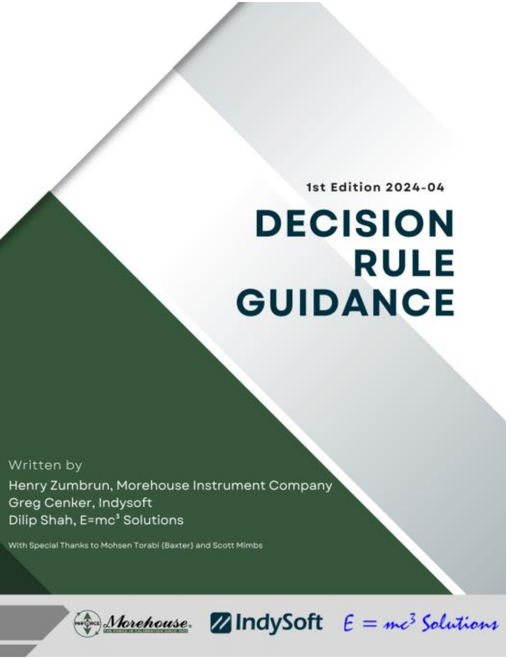
Who Needs Another Tutorial on Risk or Decision Rules?



SCAN ME

https://mhforce.com/ncsli-force-course/

Who Needs Another Tutorial on Risk or Decision Rules?

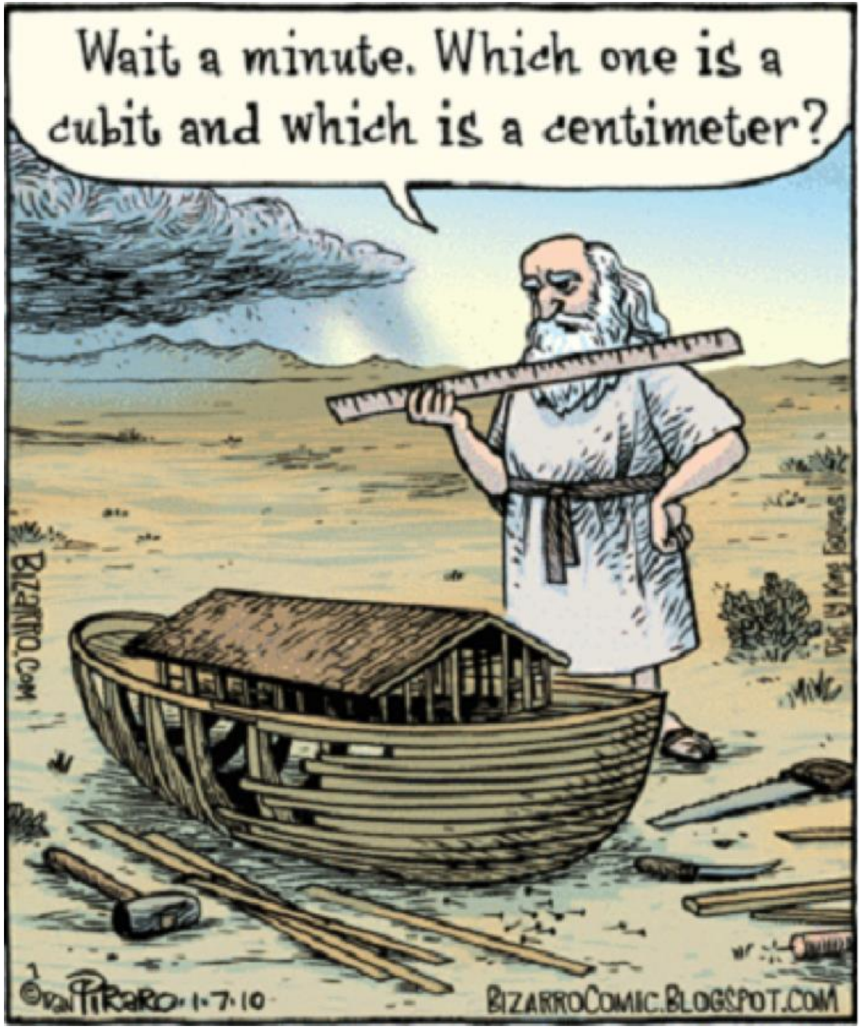
- Henry Zumbrun II, Morehouse Instrument Company
- 1742 Sixth Ave
- York, PA 17403
- PH: 717-843-0081   web: www.mhforce.com
- sales: hzumbrun@mhforce.com

Who Needs Another Tutorial on Risk or Decision Rules?

Greg Cenker, Indysoft

- 146 Fairchild St, Suite 202.
- Daniel Island, SC 29492
- web: www.indysoft.com
- greg.cenker@indysoft.com

IndySoft


Wait a minute. Which one is a cubit and which is a centimeter?
BIZARRO.COM
BIZARROCOMIC.BLOGSPOT.COM

Abstract

This 4-hour tutorial will help the participant eliminate much of the noise on decision rules. It will provide guidance anyone can take away and implement in their laboratory. This session aims to give guidance beyond simply requesting a 4:1 TUR (antediluvian) or accepting a shared-risk scenario as with simple acceptance.

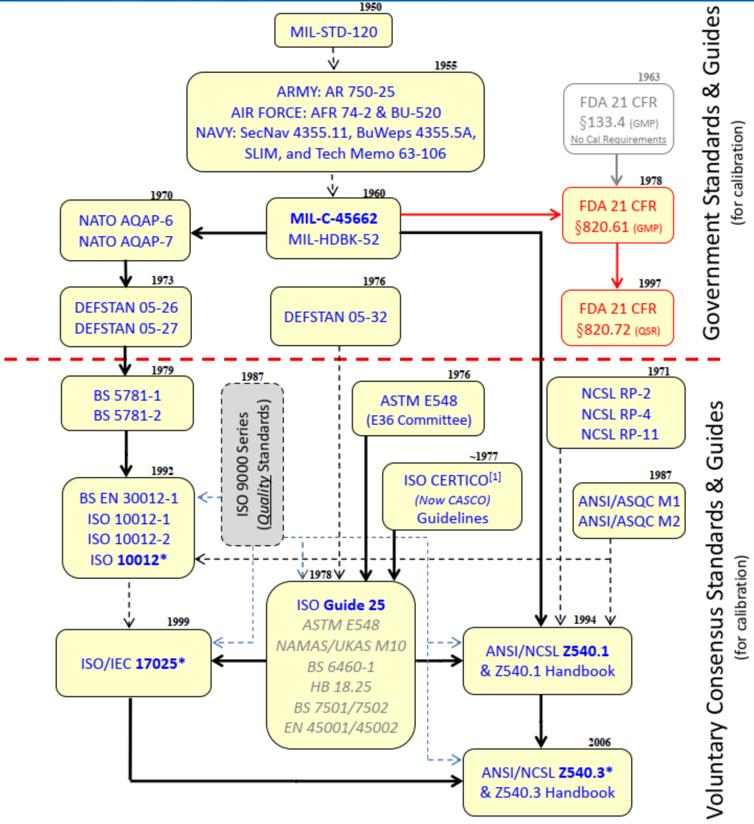
When a calibration report is provided, a typical concern for the customer is to know if the item calibrated is within the tolerance specified so they can continue using the device (i.e., many want a new sticker ☺).

While this long-established approach has been in service since 1955, measurement science has evolved (or not?). With our technological evolution, a simple “pass/fail” may no longer be enough.

Learning Objectives

1. Understanding Measurement Traceability Requirements.
2. Know the Fundamentals of Measurement Uncertainty.
3. Understanding Measurement Data Sampling Requirements.
4. Understanding the Basics of Decision Rules.
5. Know the Differences and Applications for Specific and Global Risk Models.
6. Introduction to Metrology Costing Models.

History



* Asterisk designates active voluntary consensus standards for calibration.

[1]CERTICO was created in 1970 as the "Committee on Certification". In 1985, it became CASCO or "Committee on Conformity Assessment"

Dates shown reflect initial publication of earliest document; see Table 1 for revision histories. Dotted lines indicate "indirect" predecessors.

Figure 1. Simplified evolution of international and national *calibration* standards and guidelines

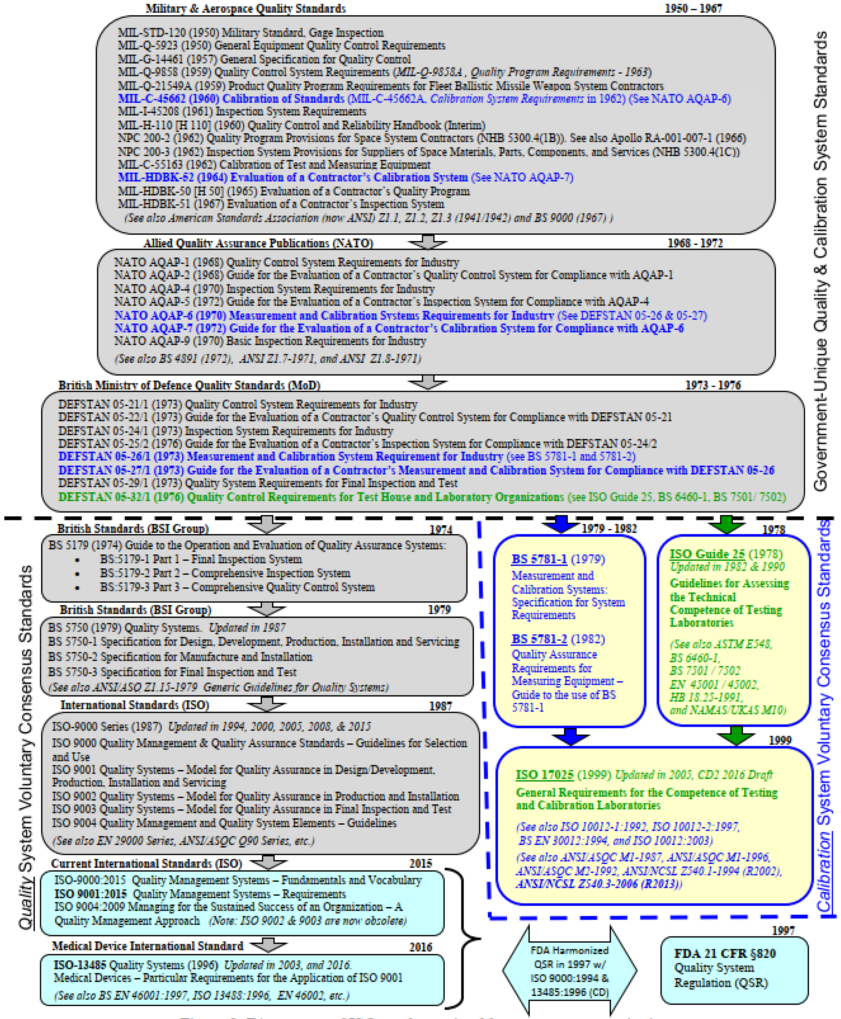


Figure 2. Divergence of ISO *quality* and *calibration* system standards

From Calibration in Regulated Industries: Federal Agency use of ANSI Z540.3 and ISO 17025 – P.Reese

Where did the famous 4:1 Requirement Come From?

Origins of the ubiquitous 4:1 test accuracy ratio have been and attributed to Hayes and Crandon of the U.S. Navy in the mid-1950s. Hayes published the first known statistical analysis of calibration quality (false accept/reject) in 1955, invoking the concept of accuracy ratios, based primarily on the 1954 seminal works of Eagle and Grubbs & Coon.

A brief summary of the TAR requirements throughout the life of MIL-STD-45662A and its associated handbook is given, including its replacements ANSI Z540.1-1994 and Z540.3-2006.

- 1960: MIL-C-45662: Required a **10:1** TAR [12]
- 1962: MIL-C-45662A: No TAR specified [7]
- 1964: MIL-HDBK-52: Required **4:1** to **10:1** TAR [13]
- 1980: MIL-STD-45662: No TAR specified [30]
- 1984: MIL-HDBK-52A: *Mentions* 10:1, 4:1, 3:1, 2:1, & 1:1 TAR *examples only* [31]
- 1988: MIL-STD-45662A: Required **4:1** TAR [29] (replaced by Z540.1-1994 in 1995)
- 1989: MIL-HDBK-52B: Provided guidance on **4:1** TAR [177]
- 1994: Z540.1-1994 Required uncertainty analyses or **4:1** TAR [32]
- 1995: Z540.1 Handbook: Provided guidance on **4:1** TAR, guardbanding, etc [181A]
- 2006: Z540.3-2006 Requires **<2%** false accept risk (FAR) *or* **4:1** TUR [44]
- 2009: Z540.3 Handbook: Provides guidance on **<2%** FAR and **4:1** TUR [45]

Condon: NCSL (1966): [4]

"NASA's policy on ratio-of-accuracy as stated in NPC 200-2 requires that 'Within the state-of-the-art limitation, the standards used for calibration of inspection, measuring, and test equipment shall have a tolerance no greater than 10 % of the allowable tolerance for the equipment being calibrated.' …many measurement requirements are becoming so sophisticated that they approached the limits of the science of metrology. In such cases, it becomes impossible to maintain the 10 to 1 ratio of accuracy in the calibration of the instrument"

Russell: NCSL (1966): [5]

"…our discussion centered around the accuracy ratio of standard to instrument during measurement and calibration operations… Basically, the problem revolves around the actual or implied requirement that the accuracy of an instrument or standard used to measure a quantity, or to calibrate another instrument, shall be 10 times as accurate as the quantity or the instrument being calibrated.

There is also the implication that the 10-to-1 ratio of accuracy shall exist between every level or echelon in the traceability chain for product to National Standards. This requirement could create an impossible situation…

Most contractors indicated that the ratio-of-accuracy requirements imposed upon them ranged from 10:1 to 4:1, or 'state-of-the-art'. In nearly every case, they stated that the 10-to-1 requirement was considered unrealistic from an economic as well as a practical point of view…

I am of the opinion that that there are too many documents that basically state parallel requirements, the majority of which are, to a degree, unrealistic… Where 4-to-1 is maintained… the reliability of the calibrated instrument accuracy is assured…"

Introduction - First efforts at reducing measurement risk

Modern measurement decision risk traces its roots to the late 1940’s and early 1950's.

- Alan Eagle
- Frank Grubbs, and Helen Coon

Eagle's 1954 paper focused on methods to analyze, quantify, and mitigate “test errors”

- methods for calculating consumer and producer risk.
- methods for establishing “test limits,” referred to as Guard Bands today.

Grubbs and Coons, in the same publication, expanded on Eagle’s paper

- methods for balancing consumer and producer risk for measurement-based decisions.

These papers were the genesis of requirements found in standards such as MIL-STD 45662, ANSI/NCSL Z540.1, ANSI/NCSL Z540.3, ASME B89.7.4.1-2005, and JCGM 106:2012, as well as other National and International standards and papers.

Introduction - First efforts at reducing measurement risk

In 1955, the U.S. Navy needed improved measurement reliability in their guided missile program.

In response, Jerry Hayes authored Technical Memorandum No. 63-106. Aspects of this document are still relevant today.

- Calibrated equipment needed for testing
- Establishment of reasonable testing risk levels
- Reasonable design tolerances
- Adequate procedures for testing

Using Eagle’s work, Hayes proposed a “family of curves” to determine specific testing risk

- A new family of curves had to be established for each change in process or design tolerance
- Computing consumer risk was very arduous with slide rules

With a lack of computing power in 1955, a 4:1 accuracy ratio was established as Navy policy

- Established for a 1% Consumer Risk objective
- Assumed the test equipment and calibration standard manufacturing specifications were developed for a 95% confidence level

This is the origin of the Test Accuracy Ratio (TAR) used in calibration and testing for decades.

Introduction – The 4:1 Rule and Measurement Risk

The accuracy ratio in the Eagle equation was a ratio of two standard deviations.

- The numerator σ_x is "the true standard deviation of the product distribution."
- The denominator σ_e is "the standard deviation of the errors of measurement."

$$r = \frac{\sigma_x}{\sigma_e}$$

$$CR = \frac{1}{\pi}\int_k^{\infty}\int_{-r(k+t)+b}^{r(k-t)+b} e^{\frac{-(t^2+s^2)}{2}}\,ds\,dt$$

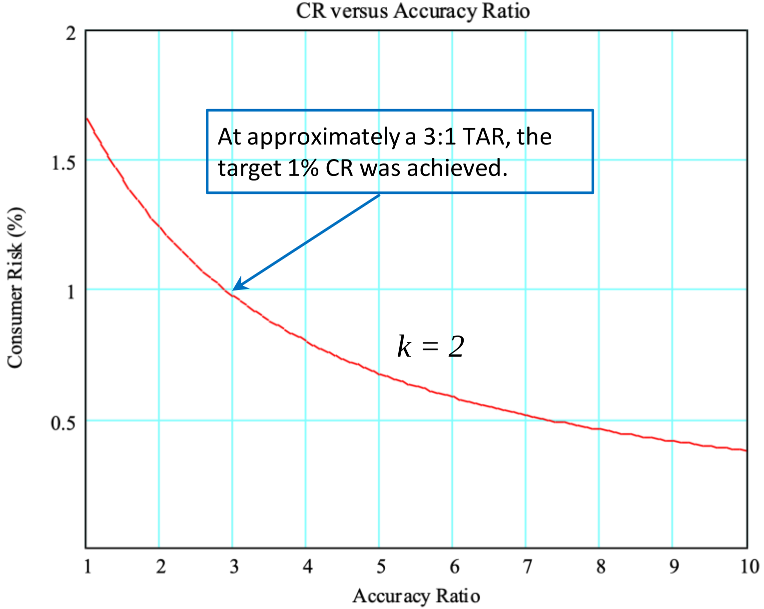
Assuming a 95% confidence level for each, Hayes used specifications for the "test ratio."

- The numerator would be UUT specifications, and
- the denominator the Calibration Standards.

