Estimation of platform process variation from large-portfolio datasets

Roger A. Hart, John Hixon, Mark DiMartino, Keith Britt, and Tom Mistretta, Amgen Inc.

4th Statistical and Data Management Approaches for Biotechnology Drug Development; USP Headquarters, Rockville, MD, **Oct 30 - Nov 1, 2017**

Outline

1. Background and Motivation
2. Rationale / Sources of Performance Variance
3. Methodology for Quantifying and Combining Variance
4. Ppk estimation with large-portfolio datasets
5. Points to Consider and Conclusions

Variation from Similar Processes can be used to Establish Product Specific Limits

Problem	• Developmental programs have small production datasets • Need to set appropriate limits for process control
Objective	• Quantify variance for process and product attributes • Use pooled variance to establish limits (Opportunity)
Value	• Improved understanding of variation for low sample size • Enhance control limit reliability (by including more sources of variability)

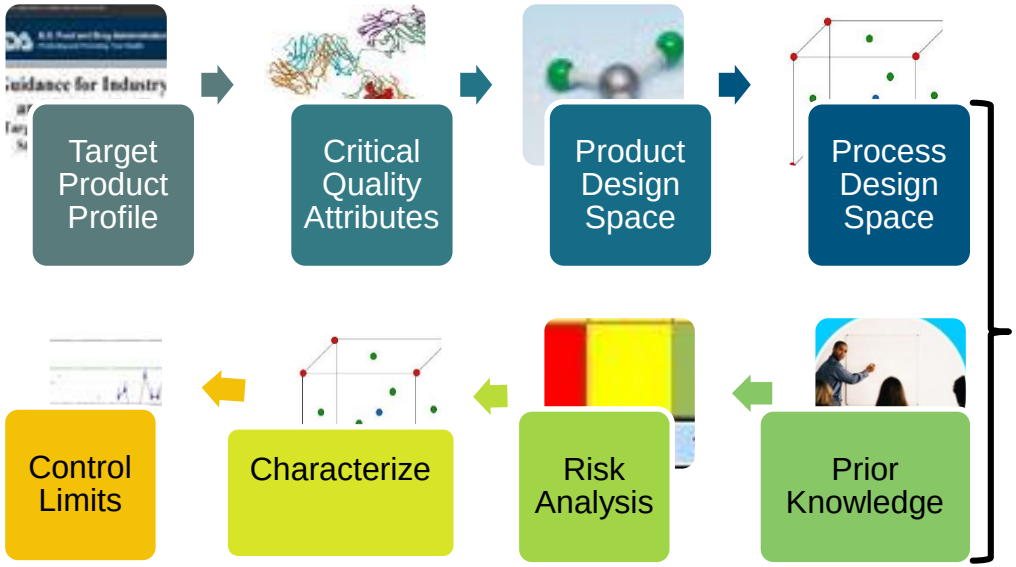
Variation can be analyzed across similar processes provided they can be objectively assessed for similarity (to ensure relevance)

Similarities in Process Design Present an Opportunity to Leverage Knowledge across Products

Process Validation Framework (FDA Guidance for Industry Process Validation: General Principles and Practices, 2011)

Process Design → Process Performance Qualification (PPQ) → Continued Process Verification (CPV)

Translation of Framework into actions within product development cycle



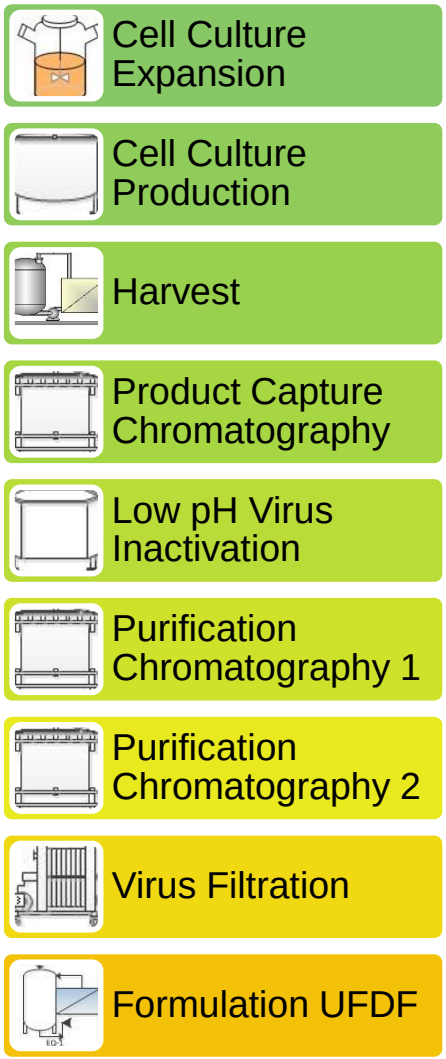
- **Processes can have similar design objectives**
  - ***TPP and CQA***
- **Consistent business and technical practices can produce similar process architectures**
  - ***Design Space, Risk Management, Characterized and Normal Operating Range***
- **Control limits establish expected ranges of performance for monitoring**

Common unit operations, performance parameters, and quality attributes allow processes to be considered similar for combining data

