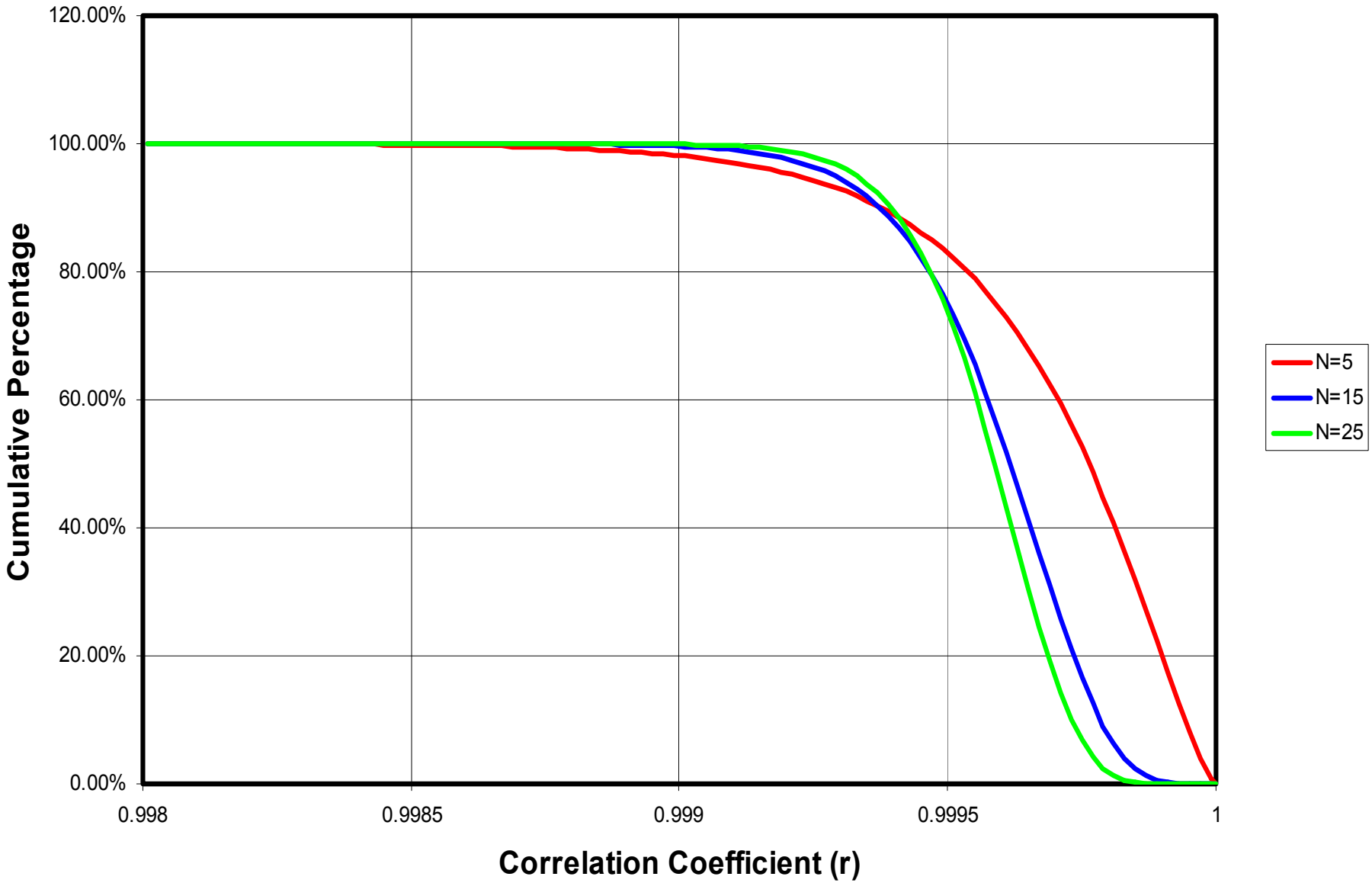
- The standard error calculation assumes the distribution of correlation coefficient values is Gaussian (it's not, but it is close).
- About the only thing this value is useful for is determining how many significant figures of the correlation coefficient are statistically meaningful.
- The decimal place at which the second significant digit of s_r is located may be noted and r may be expressed rounded off to that decimal place.