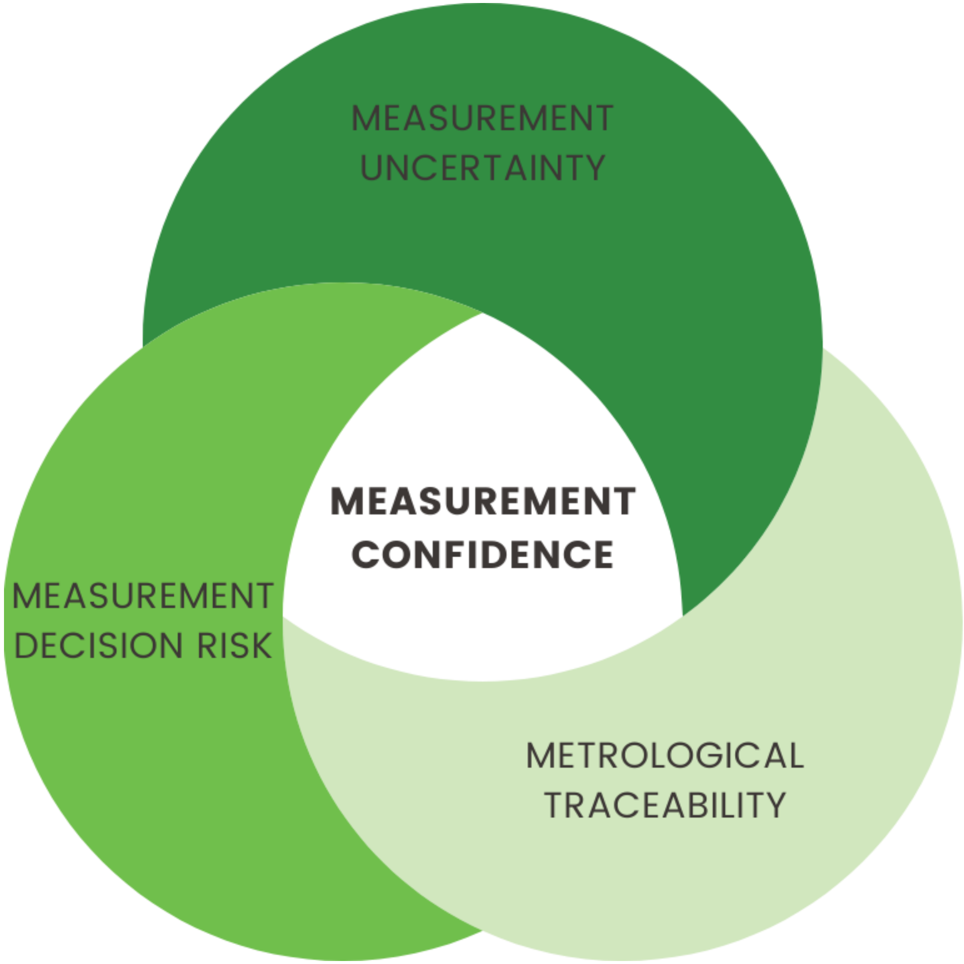
The intent was to use procedure controls to mitigate other "measurement errors" during testing.

Not fully trusting all specs were equal, Hayes and Crandon increased the ratio to 4:1.

Therefore, the 4:1 Rule provides a specific level of risk **<u>only under explicit assumptions and conditions</u>**.



Measureme Confidence



ISO/IEC 17025:2017 Requirement

6.5 Metrological traceability

6.5.1 The laboratory shall establish and **maintain metrological traceability of its measurement results** by means of a documented unbroken chain of calibrations, each contributing to the measurement uncertainty, linking them to an appropriate reference.

NOTE 1 In ISO/IEC Guide 99, metrological traceability is defined as the “property of a measurement result whereby the result can be related to a reference through a **documented unbroken chain of calibrations, each contributing to the measurement uncertainty**”.

NOTE 2 See Annex A for additional information on metrological traceability.

Measurement Related Terms

Metrological Traceability: Property of a measurement result whereby the result can be related to a reference through a ***documented unbroken chain of calibrations, each contributing to the measurement uncertainty.***

NOTE 1 For this definition, a ‘reference’ can be a definition of a measurement unit through its practical realization, or a measurement procedure including the measurement unit for a non-ordinal quantity, or a measurement standard.

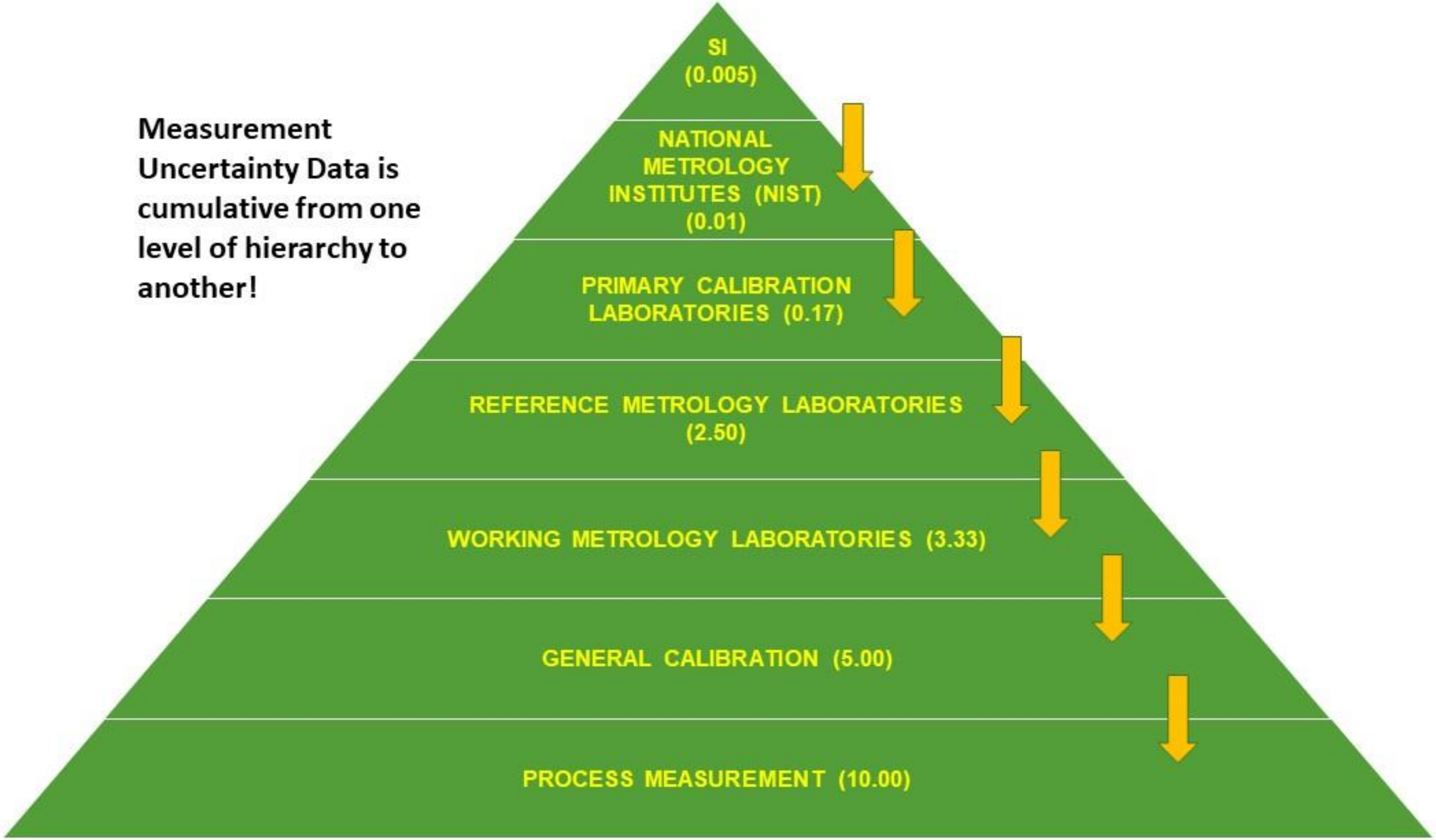
NOTE 2 Metrological traceability requires an established calibration hierarchy.

NOTE 3 Specification of the reference must include the time at which this reference was used in establishing the calibration hierarchy, along with any other relevant metrological information about the reference, such as when the first calibration in the calibration hierarchy was performed.

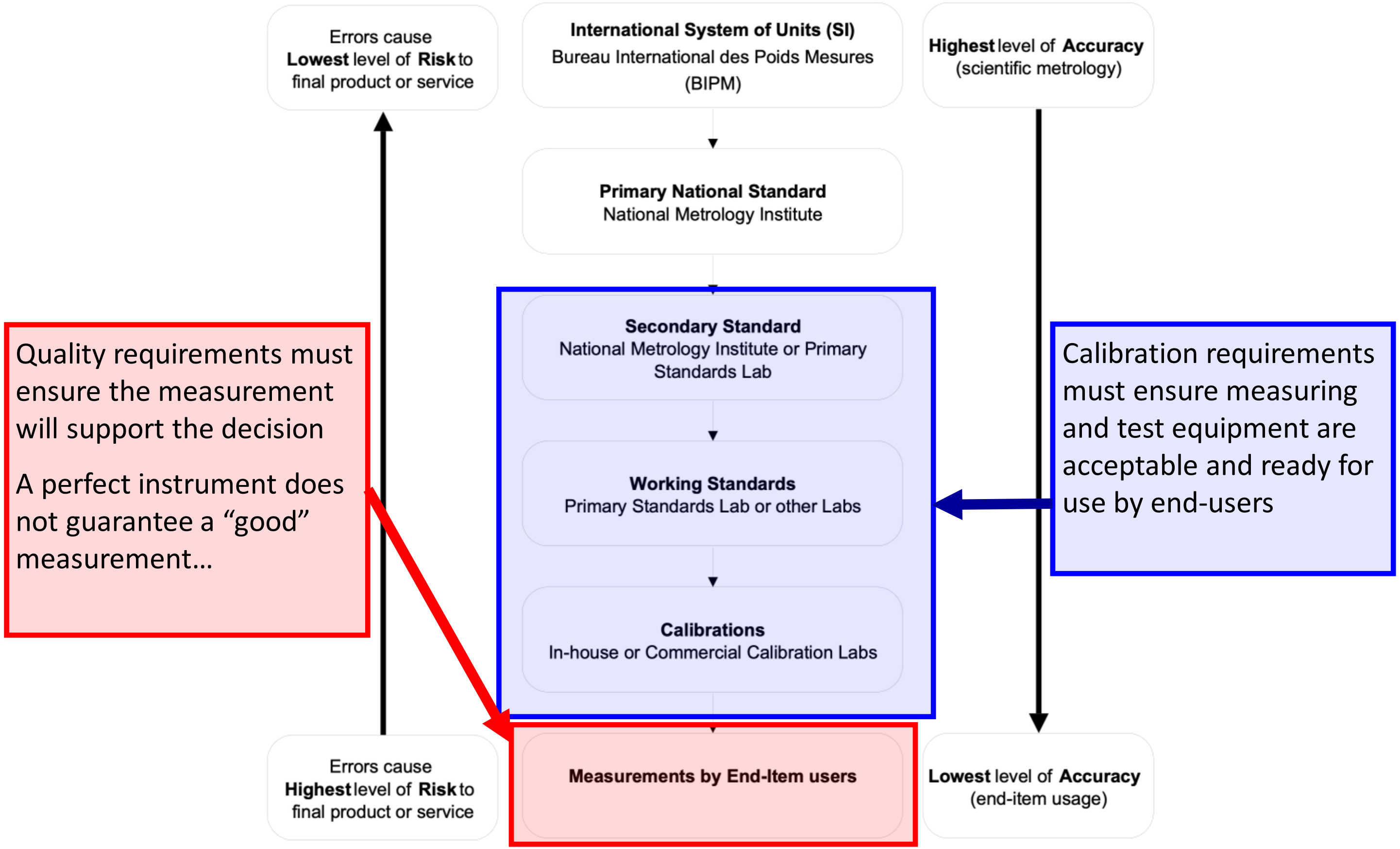
NOTE 4 For measurements with more than one input quantity in the measurement model, each of the input quantity values should be metrologically traceable.

Measurement Traceability

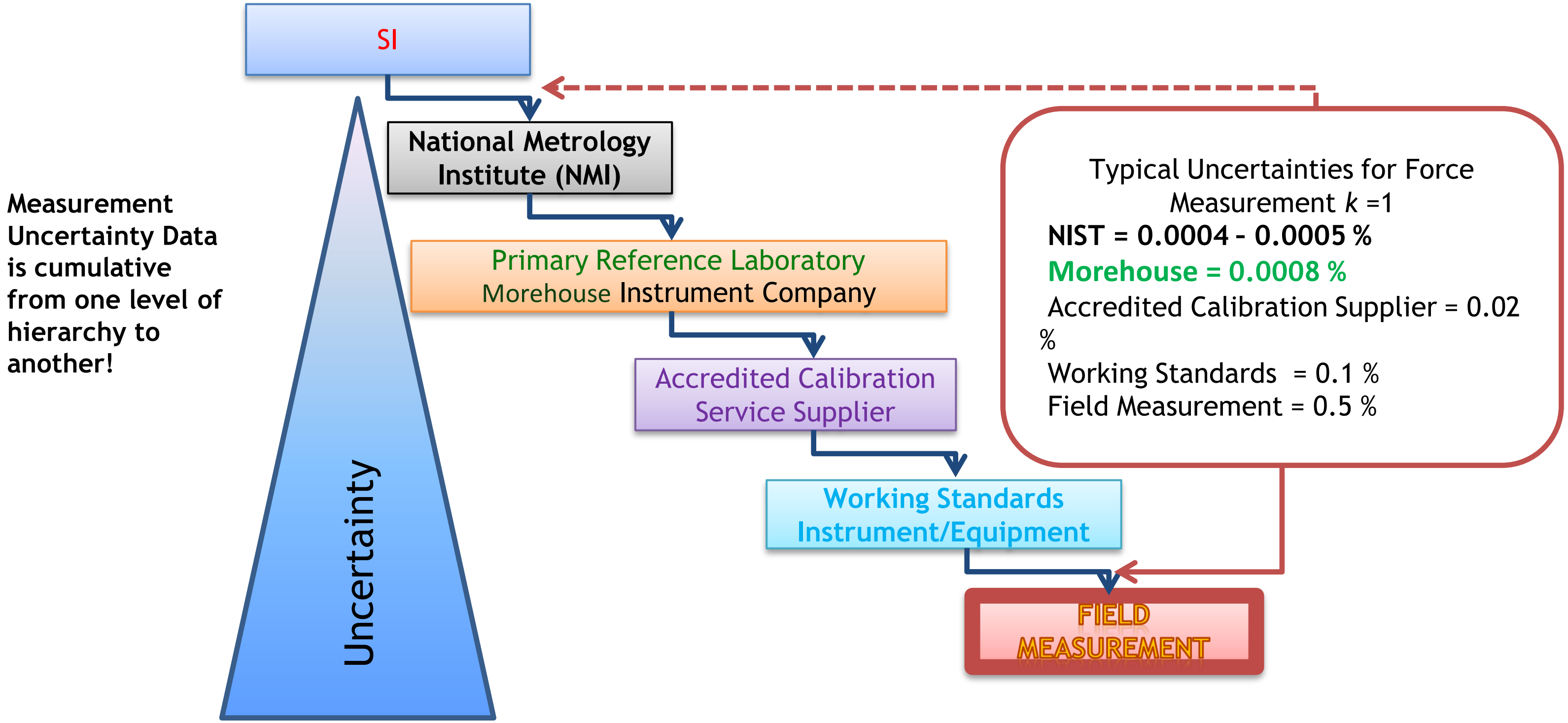
Measurement Uncertainty Data is cumulative from one level of hierarchy to another!



Measurement Accuracy, Risk, and the Metrology chain



Measurement Uncertainty's Relation to Measurement Hierarchy



Measurement Uncertainty

7.6 Evaluation of measurement uncertainty

7.6.1 Laboratories shall identify the contributions to measurement uncertainty. When evaluating measurement uncertainty, all contributions that are of significance, including those arising from sampling, shall be taken into account using appropriate methods of analysis.

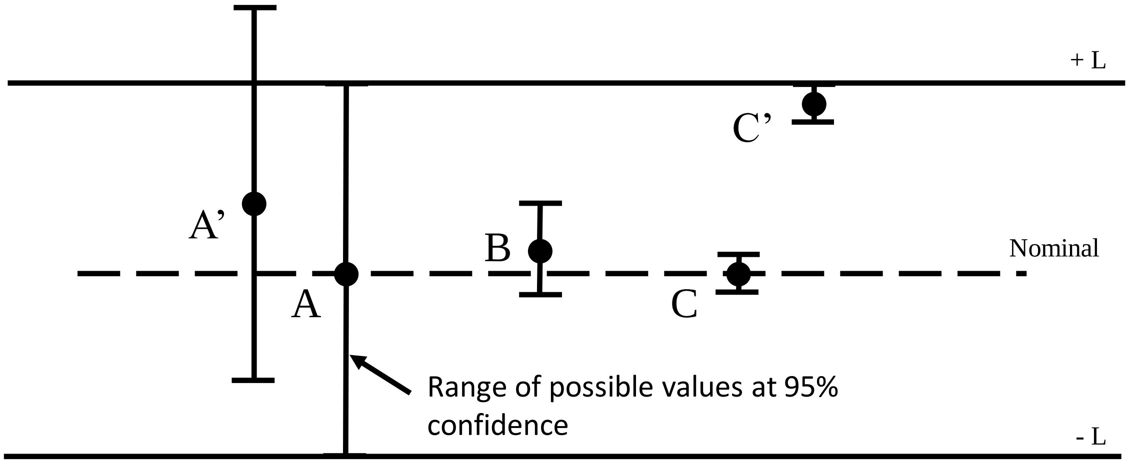
7.6.2 A laboratory performing calibrations, including of its own equipment, shall evaluate the measurement uncertainty for all calibrations.

7.6.3 A laboratory performing testing shall evaluate measurement uncertainty. Where the test method precludes rigorous evaluation of measurement uncertainty, an estimation shall be made based on an understanding of the theoretical principles or practical experience of the performance of the method.