AMGEN®

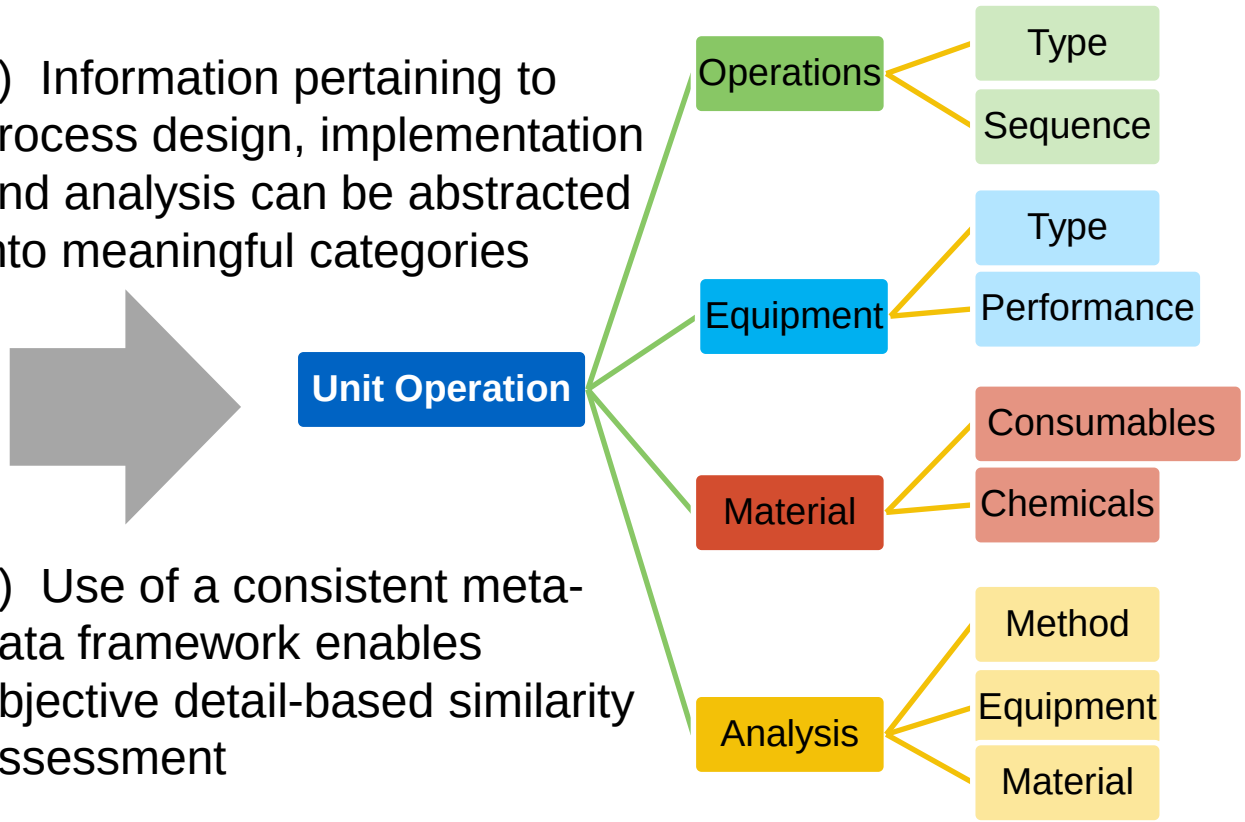
Process Designs can be Objectively Assessed for Similarity by Subject Matter Experts

Process Sequence



1) Information pertaining to process design, implementation and analysis can be abstracted into meaningful categories

2) Use of a consistent meta-data framework enables objective detail-based similarity assessment

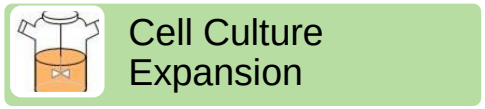


Processes with similar design have many sources of variation, some of which, may be similar (ex Equipment); yet others may be shared (ex. Raw Material)

AMGEN

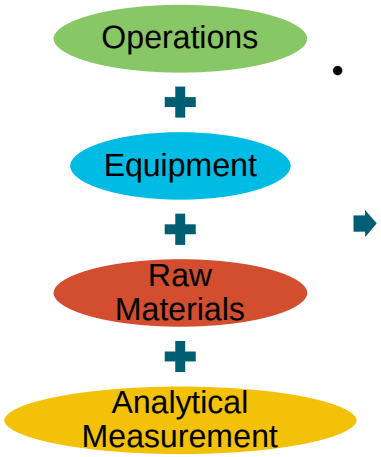
Information can be Aggregated to Understand Variance Across Similar Process Unit Operations

Principals of Aggregating Information



- **Specific attribute**
- **Specific process step**

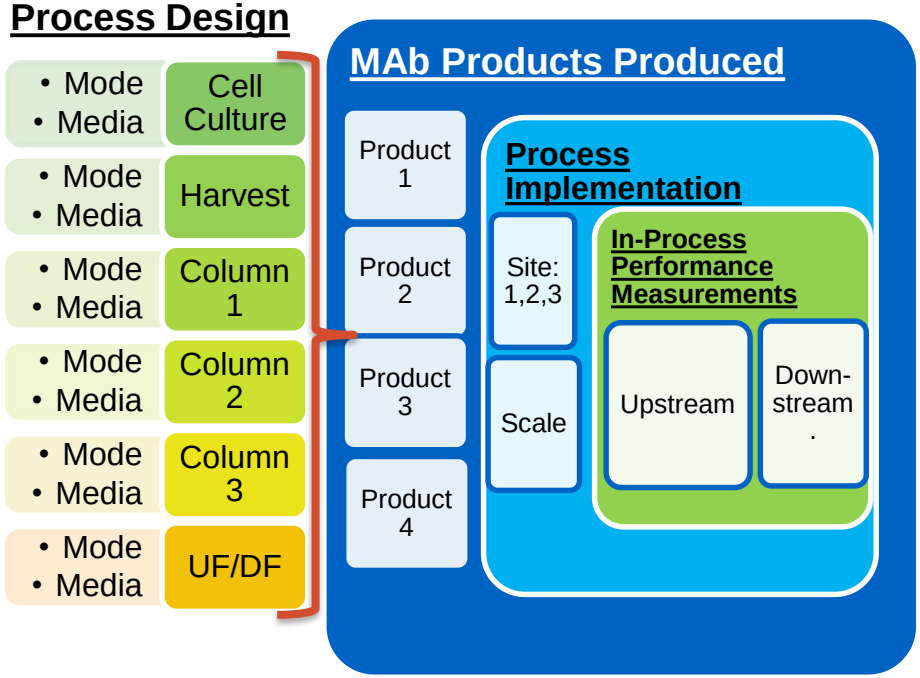
Attribute	Product Modality	Process Design	Unit Operation	Raw Materials
Product	MAb	Similar	Similar	Common
Group				