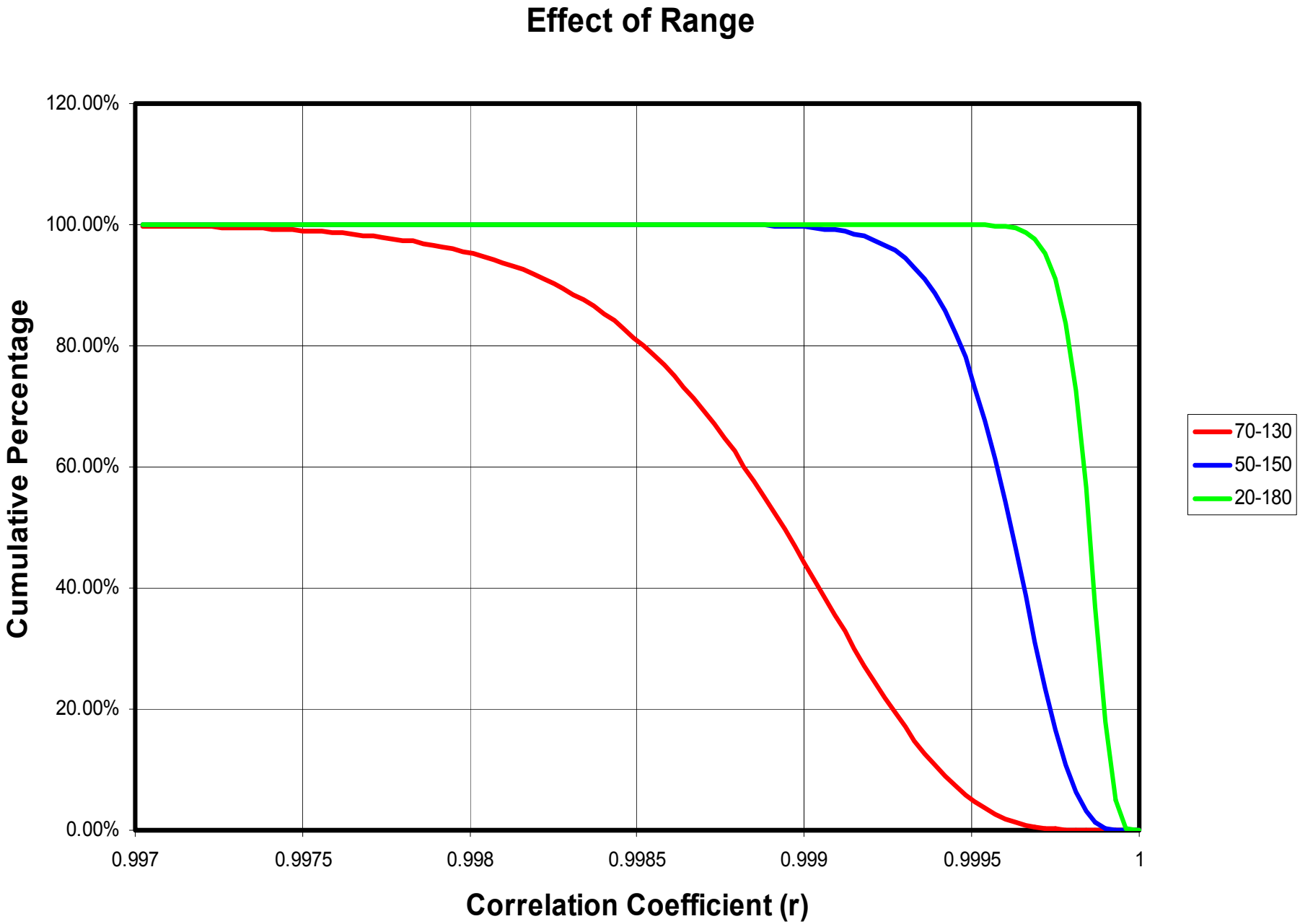

Correlation Coefficient (r) vs Decimal Places
Decimal Places
4
3
2
1
0.9
0.91
0.92
0.93
0.94
0.95
0.96
0.97