Introduction - Measurements, Uncertainty, and Specifications

Measurement Uncertainty: The doubt that exists about a measurement's result

- Every measurement—even the most careful—always has a margin of doubt
- Uncertainty is the inherent limitation of a measurement process, due to instrumentation and process variation
- Measurement uncertainty does not include mistakes



Measurement Uncertainty

CMC is defined as Calibration and Measurement Capability. It often includes the following standard uncertainty contributors:

- Repeatability
- Resolution
- Reproducibility
- Reference Standard Uncertainty
- Reference Standard Stability
- Environmental Factors

7.6.1 Laboratories shall identify the contributions to measurement uncertainty. When evaluating measurement uncertainty, all contributions that are of significance, including those arising from sampling shall be taken into account using appropriate methods of analysis.

Measurement Uncertainty

Let us examine CMC (Calibration Measurement Capability) using a primary standard as the reference and how it affects the Expanded Uncertainty. A **Primary Standard as the Reference (CMC 0.0016 % for *k* = 2 or 0.16 lbf @ 10K)**

Measurement Uncertainty Budget Worksheet									
Laboratory	Morehouse Primary Standards								
Parameter	FORCE	Range	10K	Sub-Range					
Technician	HZ	Standards Used							
Date									
Uncertainty Contributor	**Magnitude**	**Type**	**Distribution**	**Divisor**	**df**	**Std. Uncert**	**Variance (Std. Uncert^2)**	**% Contribution**	**u^4/df**
Reproducibiliy	000.0000E+0	A	Normal	1.000	10	000.00E+0	000.00E+0	0.00%	000.0E+0
Repeatability	57.7350E-3	A	Normal	1.000	5	57.74E-3	3.33E-3	7.51%	2.2E-6
U-7643 LLF	65.0000E-3	A	Normal	1.000	200	65.00E-3	4.23E-3	9.52%	89.3E-9
Resolution of UUT	100.0000E-3	B	Resolution	3.464	200	28.87E-3	833.33E-6		
Environmental Conditions	75.0000E-3	B	Rectangular	1.732	200	43.30E-3	1.88E-3		
Stability of Ref Standard	288.0000E-3	B	Rectangular	1.732	200	166.28E-3	27.65E-3		
Ref Standard Resolution	24.0000E-3	B	Resolution	3.464	200	6.93E-3	48.00E-6		
			None	0.000					
Morehouse CMC	160.0000E-3	B	Expanded (95.45% k=2)	2.000	200	80.00E-3	6.40E-3	14.43%	204.8E-9
			Combined Uncertainty (u_c)=			**210.62E-3**	**44.36E-3**	**100.00%**	**6.4E-6**
			Effective Degrees of Freedom			**309**			
			Coverage Factor (k) =			**1.97**			
			Expanded Uncertainty (U) K =			**0.41**	**0.00414%**		

14.43 % Contribution

Measurement Uncertainty

Let’s examine CMC (Calibration Measurement Capability) using a secondary standard as the reference and how it affects the Expanded Uncertainty.

Accredited Calibration Supplier with Secondary Standards as the Reference (CMC 0.04 % for *k* = 2 or 4 lbf)

Measurement Uncertainty Budget Worksheet									
Laboratory	Morehouse Primary Standards								
Parameter	FORCE	**Range**	10K	**Sub-Range**					
Technician	HZ	**Standards Used**							
Date									
Uncertainty Contributor	**Magnitude**	**Type**	**Distribution**	**Divisor**	**df**	**Std. Uncert**	**Variance (Std. Uncert^2)**	**% Contribution**	**u^4/df**
Reproducibiliy	000.0000E+0	A	Normal	1.000	10	000.00E+0	000.00E+0	0.00%	000.0E+0
Repeatability	378.5939E-3	A	Normal	1.000	5	378.59E-3	143.33E-3	3.43%	4.1E-3
U-7643 LLF	65.0000E-3	A	Normal	1.000	200	65.00E-3	4.23E-3	0.10%	89.3E-9
Resolution of UUT	100.0000E-3	B	Resolution	3.464	200	28.87E-3	833.33E-6		
Environmental Conditions	75.0000E-3	B	Rectangular	1.732	200	43.30E-3	1.88E-3		
Stability of Ref Standard	288.0000E-3	B	Rectangular	1.732	200	166.28E-3	27.65E-3		
Ref Standard Resolution	24.0000E-3	B	Resolution	3.464	200	6.93E-3	48.00E-6		
			None	0.000					
Accredited Cal Supplier CMC	4.0000E+0	B	Expanded (95.45% k=2)	2.000	200	2.00E+0	4.00E+0	95.74%	80.0E-3
			Combined Uncertainty (u_c)=			**2.04E+0**	**4.18E+0**	**100.00%**	**84.1E-3**
			Effective Degrees of Freedom			**207**			
			Coverage Factor (k) =			**1.97**			
			Expanded Uncertainty (U) K =			**4.03**	**0.04030%**		

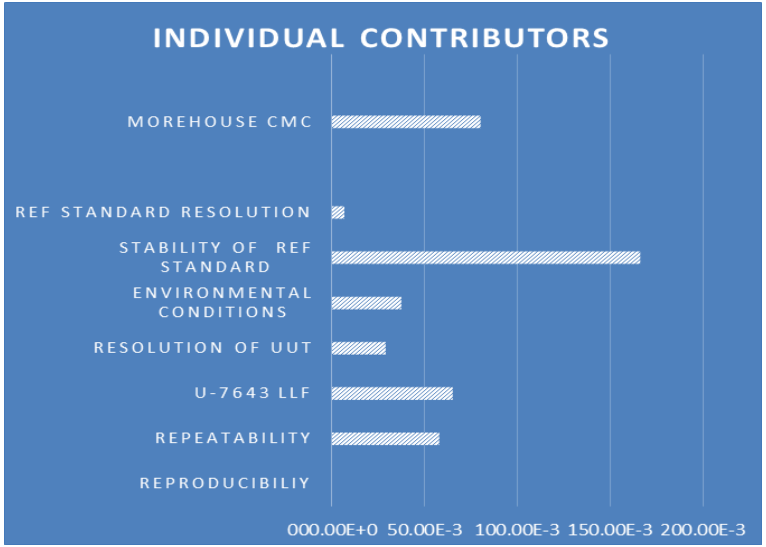
95.74 % Contribution

Measurement Uncertainty

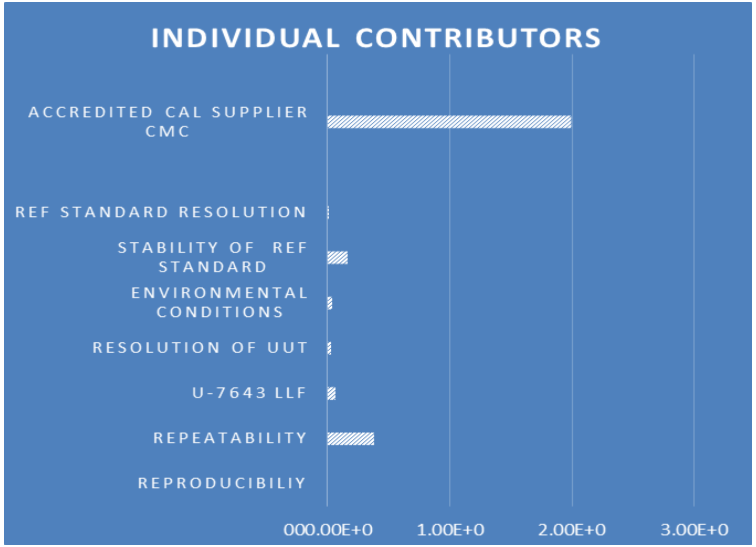
Let's examine CMC (Calibration Measurement Capability) and what the Reference CMC does to the calibration results. Deadweight Primary Standard Versus Secondary Standards

Expanded Uncertainty when calibrated with Primary Standards is approximately 10 times lower than using secondary standards

Expanded Uncertainty @ 10K = 0.41 lbf
Morehouse CMC = 0.16 lbf
Repeatability = 0.057 lbf



Expanded Uncertainty @ 10K = 4.03 lbf
Accredited Cal Supplier CMC = 4.00 lbf
Repeatability = 0.379 lbf



Metrological Traceability Review

Diameter Tolerance Chart per (ANSI/ASME B89.1.5)					
Diameter Range	"XX"	"X"	"Y"	"Z"	"ZZ"
Above - Including		*Inch*			
.010" - .825"	.000020"	.000040"	.000070"	.00010"	.00020"
.825" - 1.510"	.000030"	.000060"	.000090"	.00012"	.00024"
1.510" - 2.510"	.000040"	.000080"	.00012"	.00016"	.00032"
2.510" - 4.510"	.000050"	.00010"	.00015"	.00020"	.00040"
4.510" - 6.510"	.000065"	.00013"	.00019"	.00025"	.00050"
6.510" - 9.010"	.000080"	.00016"	.00024"	.00032"	.00064"
9.010" - 12.010"	.00010"	.00020"	.00030"	.00040"	.00080"

Metrological Traceability Review

- The resolution is1 µ inch
- Master thread gauge had Measurement Uncertainty of 35 µ inch
- Gage R & R was 7 µ inch

What would be the minimum uncertainty?

Metrological Traceability Review

$$Resolution(R) = \frac{0.000\,001}{\sqrt{12}}$$

$$Master\ Thread\ Gauge(MTG) = \frac{0.000\,035}{2}$$

$$Gage\ R\&R(G) = 0.000\,007$$

$$Uc = \sqrt{R^2 + MTG^2 + G^2}$$

$$Uc95 = 2 \times Uc$$

$$Uc95 \cong 37.7\mu in$$

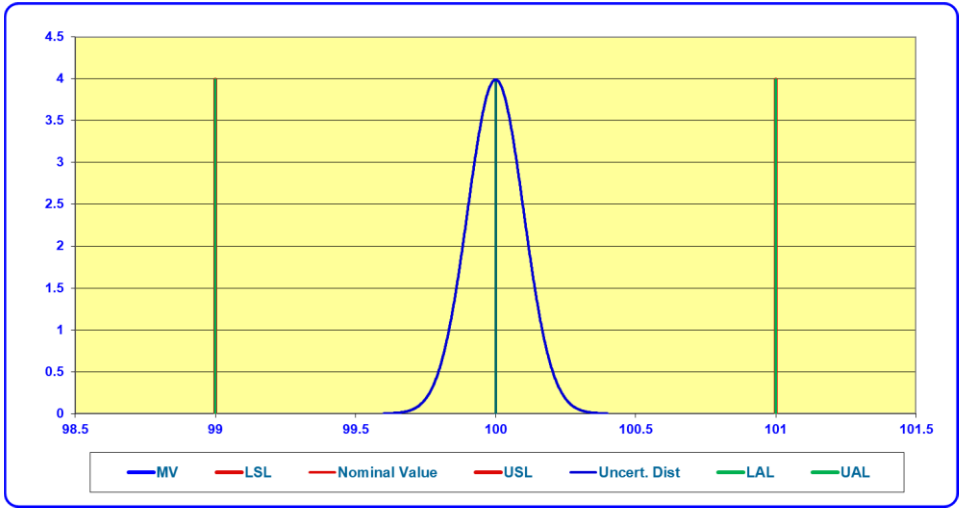
Tolerances

Tolerance: the total amount by which a specific characteristic is permitted by specifications to vary.

NOTE: The tolerance is the difference between the upper and lower specification limits tolerance interval: region between, and including, the tolerance

Tolerance limits: specified values of the characteristic, giving upper and/or lower bounds of the **permissible value**

A measurement quantity of 100 Volts has a tolerance of ±1 Volt. The measurement process used for calibration has an estimated 95 % expanded uncertainty of 0.2 Volts.



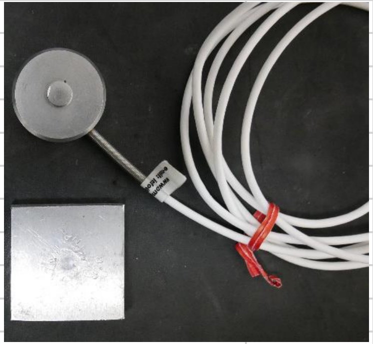
Tension Links – Tolerance?
PROPER PIN DIAMETER

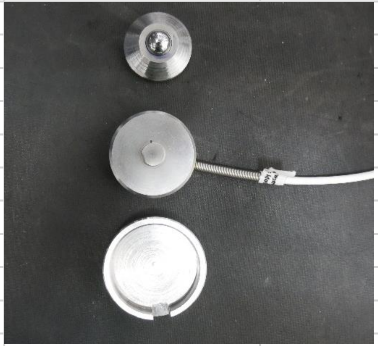
Pin B (2.0030 to 2.0060)	Pin A (2.0005 to 2.0045)
50,070	50,010
50,050	50,020
50,040	50,010
50,070	50,020
50,090	50,020
50,060	50,030
50,080	50,010
50,070	50,030
50,090	50,020
50,090	50,070
50,080	50,060
50,100	50,070
17.81640375	22.74696117
Out of 24 tests 13 did not meet spec ± 50	

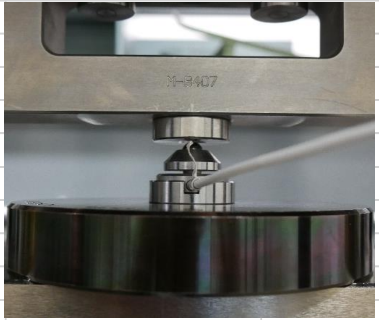
The Problem With Averages

On average the results look similar!

Standard Setup versus Morehouse Adapters in Morehouse Deadweight



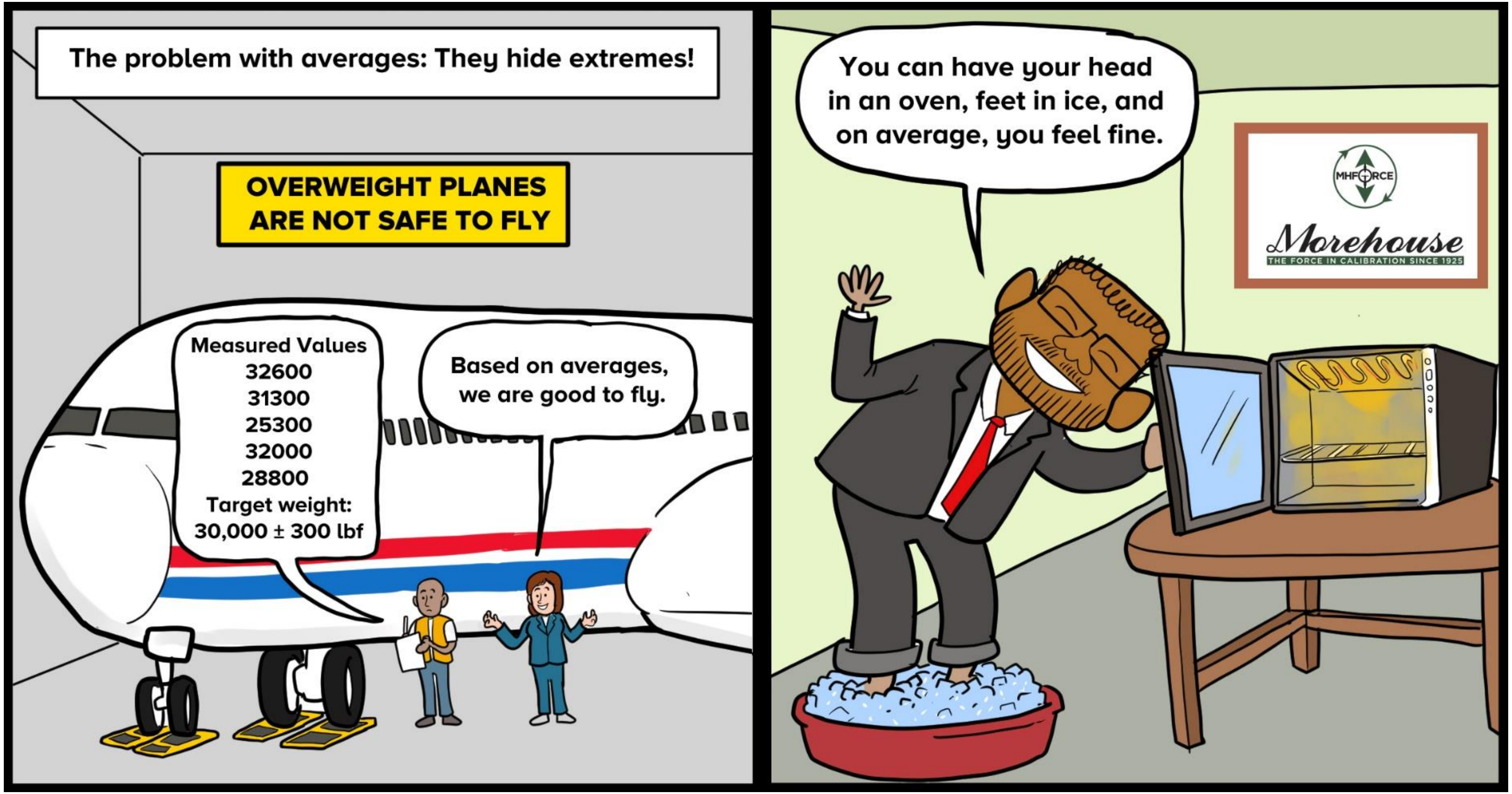




Manually Aligned	Data
0 degree	2011
120 degree	1997
240 degree	2018
Average	2008.66667
Standard Deviation	10.6926766
Max Deviation	21
% Error	1.045%

Aligned with Adapter	Data
0 degree	2008
120 degree	2006
240 degree	2010
Average	2008
Standard Deviation	2
Max Deviation	4
% Error	0.199%