- **Assessment of the similarity of process design and implementation**
- **Define products contained in the Group**

Total Variance

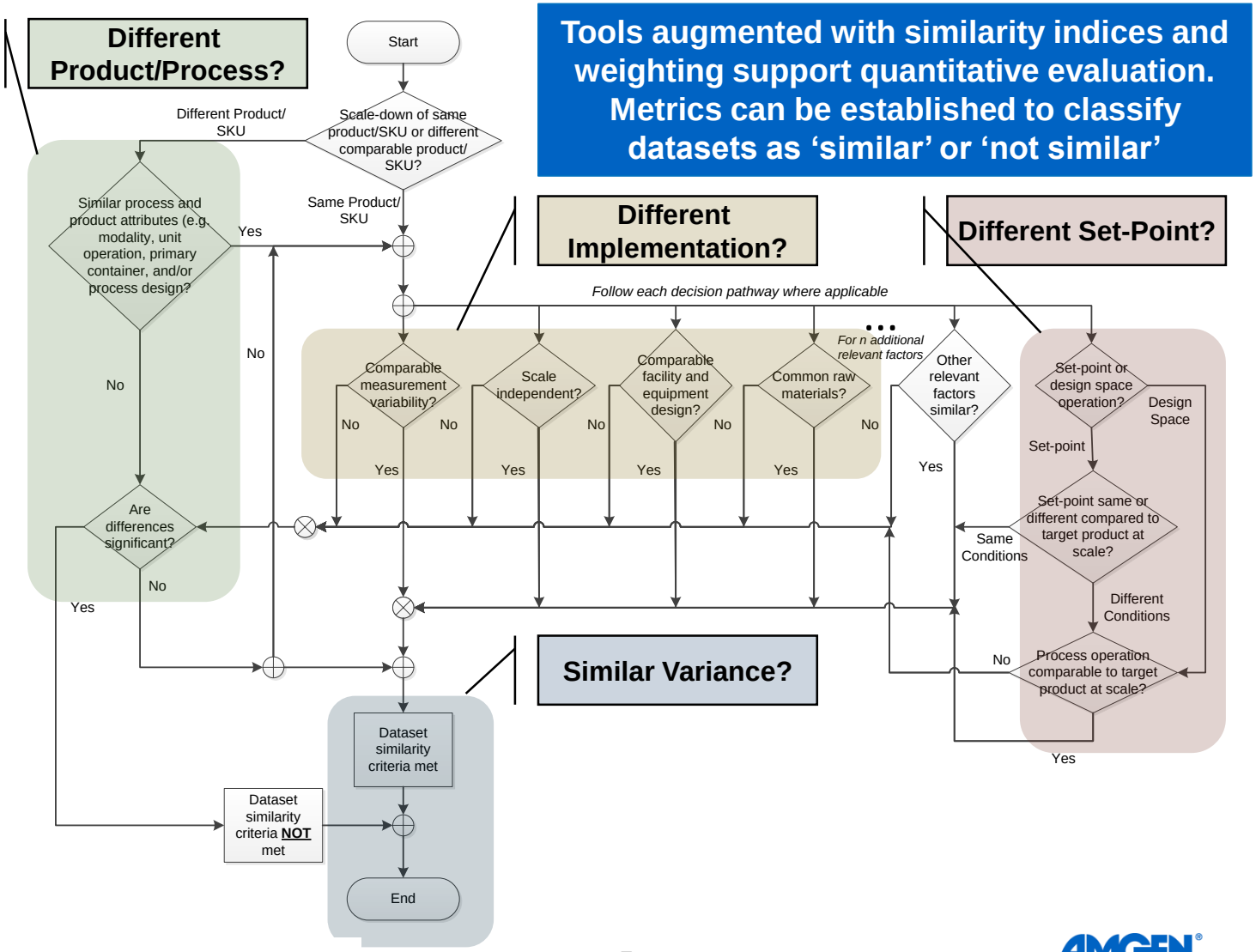
Total variance is the sum of contributing sources



Catalog of process information is needed to support assessment of process design and implementation similarity