0.98
0.99
1
Correlation Coefficient (r)
n=5
n=15
n=25
n=45

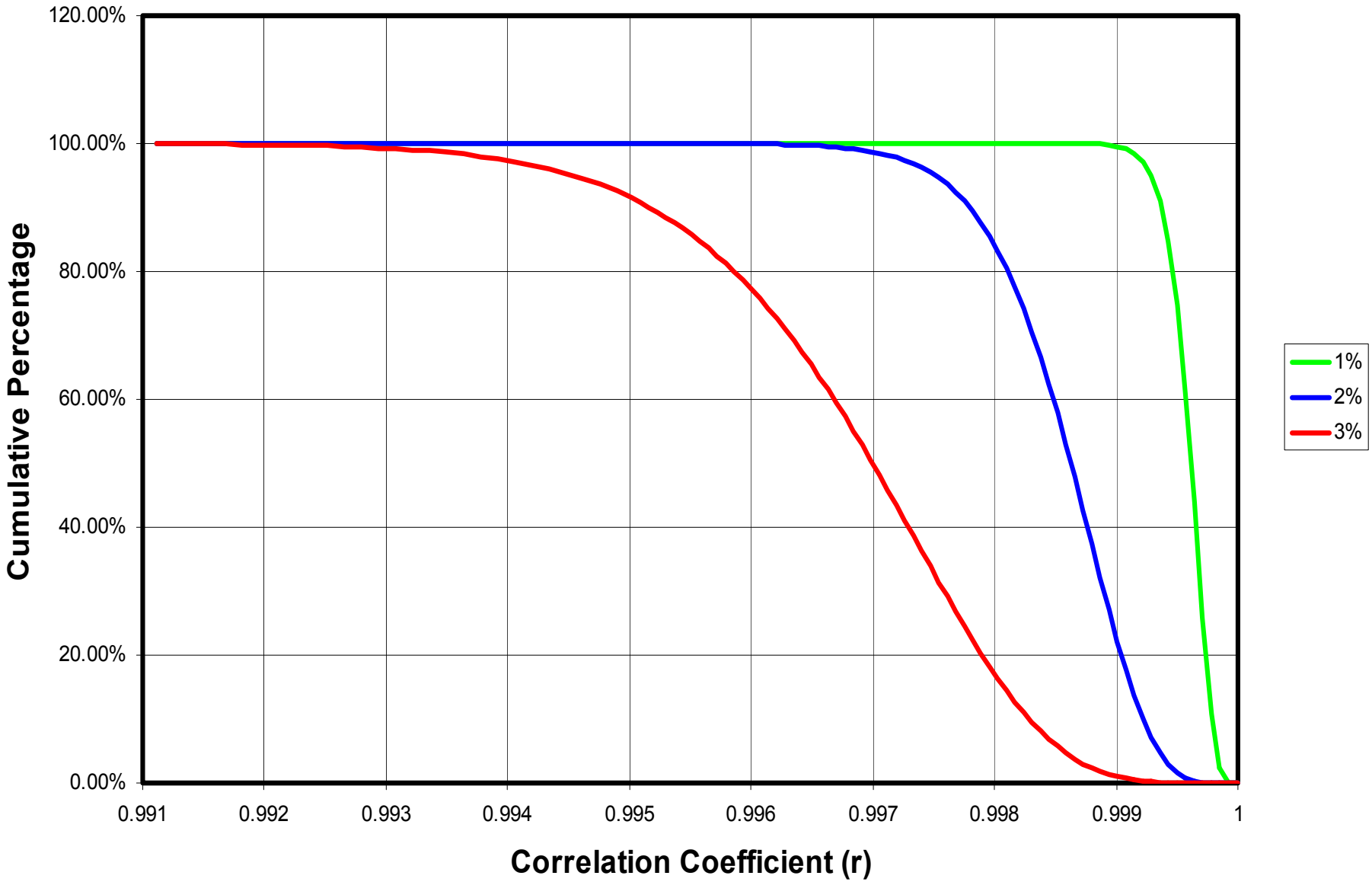
Calculation of Correlation Coefficient Distribution