The Problem with Averages

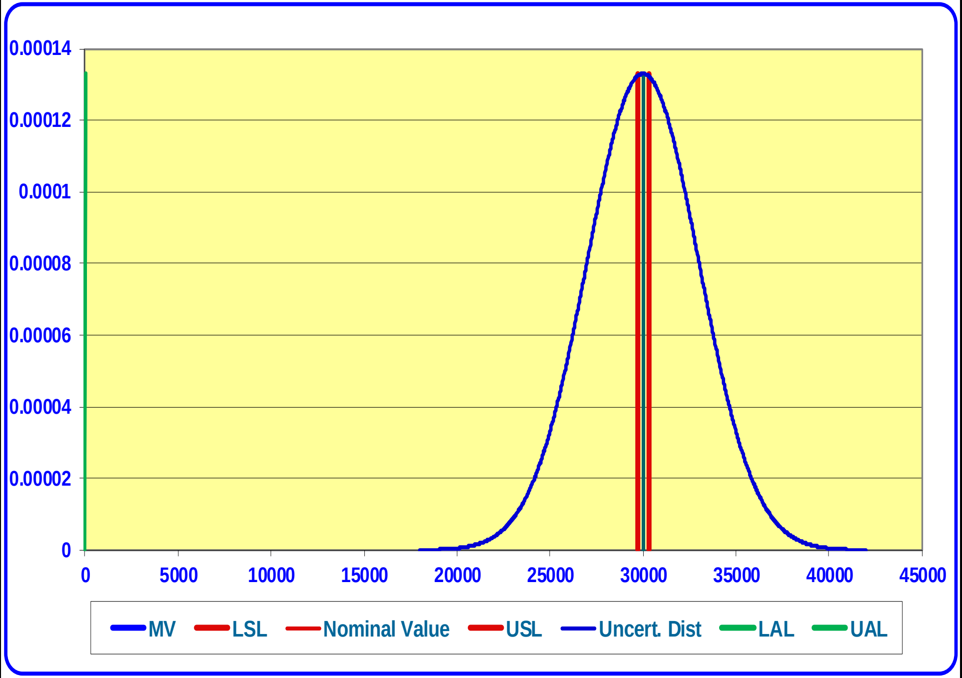


The Problem with Averages

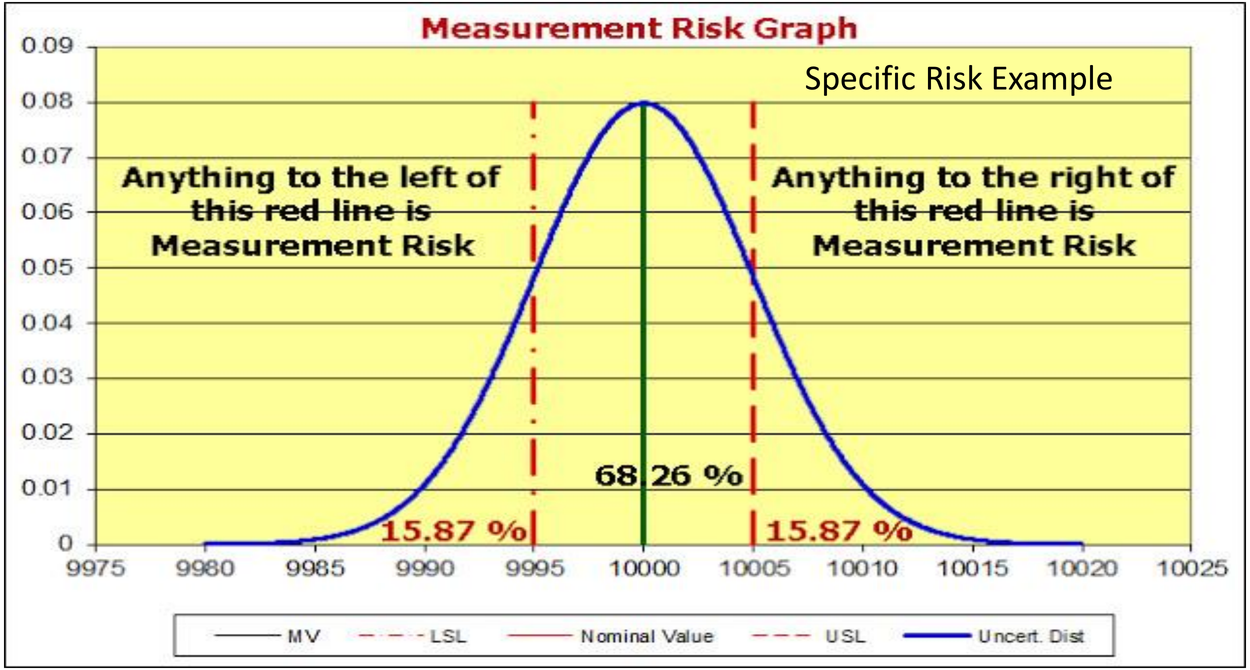
A1	B	C	D	SELECTION OF GUARDBAND METHOD				
2	E = mc^3 Solutions	Reported Result	Acceptance Limit	Choose Decision Rule >>>>>>>	LAC G8:2009 Decision Rule(95%	########	Custom Acceptance Limit	2
3	Nominal Value	30000						
4	Lower Specification Limit	29700	INDETERMINATE					
5	Upper Specification Limit	30300	INDETERMINATE					
6	Measured Value	30000.0000	30000.00					
7	Std. Uncert. (k=1)	3.00E+3						
8	Total Risk	92.032%						
9	Upper Limit Risk	46.016%						
10	Lower Limit Risk	46.016%						
11	Test Uncertainty Ratio (TUR) =	0.05						
12	Process Capability (C_{pk})	0.033						

Area below for calculations

	Sample	Measurement	
	1	32600.0	
	2	31300.0	
	3	25300.0	
	4	32000.0	
	5	28800.0	
	Sample Mean	30000.00	
	Sample Standard Deviation	2999.17	



Measurement Decision Risk



ISO/IEC 17025: 2017 Section 3.7 defines a decision rule as a rule that describes how measurement uncertainty is accounted for when stating conformity with a specified requirement.

A calibration laboratory cannot make a statement of conformity or "Pass" an instrument without violating ISO/IEC 17025:2017, as section 3.7 defines a Decision Rule as a rule that describes how <u>measurement uncertainty is accounted</u> for when stating conformity with a specified requirement. Some may argue that you can <u>take it into account by ignoring it.</u>

To that end, can we all decide to take all red stoplights into account and start ignoring them?

- UKAS LAB 48 Decision Rules and Statements of Conformity

Types of Risk (Errors)

Type I - Type II Error			
		Calibration	
		In Tolerance (GOOD)	Out Of Tolerance (BAD)
Decision Made	Called In Tolerance - ACCEPT	(1-α) *Calibration Lab's Confidence (Probability of Correct Accept - PCA)*	β *Type II Error (Probability of False Accept - PFA)*
	Called Out of Tolerance - REJECT	α *Type I Error (Probability of False Reject - PFR)*	(1- β) *End User's Confidence (Probability of Correct Reject - PCR)*

Types of Risk (Errors)

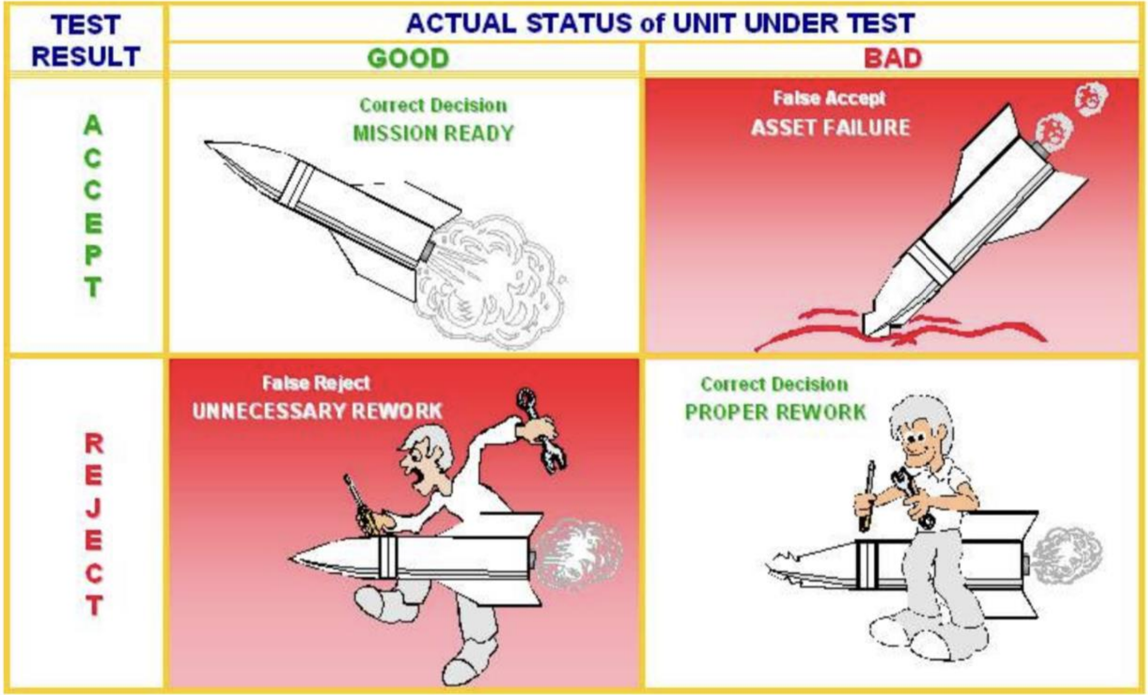


Image from NAVSEA (asq711.org)

Consumer and Producer Risk

There are two general types of risks associated with conformity decisions.

Consumer Risk:

The probability that a non-conforming item is accepted. Also known as Type II error, pass error, false accept risk (FAR), and probability of false acceptance (PFA).

Producer Risk:

The probability that a conforming item is rejected. Also known as Type I error, fail error, false reject risk (FRR), and probability of false reject (PFR).

Consumer risk can have potential negative impacts to <u>product/system</u> performance.

Producer risk has a direct impact on the <u>cost</u> of manufacturing, testing and/or calibration.

Consumer and Producer Risk

Consumer Risk, depending on the criticality of the measurement, can lead to:

- Loss of life or mission
- Reduced end-item function, capacity, or utility
- Warranty expenses
- Damage to corporate reputation
- Loss of future sales
- Punitive damages
- Legal fees, etc

Producer Risk can result in additional costs because of:

- Unnecessary rework, adjustments, repairs, and retests
- Increased scrap of good product
- Increased frequency of inspections or calibrations
- Decreased availability of the hardware
- Out-of-tolerance reports or administrative reaction (reverse traceability reports)

Consumer and Producer Risk

Specific Risk (also called bench-level risk) is based on a specific measurement result.

It triggers a response based on measurement data gathered at time of test.

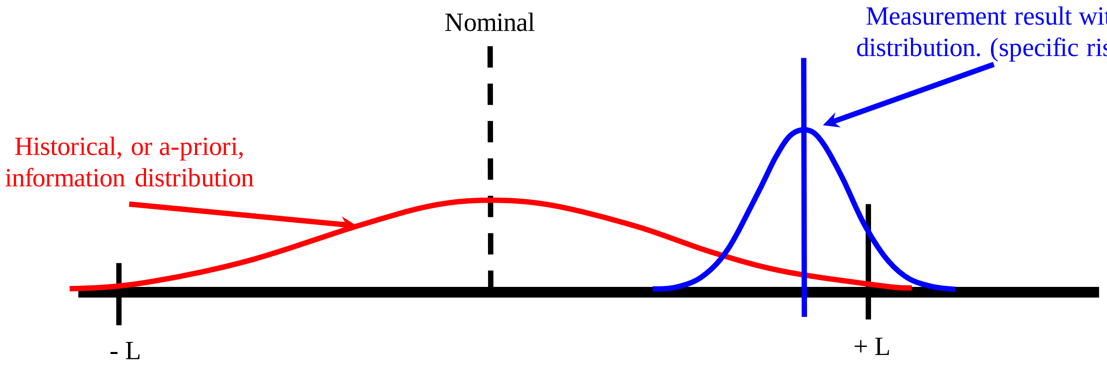
It may be characterized by one or two probability distributions, depending on the method.

Any representation with only one probability distribution is always a specific risk method.

Global Risk (also called process-level risk) is based on a future measurement result.

It is used to ensure the acceptability of a documented measurement process.

It is based on expected or historical information and is usually characterized by two probability distributions.



Consumer Risk (aka PFA)

Consumer Risk could be altered by fine-tuning of calibration system control tools like:

- Measurement reliability
- Calibration intervals
- Calibration process uncertainty
- Calibration adjustments
- Guard-bands

From: Guard-banding Methods-An Overview

Common Definitions ILAC G8

Tolerance Limit (*TL*) (Specification Limit) specified upper or lower bound of permissible values of a property.

Acceptance Limit (*AL*) specified upper or lower bound of permissible measured quantity values.

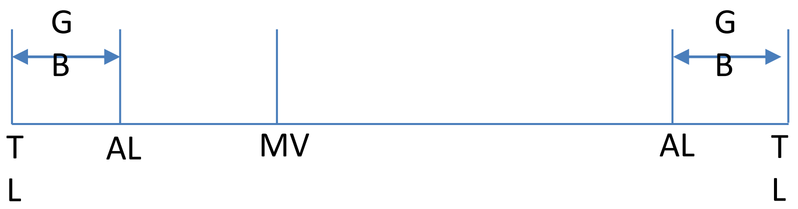
LSL – Lower Specification Limit

USL – Upper Specification Limit

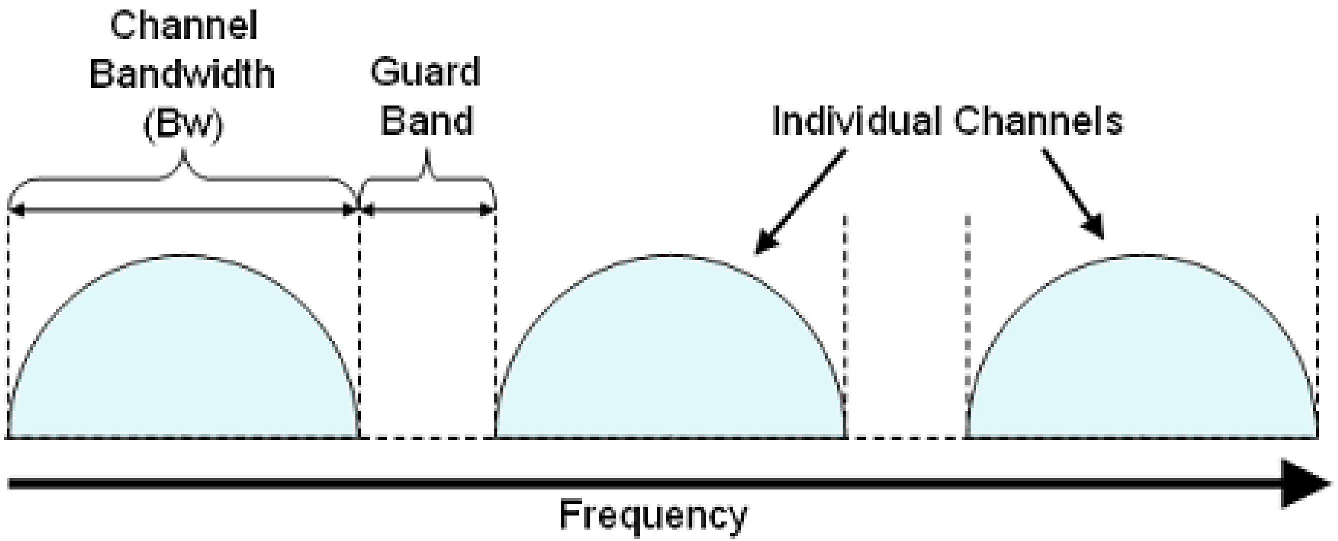
Measured Quantity Value quantity value represents a measured result.

Guard Band (*w*) interval between a tolerance limit and a corresponding acceptance limit where length $w=|TL-AL|$.