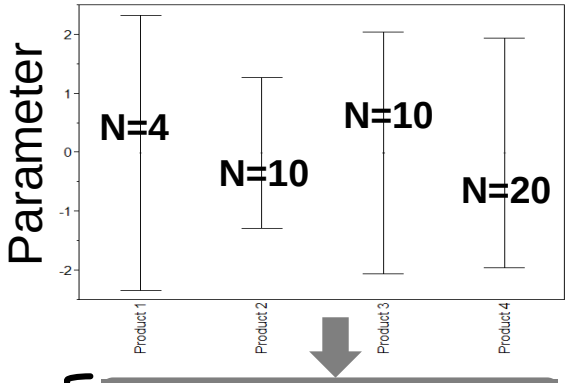
AMGEN®

Standardized Similarity Assessment Minimizes Subjectivity in Dataset Selection



Variance is Demonstrated to be Statistically Similar Prior to Combining Datasets

Objective Test for variance similarity



Brown Forsythe Homogeneity of Variance Test

Hypothesis Test
$H_o: Var1 = Var2 = Var3$

Calculate: $Z_{ij} = |y_{ij} - \hat{y}_i|$

Run an ANOVA on each set of $Z_{ij}'s$

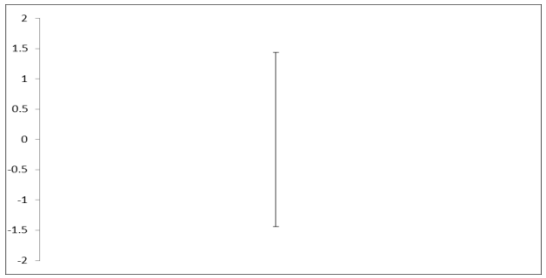
If p=value ≤ 0.05, reject H_o

y_{ij}=data value, $\hat{y}_i$= median of dataset

Output of test influences product Inclusion in Group

Determine variance to use when establishing limits

Establish Limits for parameter



Approach leads to more accurate prediction of limits for processes with few data points

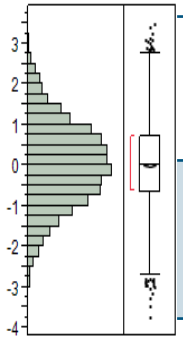
AMGEN®

Empirical Bayes approach leverages knowledge of similar attributes to improve accuracy of limits

Classical Approach

Calculate Mean and Variance for a performance parameter for a single process used to manufacture a product to determine limits.

1). Determine mean and standard deviation



Mean $\bar{x} = \frac{\sum x}{n}$

Standard Deviation $\sigma = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}}$

2). Calculate Limits Control Limits: $\mathrm{CL} = \bar{x} \pm 3\sigma$

Tolerance Interval: $TL = \bar{x} \pm k\sigma$

For low run rate products, significant amount of time may be required to establish accurate variance estimates

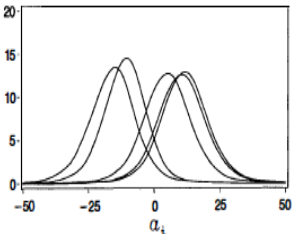
Empirical Bayes

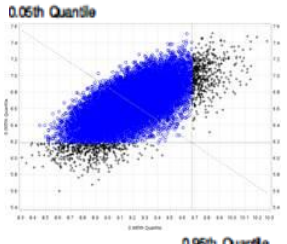
Use existing data to form a Bayesian model that predicts the tolerance interval of a performance parameter based on simulation for the performance parameter

1). Demonstrate Similarity in Variance (HOV Test)

2). Obtain Statistical Measures of Prior Information

3). Simulate to Determine Quantiles of Interest





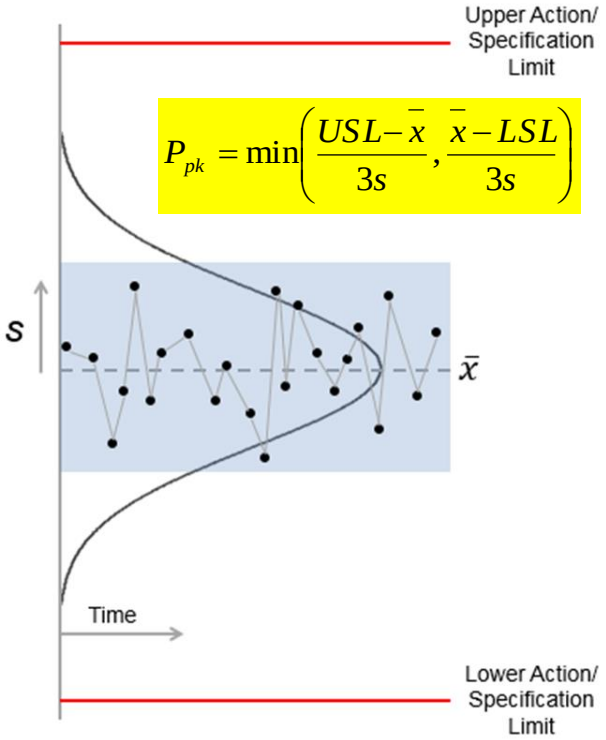
[1]Wolfinger, (1998). Tolerance Intervals for Variance Component Models Using Bayesian Simulation. *Journal of Quality Technology*, *30* (18-32).

Use portfolio data to simulate uncertainty in reference product mean and variance. Overcome concerns with small sample sizes.

AMGEN®

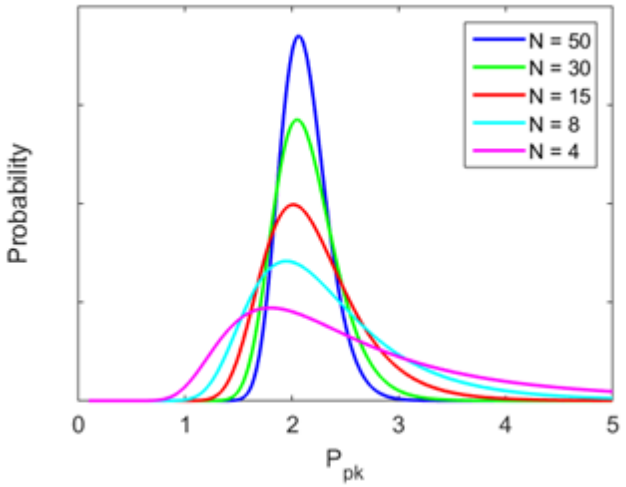
Prediction of Process Performance Indices, Ppk, requires accurate estimates of process variation

- In Biopharmaceutical Manufacturing, a 'large' dataset often refers to programs with N > 30.
- During product development, manufacturing datasets are commonly quite small (N<10)
- Data from large product portfolios, combined with Platform Variance approaches, provides a means for predicting anticipated Ppk for small datasets



- The PDF is nearly symmetrical for large sample sizes
- It becomes skewed towards higher P_{pk} as sample size decreases.