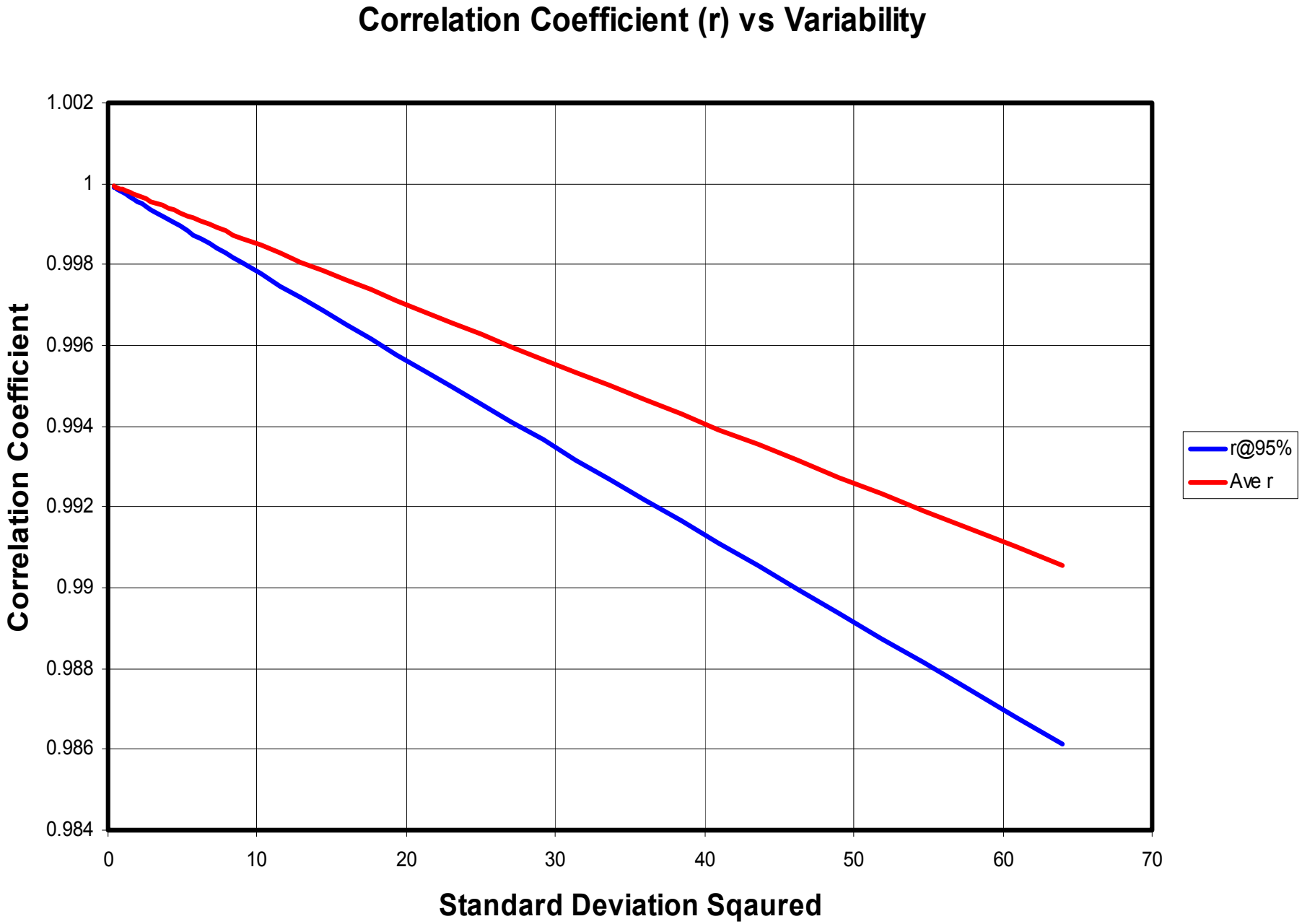
- Utilize a selected range of x values, number of data points, σ_y, and σ_C.
- Calculate distribution of r values using Monte Carlo simulation ~
  - About 350,000 linear regressions per plot
  - About 7,000,000 Gaussian distributed random units and floating operations.
- Example: N=15, 50% to 150%, σ_y=1%, σ_C=0.5%