Decision Rule describes how measurement uncertainty is accounted for when stating conformity with a specified requirement. *(ISO/IEC 17025:2017 3.7 a rule that describes how measurement uncertainty will be accounted for when stating conformity with a specified requirement).*

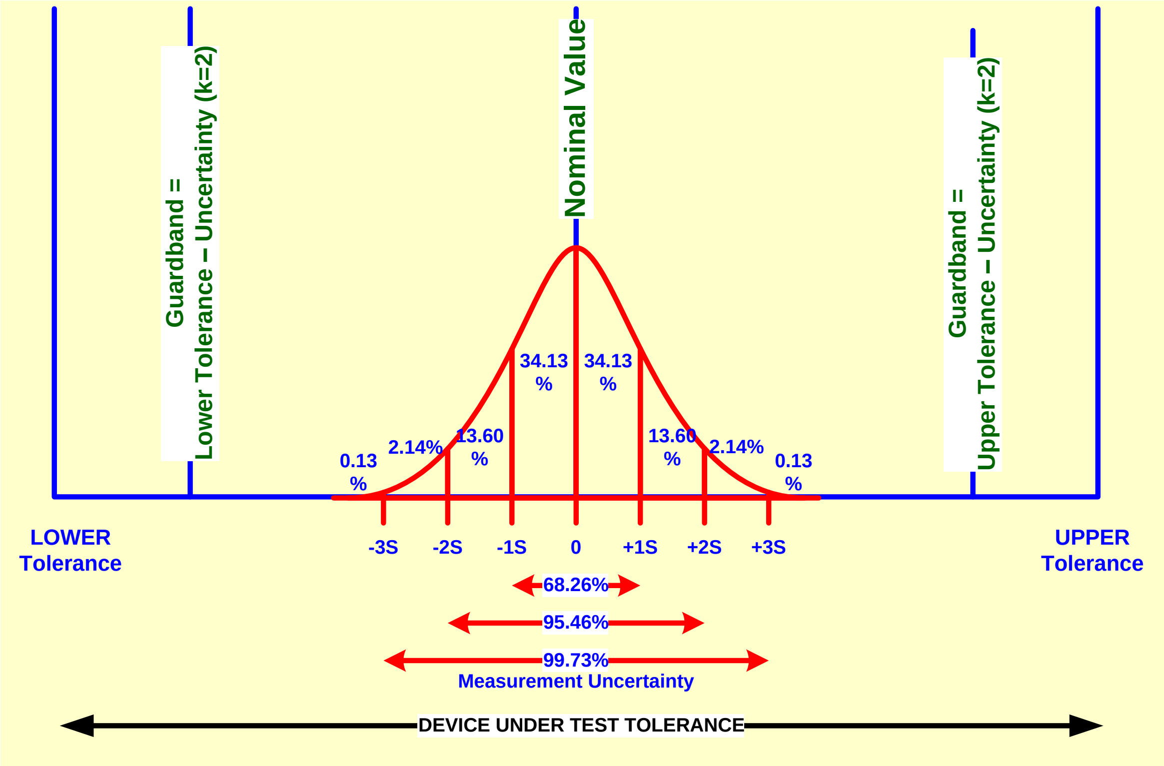


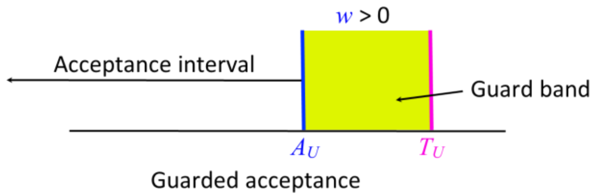
Guard Banding

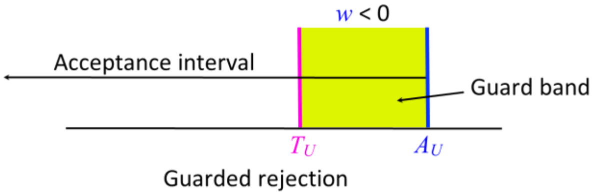


Bandwidth (Bw) = 2 / Symbol Rate (Rs)

Instrument Measurement Uncertainty Guard Banding







Statement of Conformity

When performing a measurement and subsequently making a statement of conformity, for example, in or out-of-tolerance to the manufacturer's specifications or Pass/Fail to a particular requirement, there are two possible outcomes:

a. The result is reported as conforming with the specification

b. The result is reported as not conforming with the specification

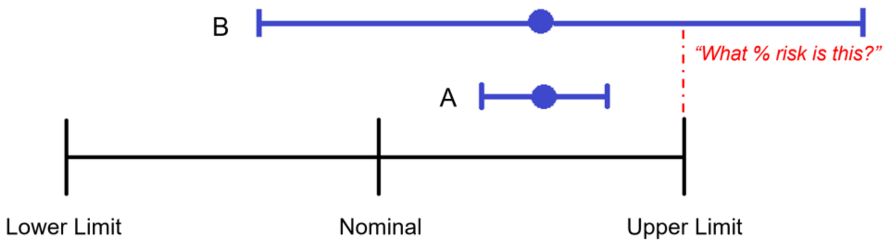


Illustration of Measurement Decision Risk

Measurement Uncertainty in Conformity Assessment

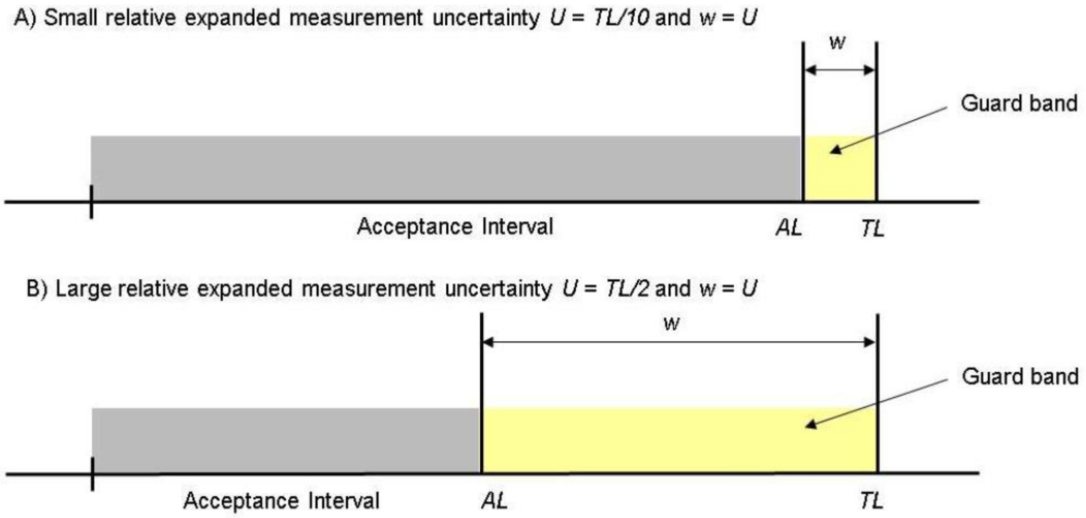
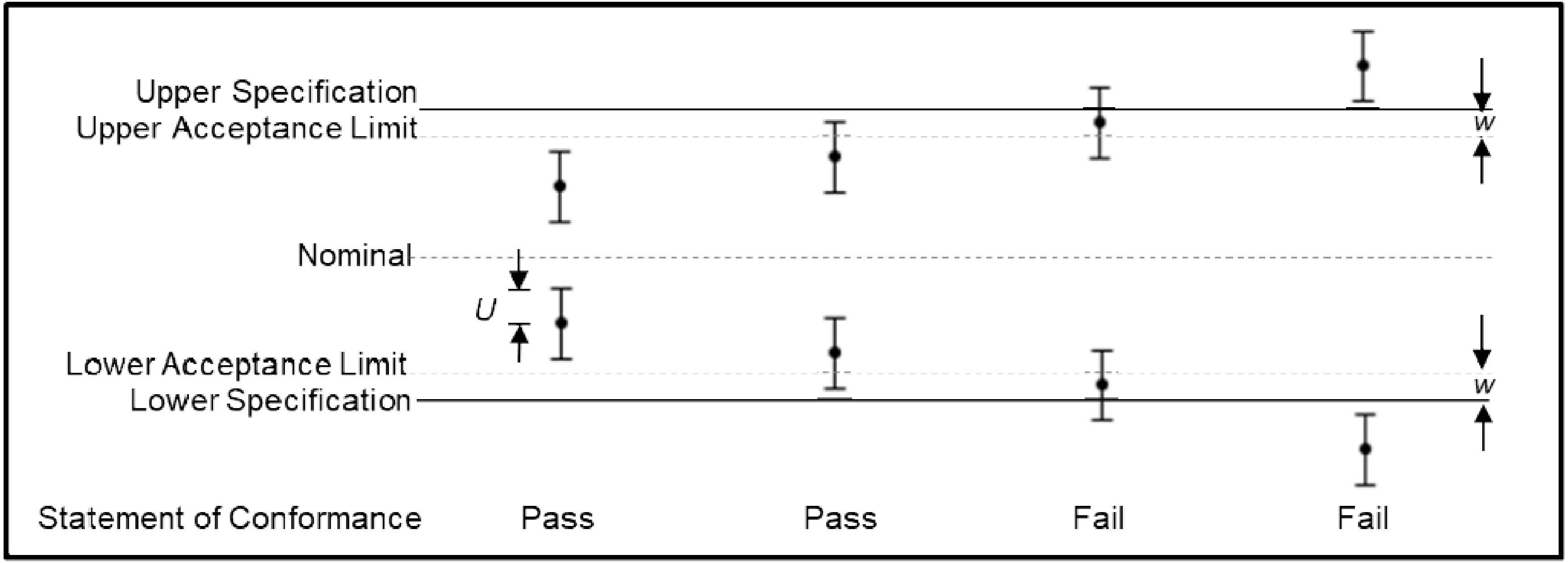


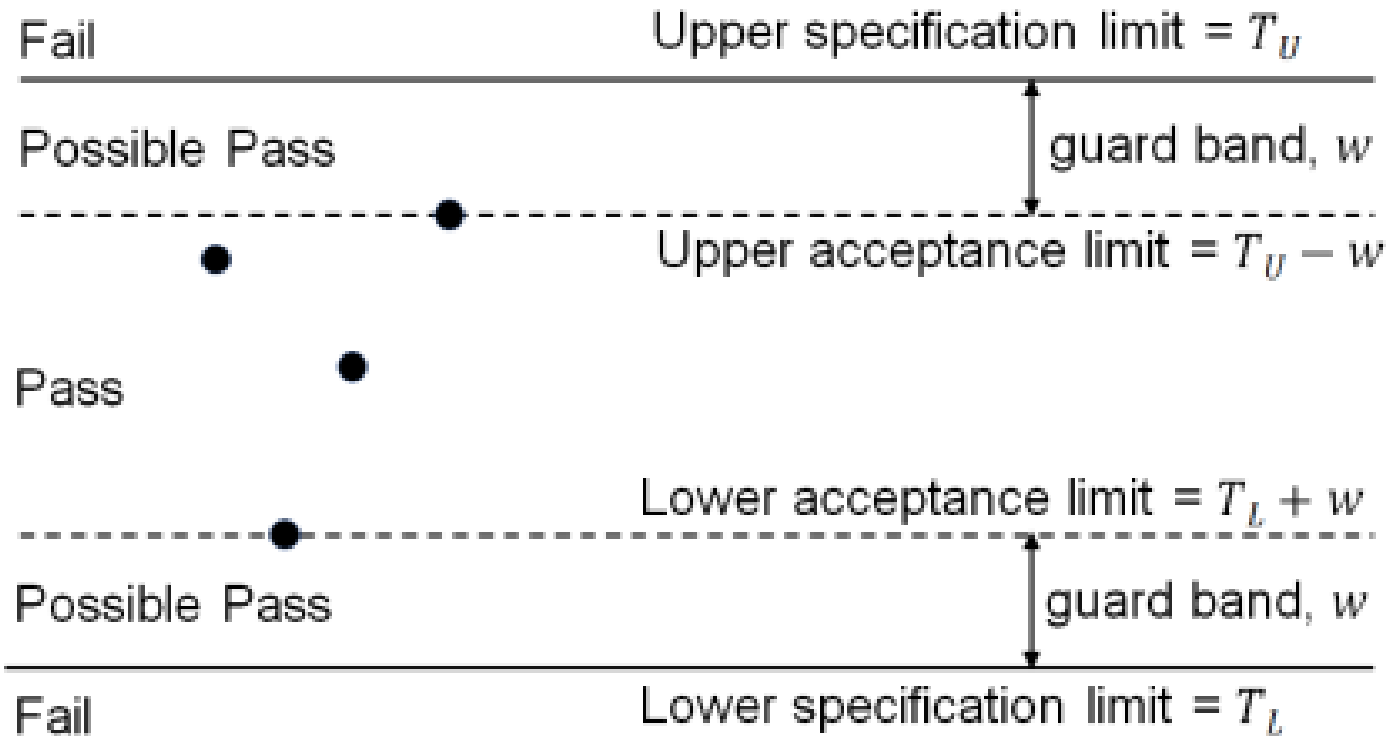
Figure 6 Acceptance interval for a case where expanded measurement uncertainty is small compared to tolerance A) and large B) for the same tolerance limit *TL*. A large guard band narrows the distribution function of accepted items.

Binary Statement with Guard Band



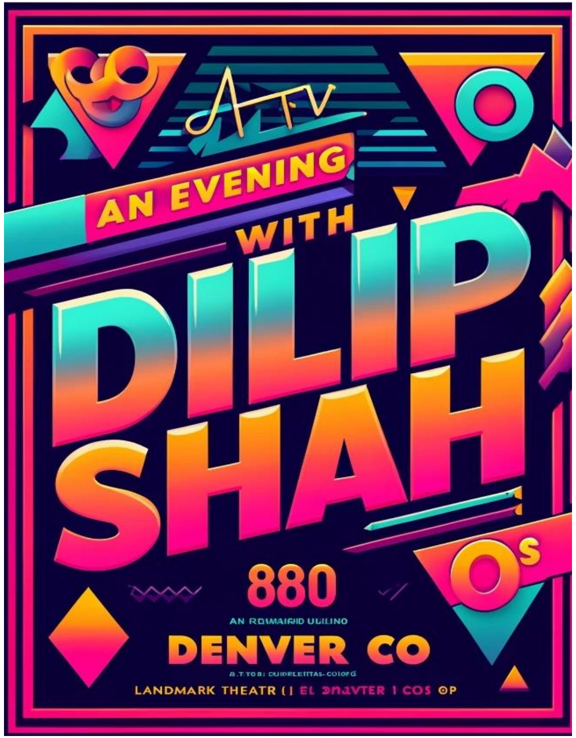
U = 95% expanded measurement uncertainty

Non-Binary Statement with Guard Band



Example from UKAS LAB 48

Binary or Non-Binary Example



Tickets are free, Tonight at 7:00 PM at Landmark Theatre, Transportation from the Hotel is Provided, Open Bar and Light Fare.

Are you going to go to this event?

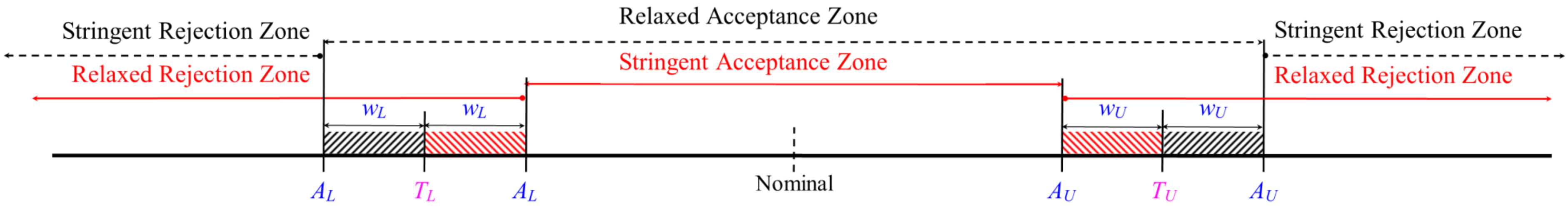
Your possible answers right now are what?

Yes Maybe No

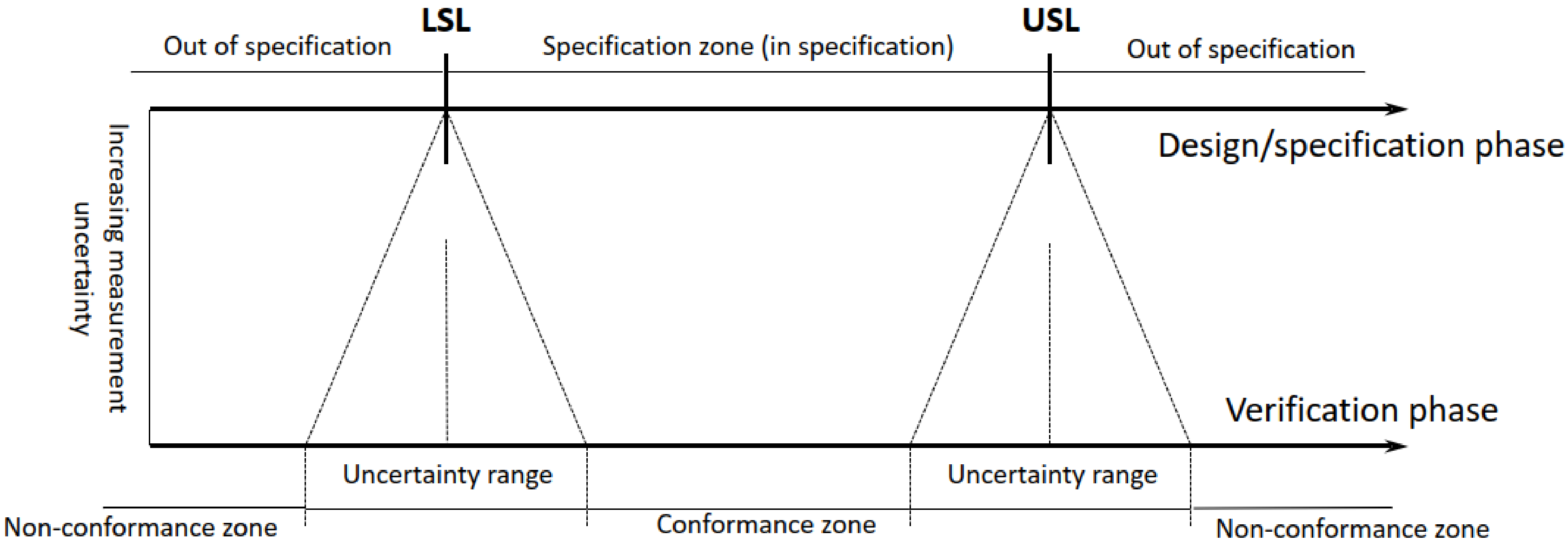
There is no Maybe Yes or Maybe No, it’s Maybe

Eventually, you have only two outcomes. You either went or you didn’t

ASME B89.7.3.1-2001 – Specific Risk



The Size of Acceptance limits is Determined by the Measurement Uncertainty and Desired Risk Level.



ISO 14253-1:2017

Types of Risk Scenarios

ASME B89.7.4.1-2005 describes both risk levels well

Specific Risk mitigation can be thought of as "controlling the quality of the workpieces," while **Program Level Risk** strategies are described as "controlling the average quality of workpieces."

Specific Risk being instantaneous liability at the time of the measurement and program level is more about the average probability that incorrect acceptance decisions will be made based on historical data

Specific Risk

Specific Risk (sometimes called bench-level risk) is based on a specific measurement result.