KEY QUESTION:
- How large should the Portfolio Dataset be to obtain a credible Ppk prediction?



Ppk Probability Density Function (Novoa and Artiles-Leon, 2009)

Simulation courtesy of Keith Britt

AMGEN

Simulation Experiments used to investigate the importance of "Reference Product Sample N"

Definitions:

- **Platform Variance Products (PVP):** A set of products having the same variability for a single measured Parameter.
- **Reference Product (RP):** The product for which we want to develop Empirical Bayes Tolerance Limits (EBTL) and predict the PPK.
- **Empirical Bayes Predictive PPK (EBPPK):**
  - Derive the distribution of the PPK parameter from 20,000 independent samples using Bayesian simulation. (Wolfinger, 1998).
  - Extract the Median and 95% Credible Intervals for the PPK from the distribution of the PPK.

Method:

a. Define Platform Variance Product Statistics

b. Create random 60 samples for the 4 PVPs and 100 random samples for the Reference Product

   a. Perform the EBPPK analyses using the fixed 60 samples from the PVP

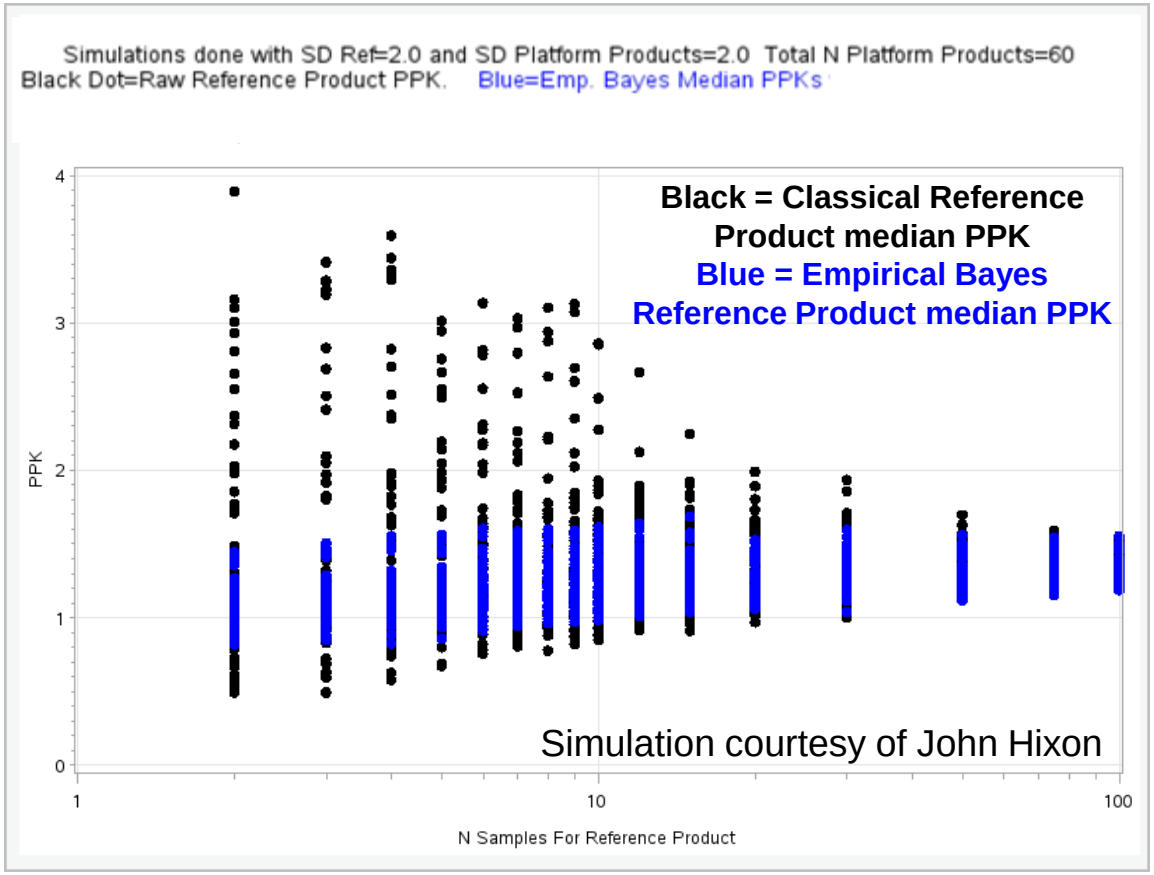
   b. Select Reference Product data from among the 100 samples

   c. Increment reference sample within the set: {2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 15, 20, 30, 50, 75, 100}

c. Repeat (b) 50 times.

Prod	N	Mean	SD	USL	LSL	PPK
Ref		100	2			
1	10	95	2	103	87	1.33
2	15	125	2	133	117	1.33
3	25	110	2	118	112	1.33
4	10	90	2	98	82	1.33

Comparison of PPK Estimated from Classical vs Empirical Bayes Platform Variance Approach