Correlation Coefficient Distribution
120%
100%
80%
60%
40%
20%
0%
Probability (Normalized to 100%)
0.999
0.9992
0.9994
0.9996
0.9998
1
Correlation Coefficient (r)
Prob
Dist


Effect of the Number of Data Points
120.00%
100.00%
80.00%
60.00%
40.00%
20.00%
0.00%
Cumulative Percentage
0.998
0.9985
0.999
0.9995
1
Correlation Coefficient (r)
N=5
N=15
N=25


Effect of Range
120.00%
100.00%
80.00%
60.00%
40.00%
20.00%
0.00%
Cumulative Percentage
0.997
0.9975
0.998
0.9985
0.999
0.9995
1
Correlation Coefficient (r)
70-130
50-150
20-180


Effect of Measurement Precision
120.00%
100.00%
80.00%
60.00%
40.00%
20.00%
0.00%
Cumulative Percentage
0.991
0.992
0.993
0.994
0.995
0.996
0.997
0.998
0.999
1
Correlation Coefficient (r)
1%
2%
3%

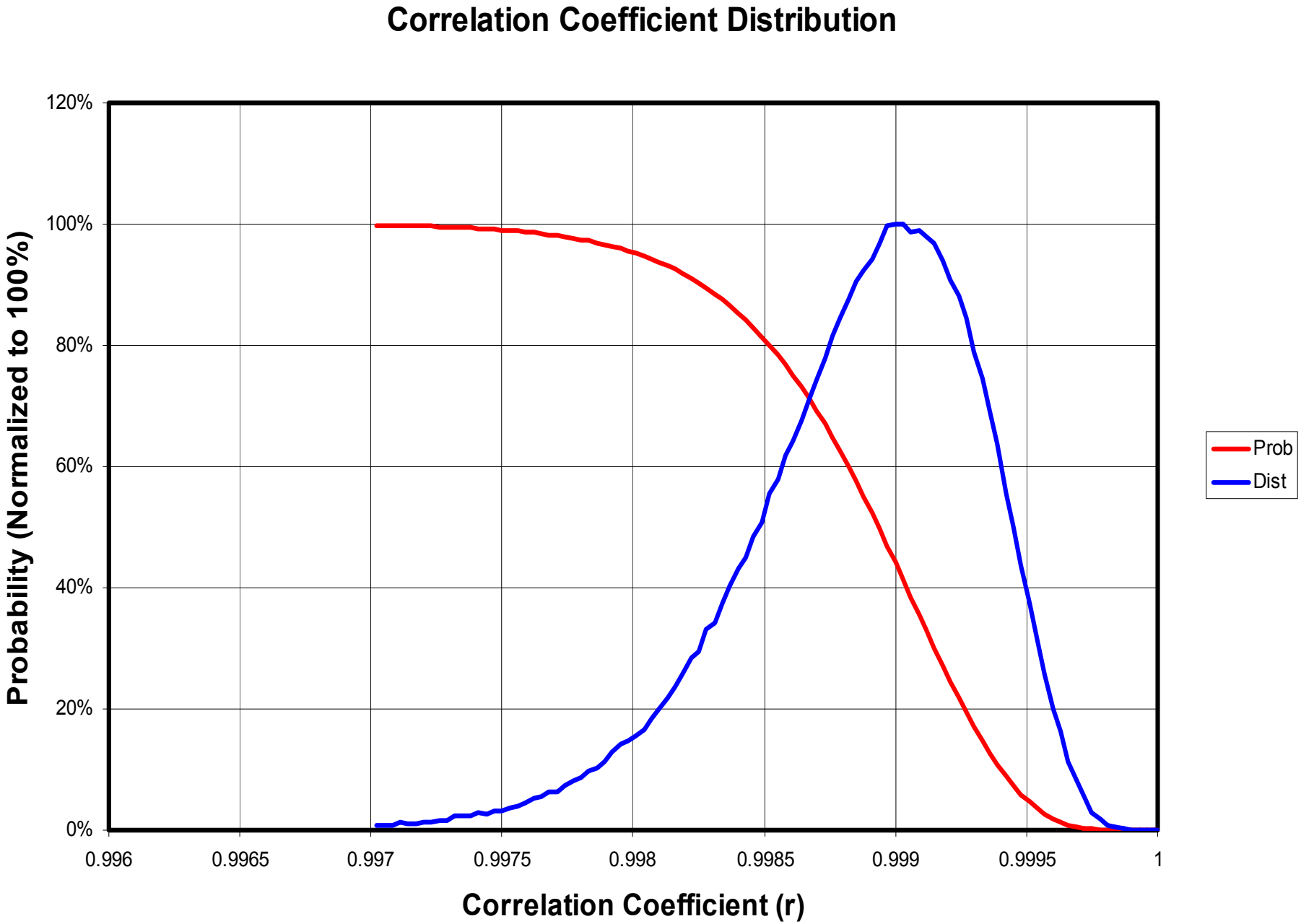

Correlation Coefficient (r) vs Variability
Correlation Coefficient
1.002
1
0.998
0.996
0.994
0.992
0.99
0.988
0.986
0.984
0
10
20
30
40
50
60
70
Standard Deviation Sqaured
r@95%
Ave r

Regression Acceptance Criteria

- Measurement precision, range of values, and number of data points significantly affect the correlation coefficient
- Therefore, a full experimental design must be defined and the matching correlation coefficient acceptance criteria must be calculated