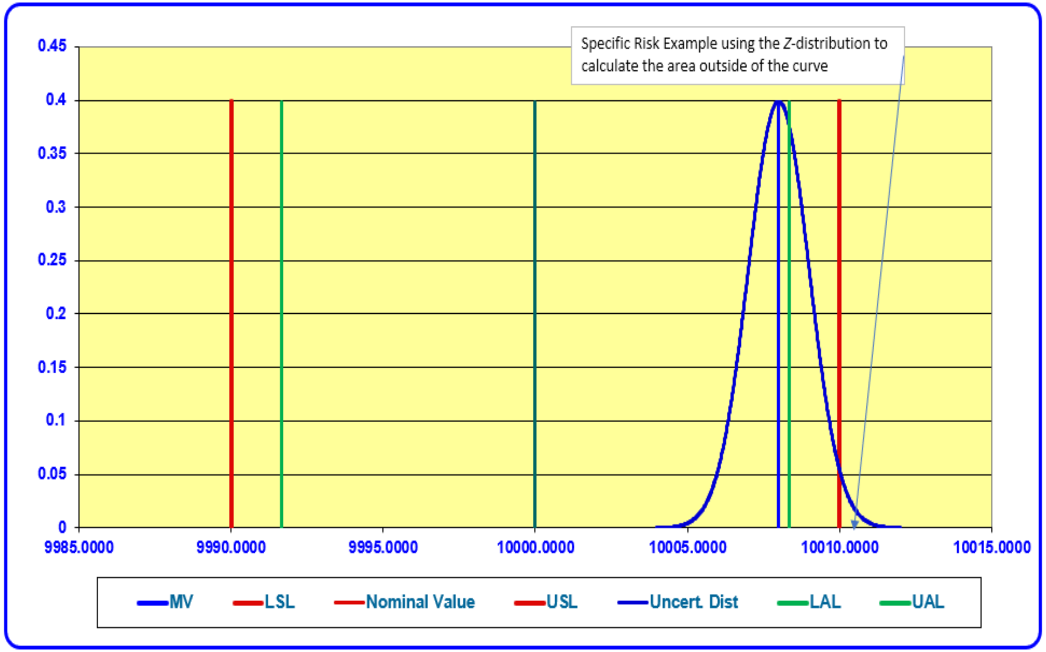
- It triggers a response based on measurement data gathered at the time of the test.
- It may be characterized by one or two probability distributions, depending on the method.
- Any representation with only one probability distribution is always a specific risk method.

Specific risk is after a measurement is made and Global risk is for future measurements

Measurement Decision Risk

Risk Calculator		
Upper Tolerance T_U	10010.0000	
Lower Tolerance T_L	9990.0000	
Nominal Value	10000.0000	
Measurement Unc um	1.0000	
Measured Value xm	10008.0000	
Tolerance T	20.00	Area Outside of USL
Z Upper	2.00	2.275%
		Area Outside of LSL
Z Lower	-18.00	0.000%

Setting the Guard Band Upper and Lower p_c	
Select Desired Conformance Probability	95.00%
Maximum Risk if within G_U & G_L	5.00%
h U (GB Multiplier)	0.8224
Guard Band Upper G_u	10008.3551
Guard Band Lower G_L	9991.6449



A customer writes a PO that states: Please calibrate "As Found" Manufacturer 10,000 N Load cell S/N XXXX with indicator Manufacturer Readout XXXX to 10,000 N in Compression only and issue a "Pass" when the PFA using Specific Risk is ≤ 2.5 %, Otherwise Fail. ☺

Classic 50 % risk scenario with “Simple Acceptance” at the bench level (w = 0), No Guard Band.

Risk Calculator	
Upper Tolerance T_U	1500.2000
Lower Tolerance T_L	1499.8000
Nominal Value	1500.0000
Measurement Unc um	0.0400
Measured Value xm	1500.2000
Tolerance T	0.40

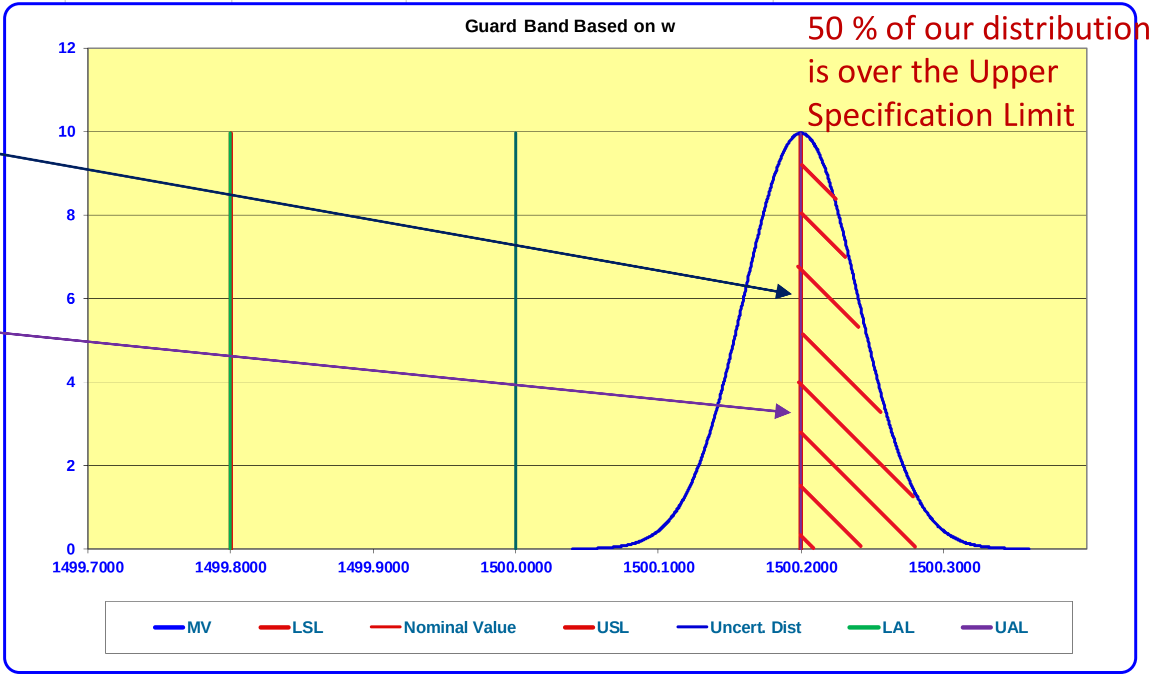
Probability of Conformance (p_c)	50.000%
Probability of NonConformance ($1 - p_c$)	50.000%

Setting the Guard Band Upper and Lower AL	
Guard Band Upper G_u (AL= TL - w)	1500.2000
Guard Band Lower G_L (AL = TL + w)	1499.8000

Setting AL based on Guard Band	
r	0.0000
$w = U_{95} * r$	0.00000
C_m	2.50000

Setting AL based on Guard Band w	
Upper Acceptance Limit	PASS
Lower Acceptance Limit	PASS

Area of Curve Outside of the AL	50.000%



Instrument Measurement Uncertainty
Guard Banding at bench-level ($w = r * U_{95}$)

Risk Calculator	
Upper Tolerance T_U	1500.2000
Lower Tolerance T_L	1499.8000
Nominal Value	1500.0000
Measurement Unc um	0.0400
Measured Value xm	1500.1069
Tolerance T	0.40

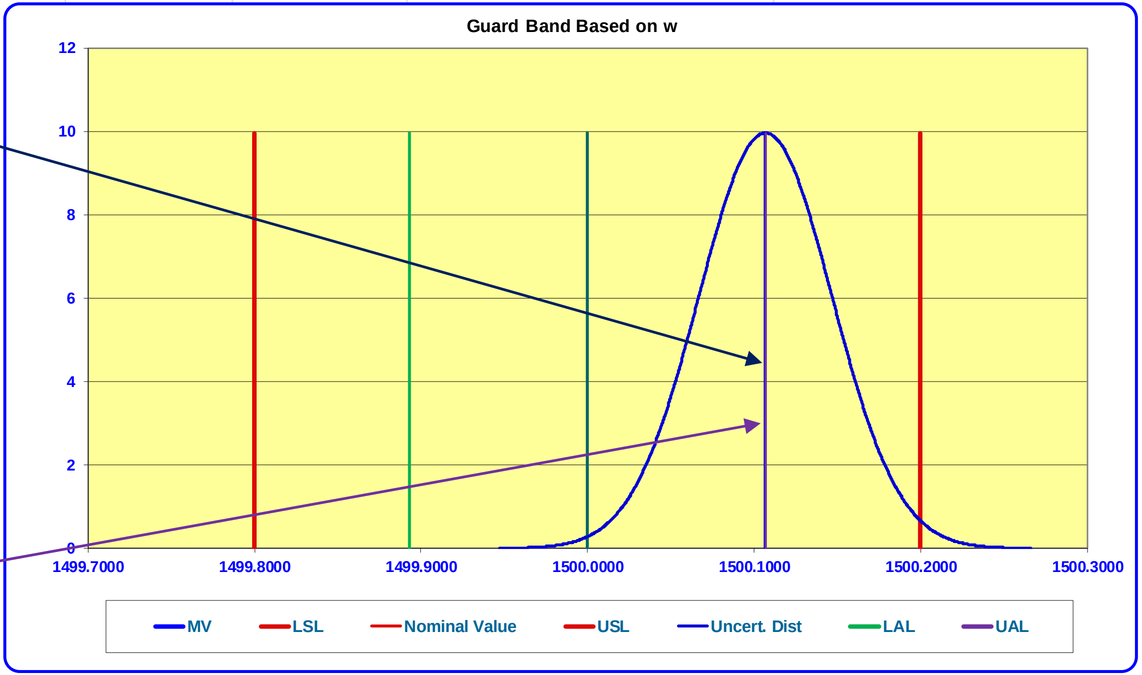
Probability of Conformance (p_c)	99.003%
Probability of NonConformance ($1 - p_c$)	0.997%

Setting the Guard Band Upper and Lower AL	
Guard Band Upper G_U (AL= TL - w)	1500.1069
Guard Band Lower G_L (AL = TL + w)	1499.8931

Setting AL based on Guard Band	
r	1.1632
$w = U_{95} * r$	0.09305
C_m	2.50000

Setting AL based on Guard Band w	
Upper Acceptance Limit	PASS
Lower Acceptance Limit	PASS

Area of Curve Outside of the AL	49.954%



Guard Band of w = 0.09305 (99 % Probability of Conformance) MV = 1500.1069

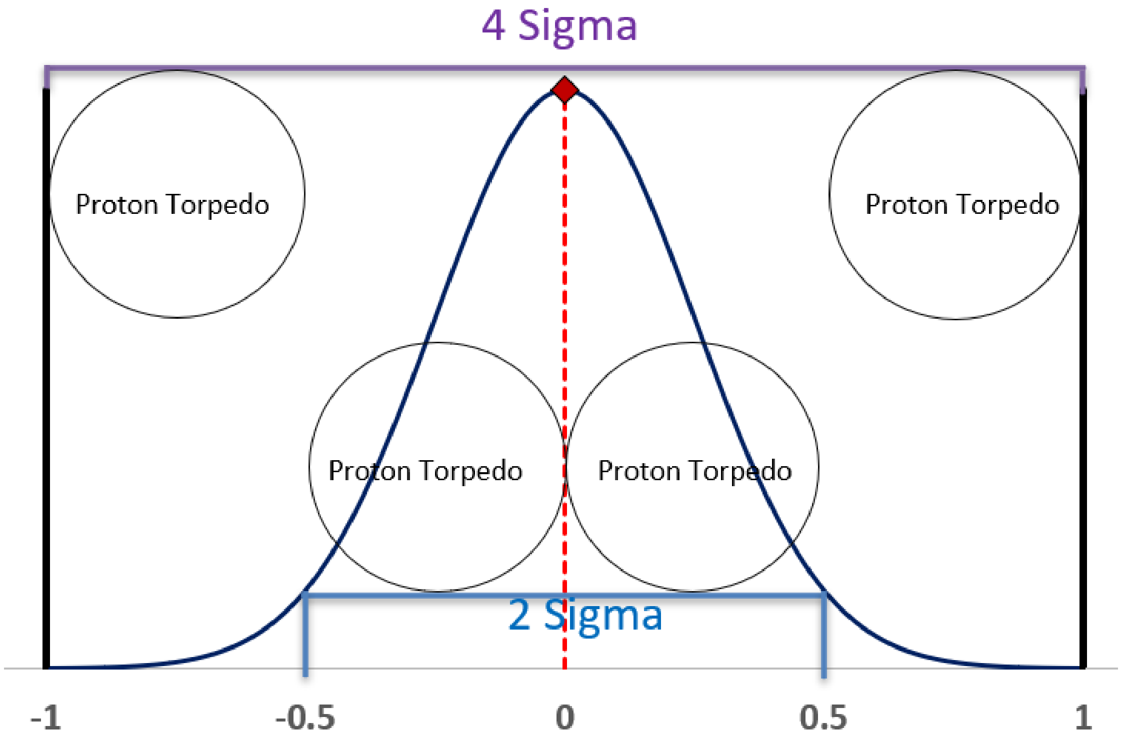
Star Wars Example

With a 2-meter hole and a 0.5-meter Photon Torpedo.

What would be the acceptance limits using a specific risk example?

Knowing our proton torpedo measures 0.5 meters and the empire does not think an X-wing can get close enough to take the shot, we need to devise a plan that will ensure if we can take the shot, we will make it.

A TUR of 4:1 Means 4 Proton Torpedo's Can Fit Between the Tolerance/Specification Limits



Star Wars Example – AL for 2.5 % Maximum risk

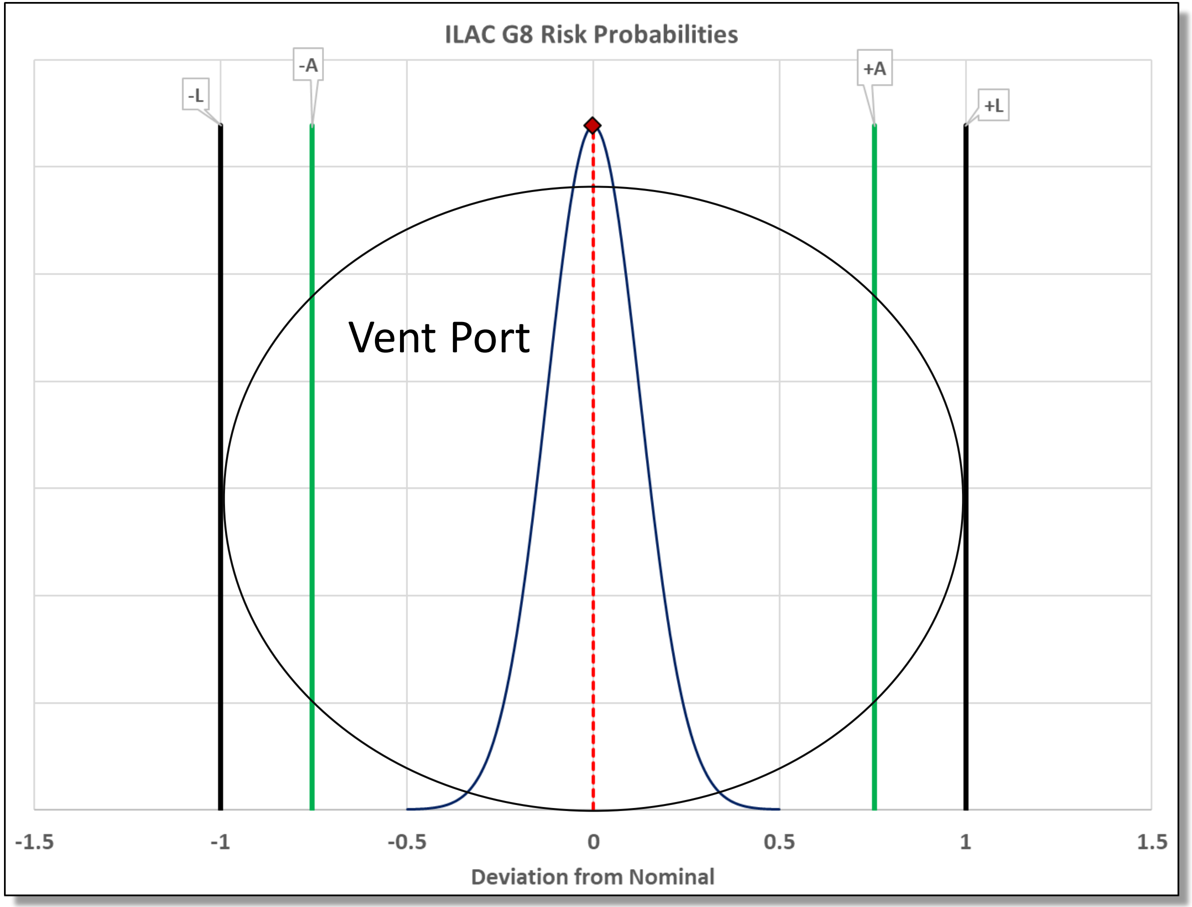
Risk Calculator	
Upper Tolerance T_U	1
Lower Tolerance T_L	-1
Nominal Value (default = blank, otherwise 0)	
Measured Value xm	0.0000
Measurement Unc um	0.1250
Maximum Allowable Risk	2.50%
Tolerance T	2.00

Probability of Conformance (p_c)	100.000%
Probability of NonConformance ($1 - p_c$)	0.000%

Setting the Guard Band Upper and Lower AL	
Guard Band Upper G_u (AL= TL - w)	0.7550
Guard Band Lower G_L (AL = TL + w)	-0.7550

Setting AL based on Probability of Conformance	
Probability of Conformance (pc)	97.50%
r	0.9800
$w = U_{95} * r$	0.24500
C_m (TUR)	4.00000

Setting AL based on Guard Band w	
Upper Acceptance Limit	PASS
Lower Acceptance Limit	PASS