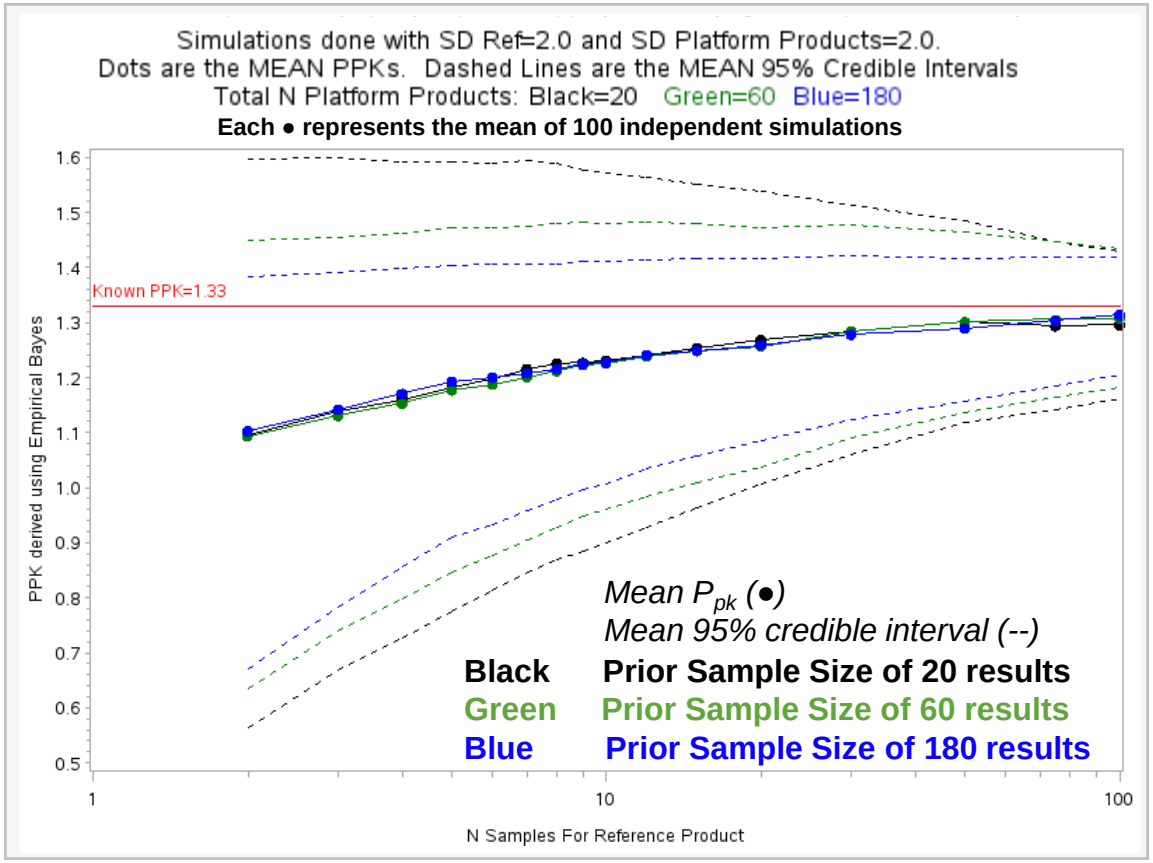


- Analysis performed under conditions that simulate data collection reality (i.e. data acquired with each new lot augments prior data).
- Simulations conducted to assess the accuracy of EB PPK estimates with increasing Reference Product samples.
- EBPPK estimates give the user more confidence that the reported result is representative of anticipated process performance, even at small sample sizes.

Empirical Bayes PPK predictions are preferred for small Reference Product Sample size

AMGEN®

Effect of Platform Product Dataset size on Empirical Bayes Platform Variance PPK Estimate



Simulation courtesy of John Hixon

- Method the same except: 100 simulations; varied random sample size for the 4 PVPs to include 20, 60, and 180 values
- The Platform Product dataset sizes evaluated did not impact the Ppk estimate, but did effect the credible interval
- Predicted mean PPK approaches known value, and credible interval narrows, with increasing Reference Product N
- PPK offset results in a conservative estimate at small sample size

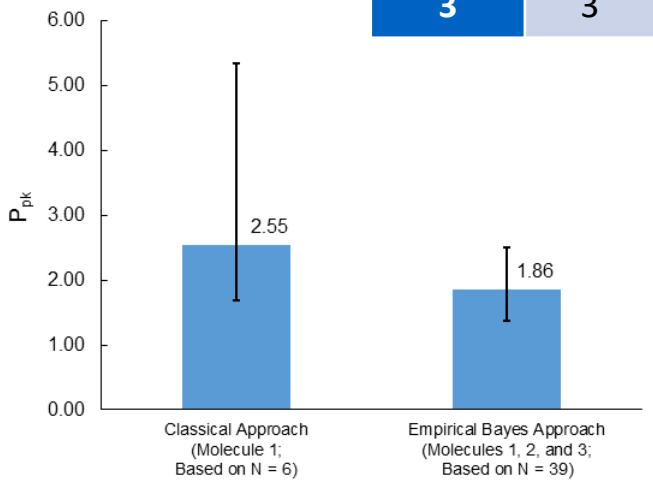
More data gives more confidence in the PPK estimate but not the conclusion of the analysis

AMGEN®

Empirical Bayes Platform Variance PPK Estimate agrees with Conventional Result for large N

DS pH – Classical (N = 6) vs EB PPK (N = 39)

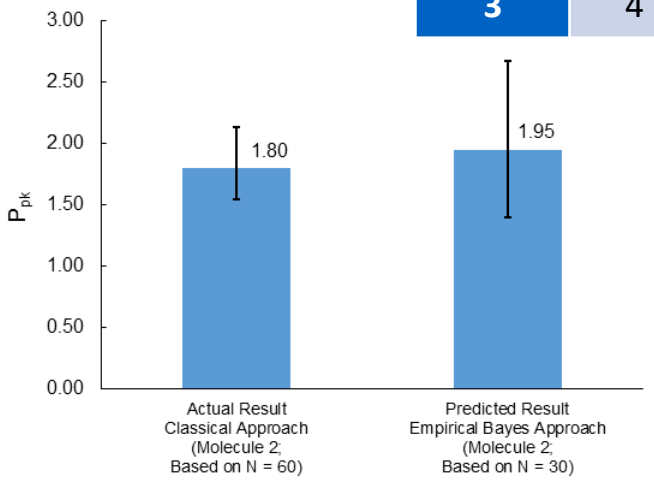
Prod	N
1	6
2	30
3	3



Classical Ppk result is higher at small sample sizes - insufficient variation to describe the underlying distribution

DS pH – Actual Ppk (Classical; N = 60) vs EB PPK (N = 30)

Prod	N
1	20
2	6
3	4



EB PPK result agrees with conventional P_{pk} result as more data are collected from maturing programs

AMGEN®

Conclusions:

Manufacturing data reflects common cause variability

Sources of variation can be quantified and combined to provide a more accurate performance assessment. Empirical Bayes predictions provide a reliable means for variation combination.