- The correlation coefficient acceptance criteria is unique to each design. Don't change the design without considering the impact on the acceptance criteria.

Experimental Design

- Category I, raw material
  - 5 Levels, 3 determinations per level (N=15)
  - Measurement RSD=1.0%
  - Linearity preparation variability=0.5%
  - Range of values=70%, 85%, 100%, 115%, 130%
  - Acceptance criteria for 95% pass rate=0.998


Correlation Coefficient Distribution
120%
100%
80%
60%
40%
20%
0%
Probability (Normalized to 100%)
0.996
0.9965
0.997
0.9975
0.998
0.9985
0.999
0.9995
1
Correlation Coefficient (r)
Prob
Dist

Linearity Experimental Design

- Category I, Product
  - 5 Levels, 3 determinations per level (N=15)
  - Measurement RSD=1.0%
  - Linearity preparation variability=0.5%
  - Range of values=50%, 75%, 100%, 125%, 150%
  - Acceptance criteria for 95% pass rate=0.999

Linearity Experimental Design

- Category II
  - 7 Levels, 3 determinations per level (N=21)
  - Measurement RSD=8.6%
  - Linearity preparation variability=1.0%
  - Range of values=10%, 25%, 50%, 75%, 100%, 125%, 150%
  - Acceptance criteria for 95% pass rate=0.98