Star Wars Example – Measured Value not Centered

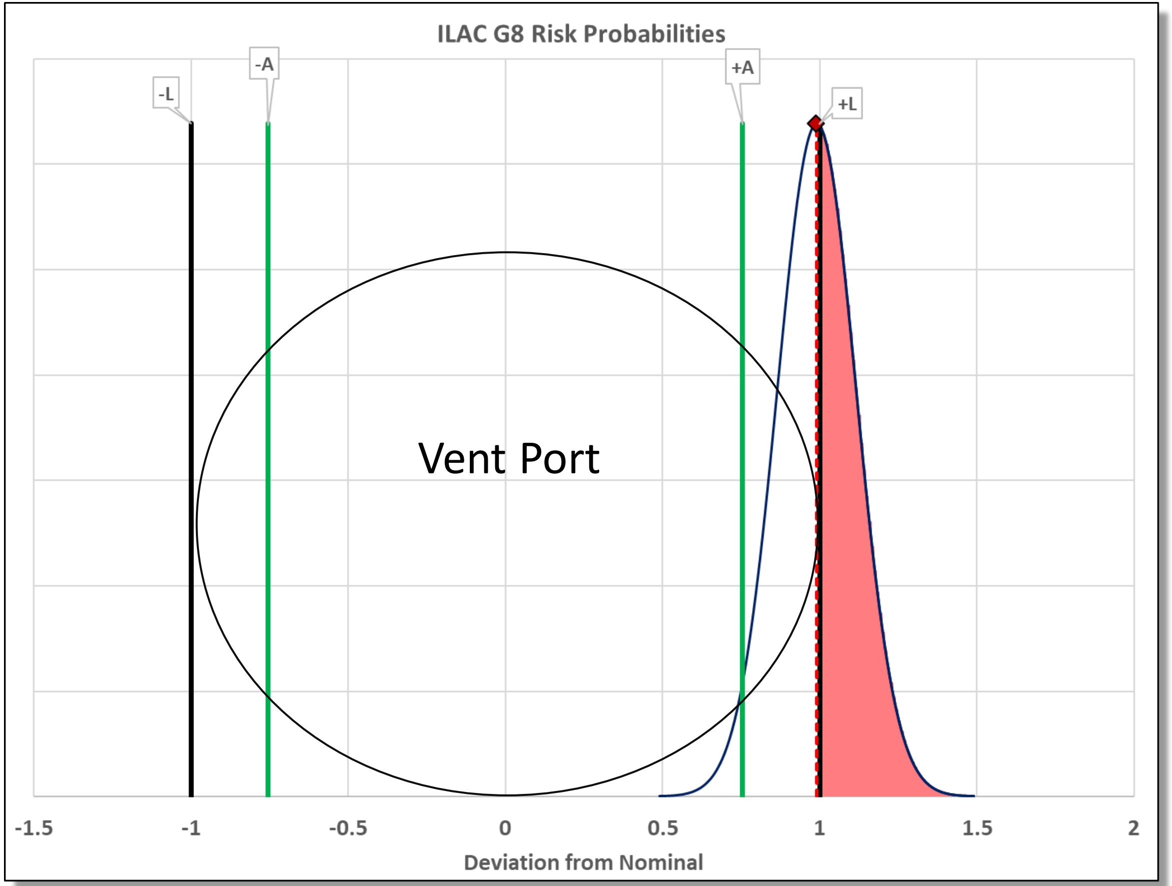
Risk Calculator	
Upper Tolerance T_U	1
Lower Tolerance T_L	-1
Nominal Value (default = blank, otherwise 0)	
Measured Value xm	0.9900
Measurement Unc um	0.1250
Maximum Allowable Risk	2.50%
Tolerance T	2.00

Probability of Conformance (p_c)	53.188%
Probability of NonConformance (1 - p_c)	46.812%

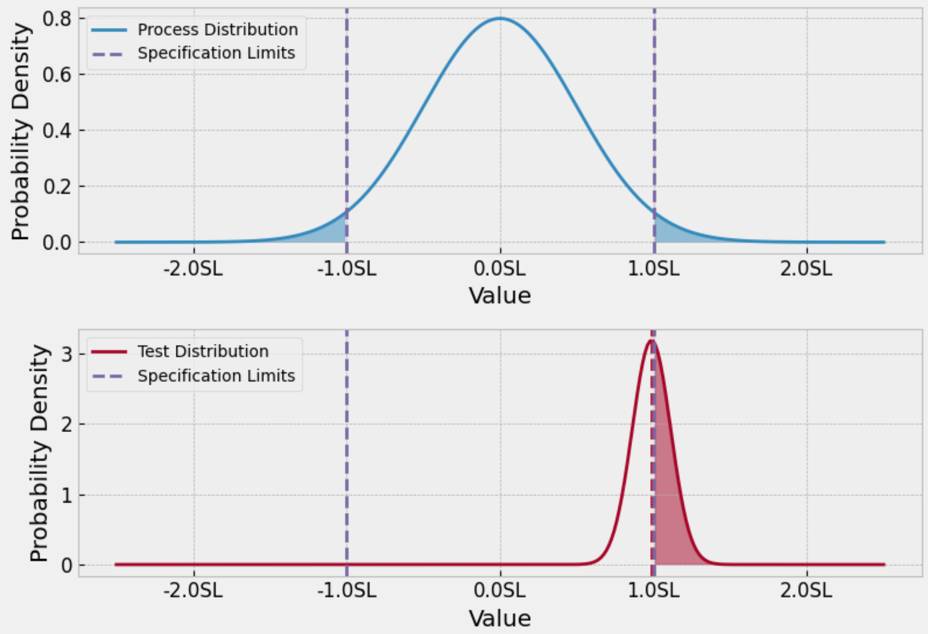
Setting the Guard Band Upper and Lower AL	
Guard Band Upper G_u (AL= TL - w)	0.7550
Guard Band Lower G_L (AL = TL + w)	-0.7550

Setting AL based on Probability of Conformance	
Probability of Conformance (pc)	97.50%
r	0.9800
w = U_{95} * *r*	0.24500
C_m (TUR)	4.00000

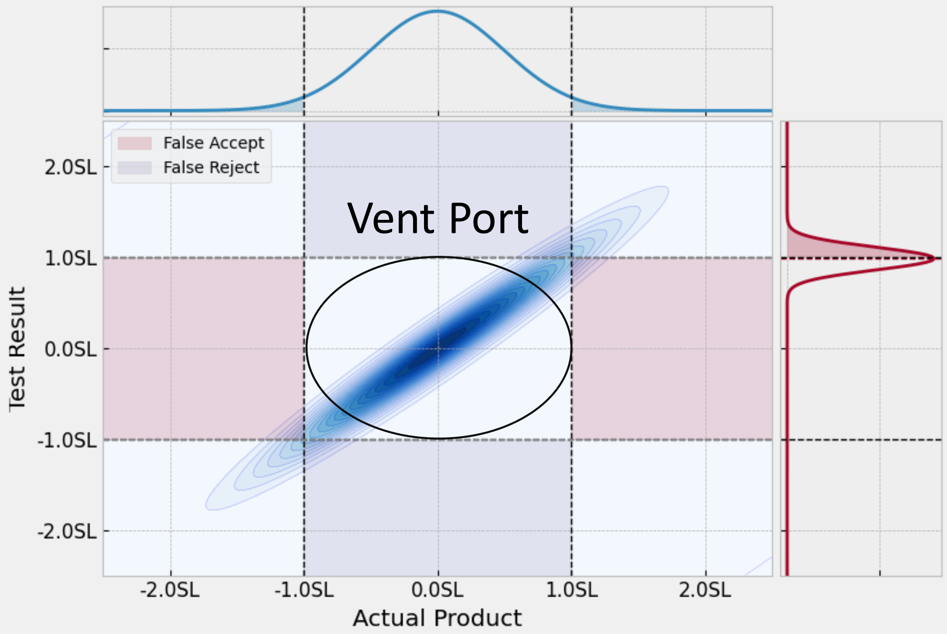
Setting AL based on Guard Band w	
Upper Acceptance Limit	FAIL
Lower Acceptance Limit	PASS



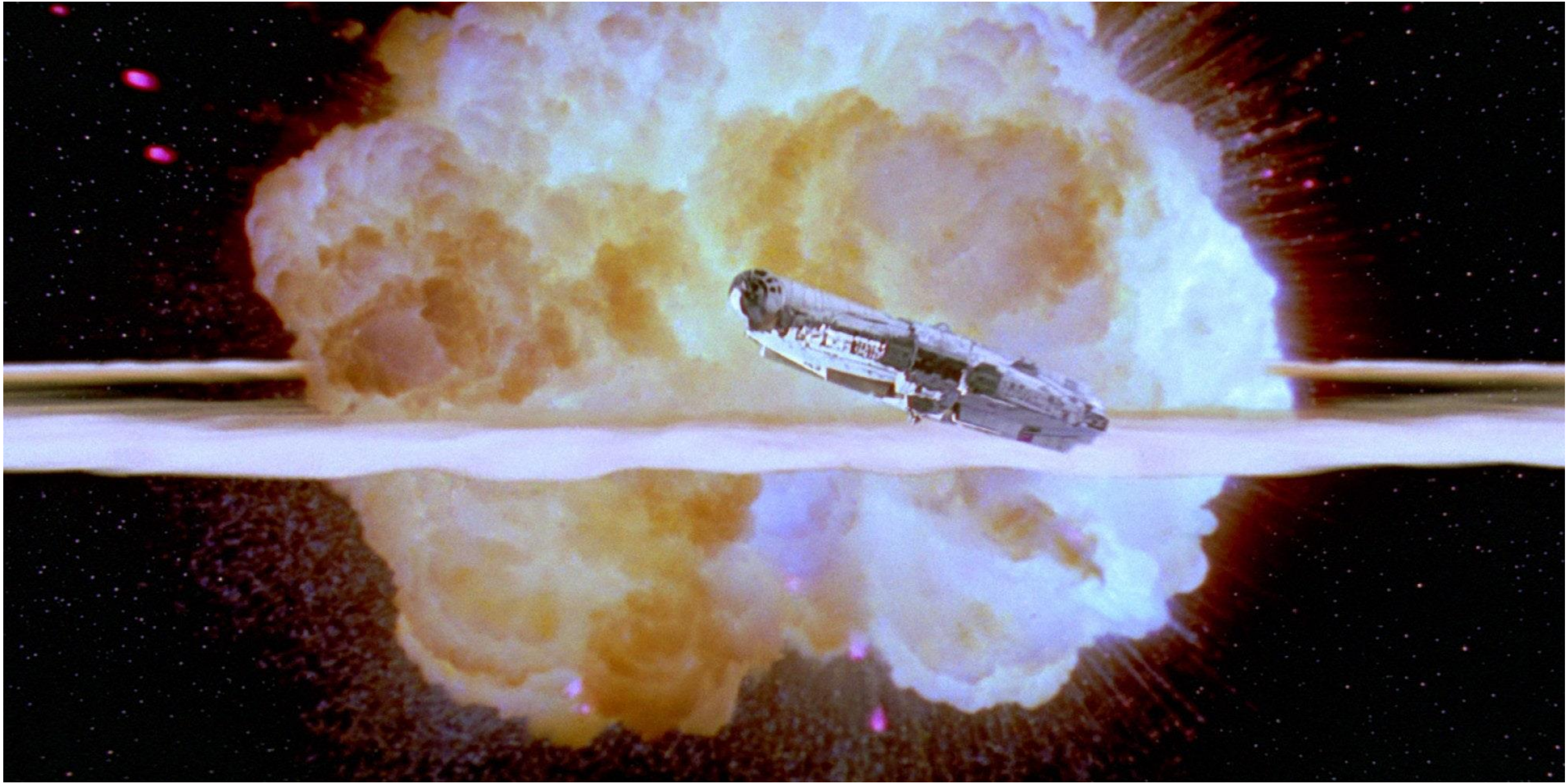
Star Wars Example



Process Risk	Specific Measurement Risk	Global Risk
Process Risk: 4.6%	TUR: 4.0	Total PFA: 0.79%
Upper limit risk: 2.3%	Measured value: 0.99	Total PFR: 1.5%
Lower limit risk: 2.3%	Result: ACCEPT	-
Process capability index (Cpk): 0.67	Specific FA Risk: 46%	-



Process Risk	Specific Measurement Risk	Global Risk
Process Risk: 4.6%	TUR: 4.0	Total PFA: 0.79%
Upper limit risk: 2.3%	Measured value: 0.99	Total PFR: 1.5%
Lower limit risk: 2.3%	Result: ACCEPT	-
Process capability index (Cpk): 0.67	Specific FA Risk: 46%	-



Global Risk

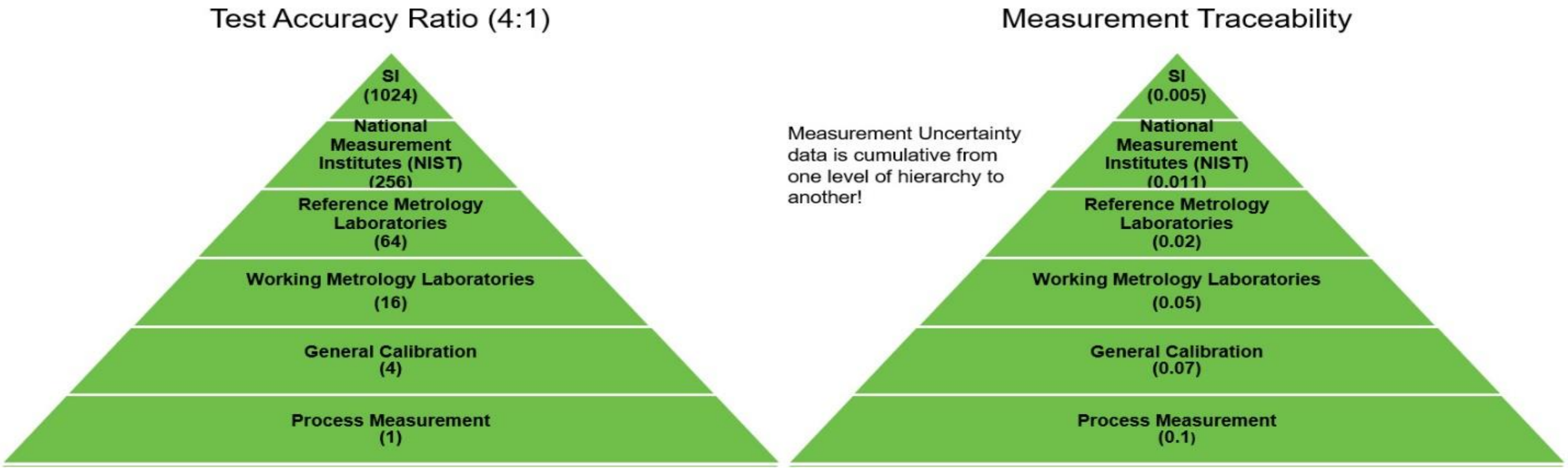
Global Risk (also called **process-level risk**) is **based on a future measurement result**.

- It is used to ensure the acceptability of a documented measurement process.
- It is based on expected or historical information and is usually characterized by two probability distributions.

The term TUR (Test Uncertainty Ratio) is commonly used as a simplified approach to evaluating global risk. When we know the tolerance, we are working to, we have a high enough sample size to know the shape and the distribution of the calibration results.

We can then use TUR with End of Period reliability to calculate the appropriate uncertainty that corresponds to the maximum amount of false accept risk we are okay with.

Outdated Practices Can Lead to Higher Risk

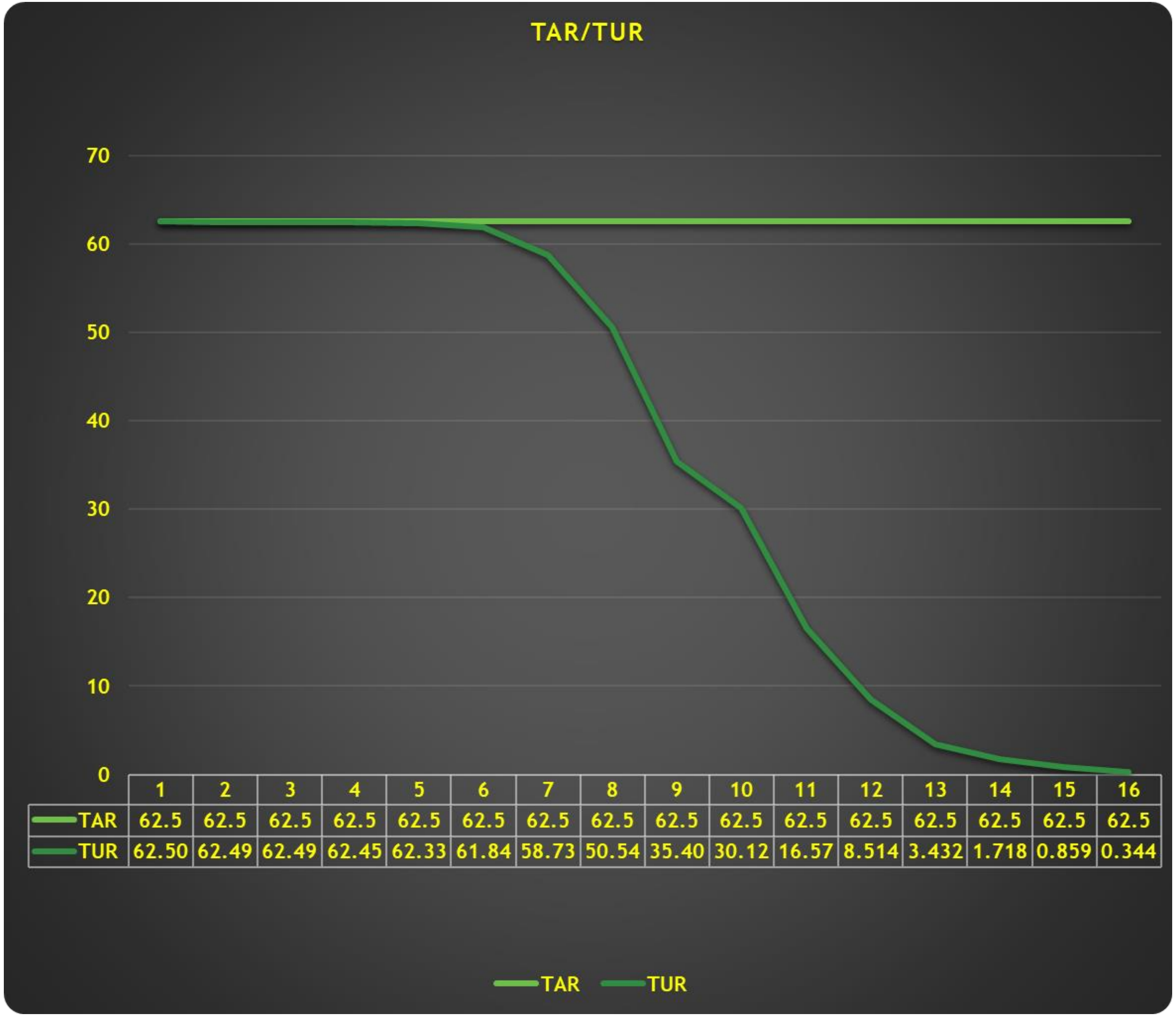


In *Measurement Decision Risk – The Importance of Definitions*, Scott M. Mimbs provides an example of a digital micrometer using a TAR 25:1 ratio. Comparing this example with the definition of TUR found in the ANSI/NCSL Z540.3 Handbook produces a 1.5:1 ratio for the same measurement.