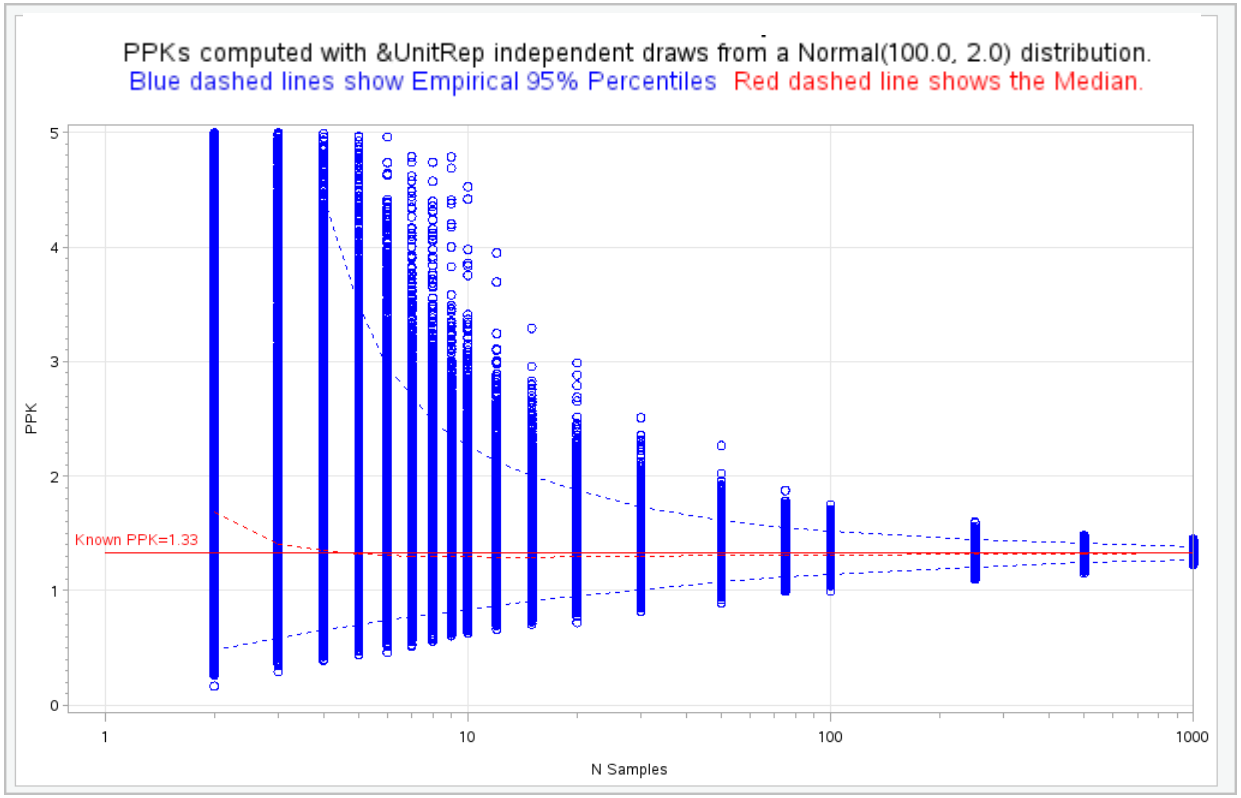
Evaluating variance in this manner allows for better product lifecycle management in a CPV Program. Empirical Bayes predicted PPK values are consistent with eventual large N conventional values.

Cross product performance variance data enhances reliability for establishing limits for product performance parameters with a low sample size

Acknowledgements

- Aine Hanly
- Margaret Faul
- Cenk Undey
- Greg Naugle
- Patrick Gammell

PPK Probability Distribution Simulation



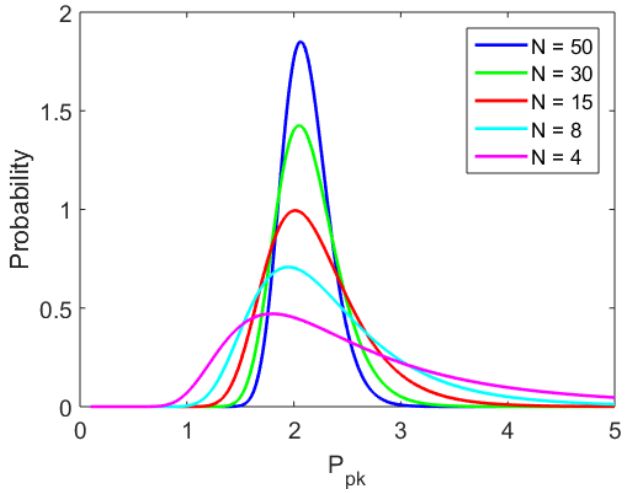
Ppk Simulation Method

- Created 10,000 samples with a size of 1,000 from a Normal Distribution with a Mean=100.0 and an SD=2.0.
- Assigned LSL=92.0 and USL=108.0 (+/- 4 SDs from the Mean)
- PPK=min(USL-Mean, Mean-LSL)/(3*SD)= 4/3 =1.33
- Computed PPK for sets of "first N observations": {2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 15, 20, 30, 50, 75, 100, 500, 1000}.