Linearity Experimental Design

- Category III
  - 5 Levels, 3 determinations per level (N=15)
  - Measurement RSD=1.0%
  - Linearity preparation variability=0.5%
  - Range of values=40%, 60%, 80%, 100%, 120%
  - Acceptance criteria for 95% pass rate=0.999

Linearity and Intercept

- Under normal circumstances, an acceptance criteria is not needed for the intercept.
  - The intercept can influence accuracy.
  - Accuracy is verified directly in the range of interest so the intercept, while statistically interesting, is not needed to support acceptability of the method.

Linearity and Intercept

- ICH, FDA, and USP indicate that accuracy may be “inferred” once precision, linearity, and specificity have been established.
  - To use this approach, an acceptance criteria for the intercept is required
  - Apply an acceptance criteria of “the 95% confidence interval of the intercept must include the origin”.
  - If origin is outside the intercept confidence interval, you may or may not be able to use the data to support accuracy.
  - This approach can be used to claim accuracy outside the range of values for which accuracy was directly confirmed.

Statistical Error

- Whenever you deal with probability distributions and a limited number of samples drawn from the population, there is a probability of drawing the wrong conclusion.
- The probabilities are dictated by the distribution and our definition of what is acceptable and not acceptable.

Statistical Error

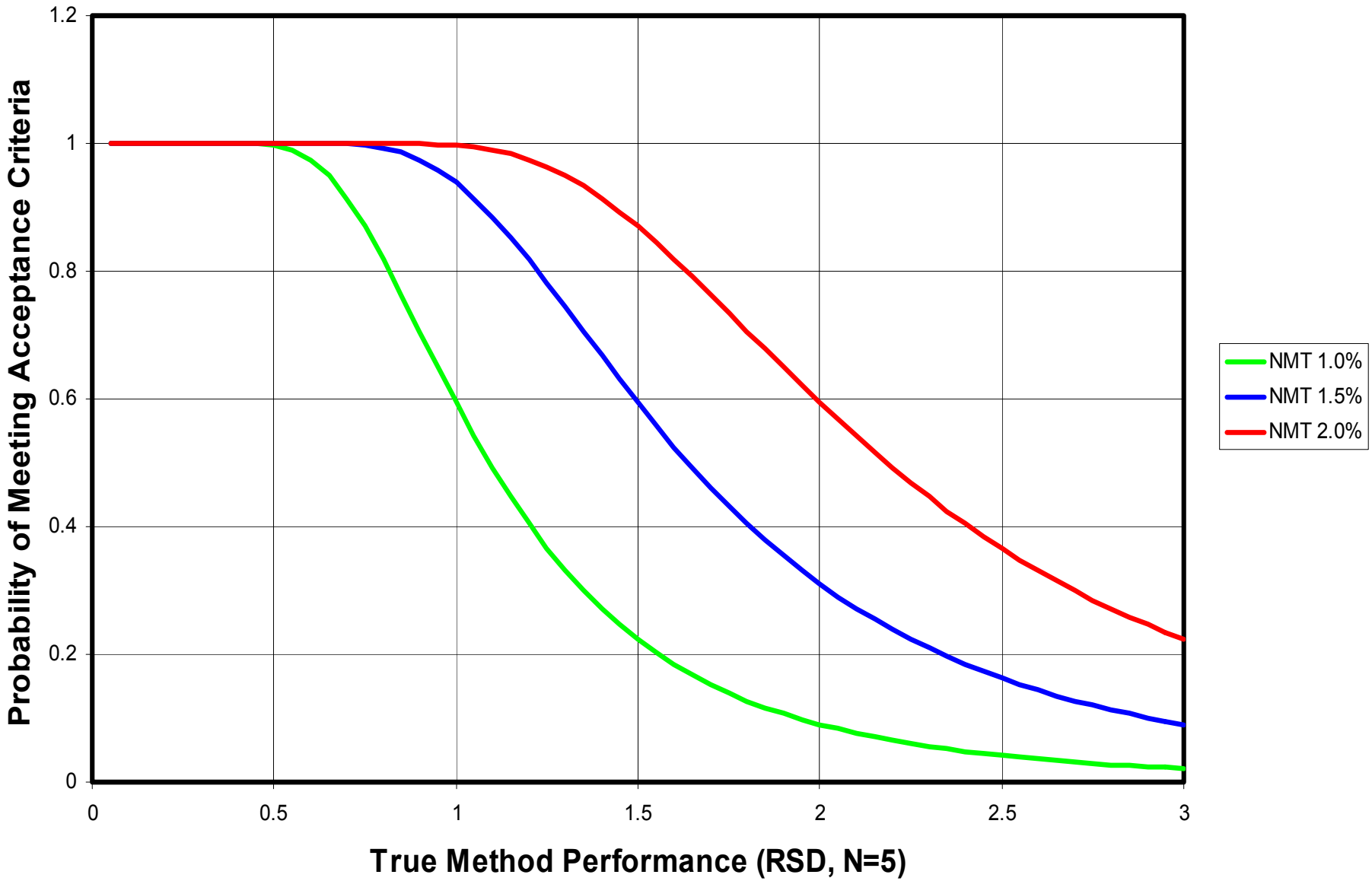
- H_0 - Null Hypothesis: The method's performance is acceptable.
- H_A - Alternate Hypothesis: The method's performance is not acceptable.
- Type I Error (α error): The method's performance is acceptable, but we conclude it is not.
- Type II Error (β error): The method's performance is not acceptable, but we conclude it is.
  - This type of error may be a concern.