AMGEN®

Uncertainty at low sample sizes necessitates the determination of confidence intervals for Ppk

Behavior of the Empirical pdf as a Function of Sample Size



- **For larger sample sizes, the pdf for Ppk is nearly symmetrical**
- **As sample size decreases, the width of the distribution increases, it becomes increasingly skewed towards higher P_{pk} values, and the peak maximum drifts towards lower P_{pk} values.**

- ## Method 1: Bissell's approximation

Bissell (1990) derived a two sided confidence interval for P_{pk} (or C_{pk}) assuming that the empirical probability density function was normal

$$CI = P_{pk}\left[1 \pm \Phi^{-1}\left(1-\frac{\alpha}{2}\right)\sqrt{\frac{1}{9nP_{pk}^2}+\frac{1}{2(n-1)}}\right]$$

- ## Method 2: Bootstrapping (with replacement)

Bootstrapping is a resampling technique that enables one to estimate a hypothetical distribution of future results based on historical observations

- ## Method 3: Empirical Probability Density Function (pdf)

Novoa and Artiles-Leon (2008) derived an expression for the underlying probability density function for P_{pk} (or C_{pk}) :

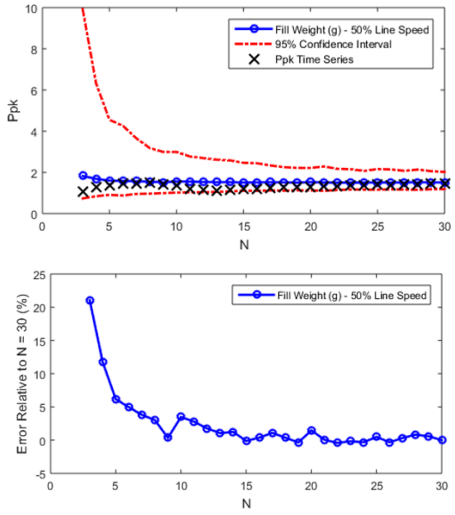
This equation is solved numerically to generate the pdf for P_{pk} (or C_{pk}). Confidence intervals can be derived from this distribution by evaluating the area under the curve associated with the 100(1-α)% confidence interval selected.

$$g(P_{pk}) = \frac{6\sqrt{\frac{n}{2\pi}\left(\frac{n-1}{2}\right)^{n-1}}}{\sigma\Gamma\left[\frac{1}{2}(n-1)\right]} \cdot \int_0^{\frac{USL-LSL}{6P_{pk}}} \left(\frac{s}{\sigma}\right)^{n-1} \cdot \exp\left[\frac{-(n-1)}{2}\left(\frac{s}{\sigma}\right)^2\right] \cdot \left\{\exp\left[\frac{-n}{2\sigma^2}\left(-3P_{pk}\cdot s + USL - \mu\right)^2\right] + \exp\left[\frac{-n}{2\sigma^2}\left(-3P_{pk}\cdot s + \mu - LSL\right)^2\right]\right\} ds$$

AMGEN®

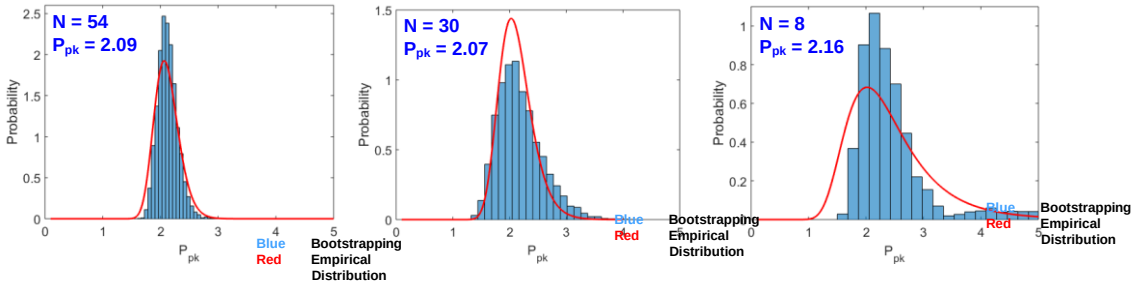
Accuracy of Confidence estimates for Ppk depend on the method employed

Median Bootstrapped P_{pk} vs. Raw P_{pk} Time Series



Smoothed Ppk profile and confidence intervals derived from N = 10000 bootstraps as a function of sample size

Parameter	Method	N	P_{pk}	Low 95% Confidence Limit	High 95% Confidence Limit
DP Plunger Depth	Bissell's Approximation	54	2.09	1.68	2.49
	Bootstrapping (with Replacement)			1.84	2.46
	Novoa and Artiles-Leon's Empirical Distribution			1.74	2.58
Production Bioreactor Titer	Bissell's Approximation	30	2.07	1.52	2.61
	Bootstrapping (with Replacement)			1.56	3.14
	Novoa and Artiles-Leon's Empirical Distribution			1.63	2.79
DSI Total Impurities	Bissell's Approximation	8	2.16	1.01	3.32
	Bootstrapping (with Replacement)			1.76	5.35
	Novoa and Artiles-Leon's Empirical Distribution			1.36	4.22



- **Results from three methods similar at large sample sizes, but become more variable as sample size is decreased.**
- **Bissell's approximation breaks down at smaller sample sizes because it assumes the P_{pk} distribution is normal.**
- **Bootstrapping resampling procedure breaks down at small sample sizes since there are fewer unique results to accurately describe the distribution.**
- **Empirical pdf yields the best results for P_{pk} confidence intervals**

AMGEN®