Statistical Error

- If the largest acceptable true population RSD is 1.3%, we will pass an acceptance criteria of NMT 2.0% 95% of the time.
- 5% of the time, a good method will fail to meet this acceptance criteria
  - This is a Type I error.
- Almost 95% of the time, a method with a true RSD slightly more than 1.3% (technically unacceptable) will meet the acceptance criteria.
  - This is a Type II error.
  - As the true RSD increases, the Type II error decreases.

Minimizing Type II Error

- It may not be necessary.
  - No regulatory source addresses it.
  - No method validation literature addresses it.
  - If one considers the acceptance criteria to be the true limit of acceptability, the probability of Type II errors drops significantly.
  - If the acceptable performance is different than the acceptance criteria, a third “designed” performance is needed, which adds another layer of complexity.


Probability of Meeting Acceptance Criteria
Probability of Meeting Acceptance Criteria
1.2
1
0.8
0.6
0.4
0.2
0
0
0.5
1
1.5
2
2.5
3
True Method Performance (RSD, N=5)
NMT 1.0%
NMT 1.5%
NMT 2.0%

Minimizing Type II Error

- **How can it be done?**
  - **Design a method with better performance than needed and set a tighter acceptance criteria. Unacceptable methods (lower than the minimum required performance) have a lower probability of passing the tight acceptance criteria.**
  - **Increase N - the distribution curve narrows and is centered on the true method performance.**

Conclusions

- Formal statistical approaches can be used to set method validation acceptance criteria.
- Validation experimental designs were presented as examples because the acceptance criteria are tied to the design
  - No particular experimental design is being advocated. Choose your design and apply the principles in this presentation to derive scientifically and statistically sound acceptance criteria.

Conclusions

- Acceptable true method precision can be derived from compendial sources and regulatory guidances.
- Precision acceptance criteria can be statistically derived from the acceptable true method precision
- Accuracy can be derived from precision acceptance criteria and allowable bias.
- Linearity acceptance criteria can be derived from the acceptable true method precision.
- LOD/LOQ acceptance criteria are directly defined by regulatory sources.
- Therefore, precision, accuracy, LOD/LOQ, and linearity acceptance criteria that have a regulatory basis can be determined.