Outdated Practices Lead to Higher Risk

Capacity	Req Tolerance	LSL	USL	Res UUT	Rep UUT	CMC	Std Unc	Exp Unc	TUR	TAR
1000	0.100%	999.0	1001.0	0.0001	0.000	0.0016%	0.01	0.02	62.500	62.5
1000	0.100%	999.0	1001.0	0.0002	0.000	0.0016%	0.01	0.02	62.498	62.5
1000	0.100%	999.0	1001.0	0.0004	0.000	0.0016%	0.01	0.02	62.493	62.5
1000	0.100%	999.0	1001.0	0.001	0.000	0.0016%	0.01	0.02	62.459	62.5
1000	0.100%	999.0	1001.0	0.002	0.000	0.0016%	0.01	0.02	62.335	62.5
1000	0.100%	999.0	1001.0	0.004	0.000	0.0016%	0.01	0.02	61.849	62.5
1000	0.100%	999.0	1001.0	0.01	0.000	0.0016%	0.01	0.02	58.737	62.5
1000	0.100%	999.0	1001.0	0.02	0.001	0.0016%	0.01	0.02	50.546	62.5
1000	0.100%	999.0	1001.0	0.04	0.001	0.0016%	0.01	0.03	35.409	62.5
1000	0.100%	999.0	1001.0	0.05	0.002	0.0016%	0.02	0.03	30.120	62.5
1000	0.100%	999.0	1001.0	0.1	0.004	0.0016%	0.03	0.06	16.573	62.5
1000	0.100%	999.0	1001.0	0.2	0.007	0.0016%	0.06	0.12	8.514	62.5
1000	0.100%	999.0	1001.0	0.5	0.018	0.0016%	0.15	0.29	3.432	62.5
1000	0.100%	999.0	1001.0	1	0.036	0.0016%	0.29	0.58	1.718	62.5
1000	0.100%	999.0	1001.0	2	0.072	0.0016%	0.58	1.16	0.859	62.5
1000	0.100%	999.0	1001.0	5	0.179	0.0016%	1.45	2.91	0.344	62.5

In this table we are only varying resolution and repeatability of the UUT.


TAR/TUR
70
60
50
40
30
20
10
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
TAR 62.5 62.5 62.5 62.5 62.5 62.5 62.5 62.5 62.5 62.5 62.5 62.5 62.5 62.5 62.5 62.5
TUR 62.50 62.49 62.49 62.45 62.33 61.84 58.73 50.54 35.40 30.12 16.57 8.514 3.432 1.718 0.859 0.344
TAR
TUR

Test Uncertainty Ratio (TUR)

Test Uncertainty Ratio: The ratio of the span of the tolerance of a measurement quantity subject to calibration to twice the 95% expanded uncertainty of the measurement process used for calibration.

NOTE: This applies to two-sided tolerances.

ANSI/NCSL Z540.3 – 2006 Definition

UUT – Unit Under Test

The Correct Definition and Calculation of TUR

$$\text{TUR} = \frac{\text{Span of the } \pm \text{Tolerance}}{2 \text{ x } k_{95\%} \left(\sqrt[2]{\left(\frac{\text{CMC}}{k_{\text{CMC}}}\right)^2 + \left(\frac{\text{Resolution}_{\text{UUT}}}{\sqrt[2]{12}}\right)^2 + \left(\frac{\text{Repeatability}_{\text{UUT}}}{1}\right)^2 + \cdots (u_{Other})^2} \right)}$$

Example of a TUR Formula (Adapted from the ANSI/NCSL Z540.3 Handbook)

In most cases, the numerator is the UUT Accuracy Tolerance. The denominator is slightly more complicated. Per the ANSI/NCSL Z540.3 Handbook, "For the denominator, the 95 % expanded uncertainty of the measurement process used for calibration following the calibration procedure is to be used to calculate TUR. The value of this uncertainty estimate should reflect the results that are reasonably expected from the use of the approved procedure to calibrate the M&TE. Therefore, the estimate includes all components of error that influence the calibration measurement results, which would also include the influences of the item being calibrated except for the bias of the M&TE. The calibration process error, therefore, includes temporary and non-correctable influences incurred during the calibration such as **repeatability**, **resolution**, error in the measurement source, operator error, error in correction factors, environmental influences, etc."

TUR (Test Uncertainty Ratio)

$$\mathbf{TUR} = \frac{\text{Span of the } \pm \text{ Tolerance}}{2 \text{ x } k_{95\%} \left(\sqrt[2]{\left(\frac{\text{CMC}}{k_{\text{CMC}}}\right)^2 + \left(\frac{\text{Resolution}_{\text{UUT}}}{\sqrt[2]{12}}\right)^2 + \left(\frac{\text{Repeatability}_{\text{UUT}}}{1}\right)^2 + \cdots (u_{Other})^2} \right)}$$

UUT Tolerance = (USL-LSL)/2

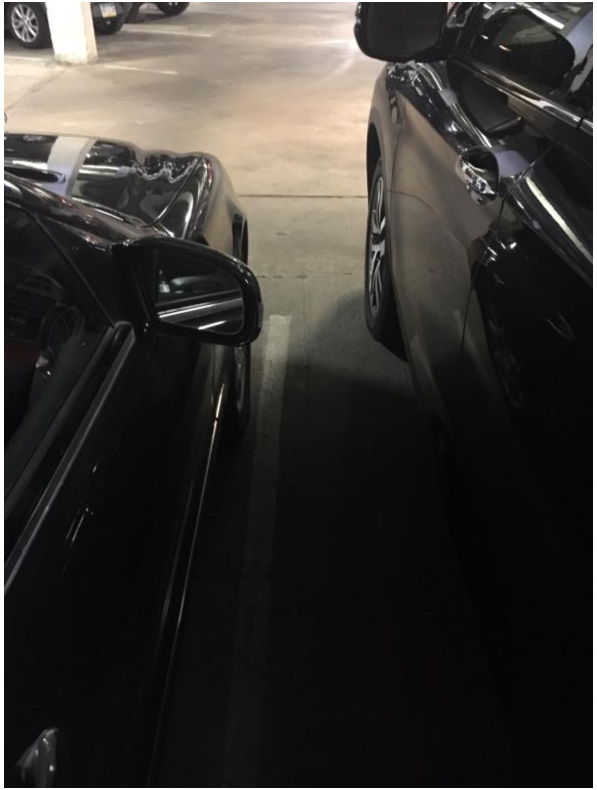
CMC = Reference labs Calibration and Measurement Capability

k = coverage factor

ANSI/NCSL Z540.3 Handbook Definition

The lab with the smaller uncertainties will typically produce larger TURs, giving you more space to be in tolerance!

The lab with the larger uncertainties will typically produce smaller TURs, giving you less space to be in tolerance!



EOPR

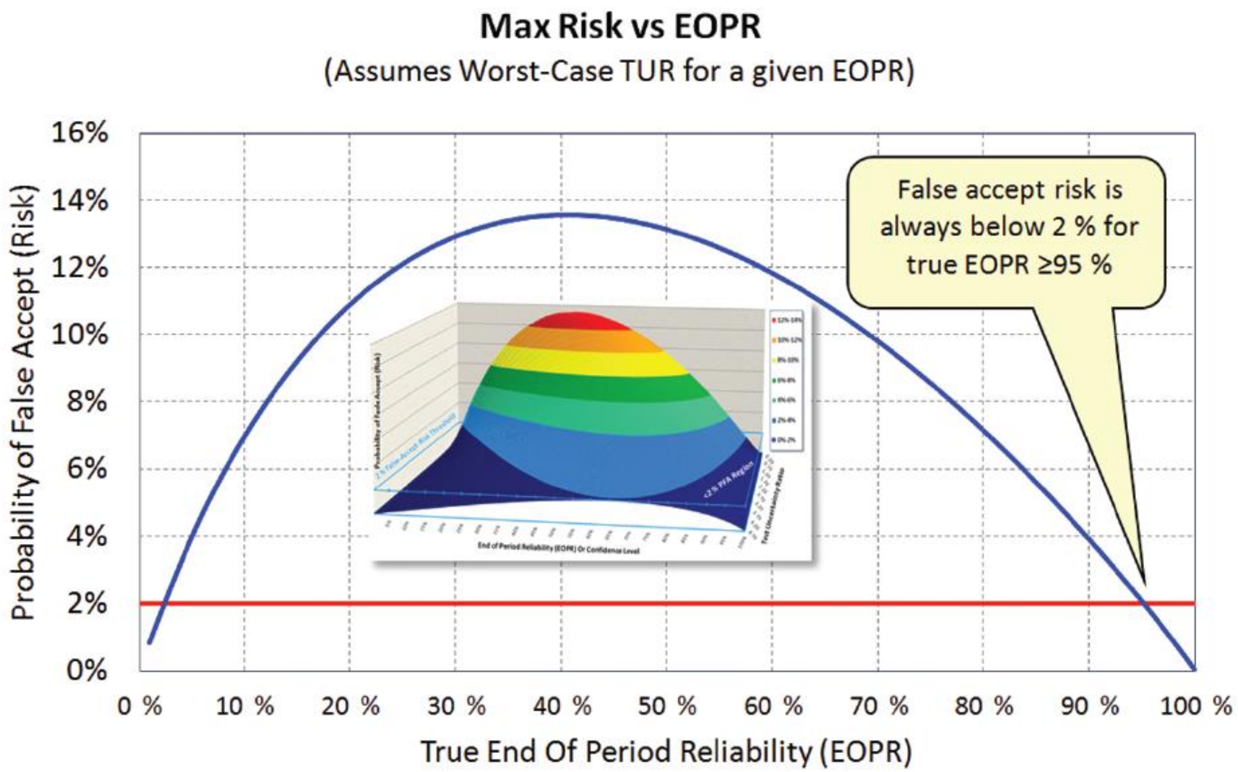
$$EOPR = \frac{\text{Number of in-tolerance results}}{\text{Total number of calibrations}}$$

In simplistic terms, End of Period Reliability is defined as the number of calibrations resulting in acceptance criteria being met divided by the total number of calibrations. This formula to determine "In-Tolerance" Reliability from historical data is easy to replicate in Excel. The formula is **Sample Size = ln(1-Confidence)/ln(Target Reliability)**

If we use the formula for Sample Size above, we will need over 59 (58.4) samples to use a joint probability distribution associated with many TUR-based methods.

There is more with EOPR as the rules to establish EOPR can be subjective. Things such as how many first-time calibrations are counted, broken instruments included, are calibrations with different due dates, or calibrations that are extended included, what about post-dating, and so on.

EOPR



From Risk Mitigation Strategies for Compliance Testing by Jonathan Harben and Paul Reese

Global Risk – EOPR Basic Overview

$$EOPR = \frac{\text{Number of in-tolerance results}}{\text{Total number of calibrations}}$$

In simplistic terms, End of Period Reliability is defined as the number of calibrations that meet acceptance criteria divided by the total number of calibrations.

Reliability Considerations may include:

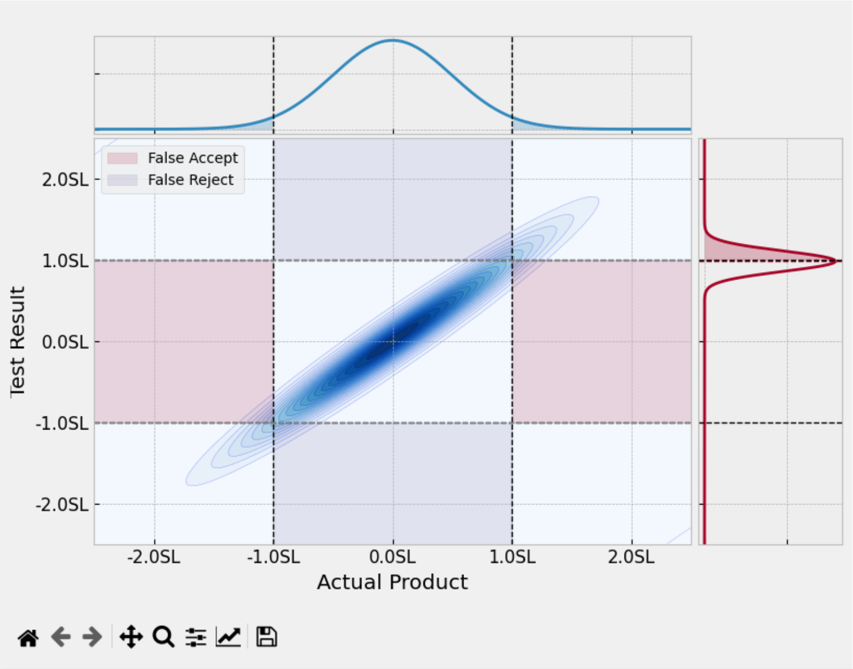
- Reliability decreases with time after calibration
- How much testing is required to demonstrate Reliability with confidence?
- *A priori* knowledge of the M&TE

Global Risk – EOPR Basic Overview

This formula to determine "In-Tolerance" Reliability from historical data is easy to replicate in Excel. The formula is **Sample Size = ln(1-Confidence)/ln(Target Reliability).**

When we use this formula for 95 % EOPR at a 95 % Confidence Interval, we need 59 samples with 0 failures or rejects as this will give us an estimation of our process.

Example: If Caitlin Clark makes 59 consecutive 3-pt shots, we would have reliability data to start using global risk models.



Process Risk	Specific Measurement Risk	Global Risk
Process Risk: 4.6%	TUR: 4.0	Total PFA: 0.79%
Upper limit risk: 2.3%	Measured value: 0.99	Total PFR: 1.5%
Lower limit risk: 2.3%	Result: ACCEPT	-
Process capability index (Cpk): 0.67	Specific FA Risk: 46%	-

Breaking down reliability with Basketball (95 %)

Binomial Operating Characteristics (OC) Curve (Enter cal failures in yellow)

Reliability Constraints	
Reliability Target =	95.0 %
Confidence Target =	95.0 %
Calculated Sample Size =	59

(Used to establish initial sample size)

☐ Correct to "True" EOPR?

Calibration History Results	
Calibrations or Sample Size, n =	100
Failures/Rejects, c =	3
Confidence Level =	95%
Failure Rate =	3.00%
Unreliability (worst case) =	7.57%
R(t) =	92.43%

95% Confidence the process is at least 92.43% reliable.

Upper Confidence Limit = 99.38%

Lower Confidence Limit = 92.43%

UUT Constraints	
Assumed EOPR =	68.27%
Use Add. Ref. Std. Samples?	☐
Max PFA =	2.00%

53 Additional samples needed

Reference EOPR with Additional Samples	
Samples needed to meet R(t) =	153
Failures/Rejects, c =	3
Confidence Level =	95%
Failure Rate =	1.96%
Unreliability (worst case) =	4.99%
Reliability (Sample size adj.) =	95.01%

95% Confidence the process is at least 95.02% reliable.

Upper Confidence Limit = 99.59%

Lower Confidence Limit = 95.01%

Breaking down reliability with Basketball (41.5 %)

Binomial Operating Characteristics (OC) Curve (Enter cal failures in yellow)

Reliability Constraints	
Reliability Target =	41.5 %
Confidence Target =	95.0 %
Calculated Sample Size =	4

(Used to establish initial sample size)

☐ Correct to "True" EOPR?

Calibration History Results	
Calibrations or Sample Size, n =	100
Failures/Rejects, c =	3
Confidence Level =	95%
Failure Rate =	3.00%
Unreliability (worst case) =	7.57%
R(t) =	92.43%

95% Confidence the process is at least 92.43% reliable.

Upper Confidence Limit = 99.38%

Lower Confidence Limit = 92.43%

UUT Constraints	
Assumed EOPR =	68.27%
Use Add. Ref. Std. Samples?	☐
Max PFA =	2.00%

-89 Additional samples needed

Reference EOPR with Additional Samples	
Samples needed to meet R(t) =	11
Failures/Rejects, c =	3
Confidence Level =	95%
Failure Rate =	27.27%
Unreliability (worst case) =	56.44%
Reliability (Sample size adj.) =	43.56%

95% Confidence the process is at least 43.57% reliable.

Upper Confidence Limit = 93.98%

Lower Confidence Limit = 43.56%

Ref Standard Stability / EOPR

$$EOPR = \frac{\text{Number of in-tolerance results}}{\text{Total number of calibrations}}$$

Sample Size = ln(1-Confidence)/ln(Target Reliability).

This is a formula for calculating with enough confidence how many samples one would need to prove a tolerance or say stability from one calibration to the next.

For our example, we needed 59 samples with 0 failures, when that did not happen, we needed more samples.

Reliability Constraints	
Reliability Target =	95.0 %
Confidence Target =	95.0 %
Calculated Sample Size =	59

(Used to establish initial sample size)

☐ Correct to "True" EOPR?

Calibration History Results	
Calibrations or Sample Size, n =	60
Failures/Rejects, c =	3
Confidence Level =	95%
Failure Rate =	5.00%
Unreliability (worst case) =	12.42%
R(t) =	87.58%

Actual Samples 60

Upper Confidence Limit = 98.96%

Lower Confidence Limit = 87.58%

UUT Constraints	
Assumed EOPR =	95.00%
Use Add. Ref. Std. Samples?	☐
Max PFA =	2.00%

93 Additional samples needed

Reference EOPR with Additional Samples	
Samples needed to meet R(t) =	153
Failures/Rejects, c =	3
Confidence Level =	95%
Failure Rate =	1.96%
Unreliability (worst case) =	4.99%
Reliability (Sample size adj.) =	95.01%

Ref Standard Stability / EOPR

What we found was 171 load cells and sampled the stability criteria at 10 %, 50 %, and 100 % test points with 95.77 % confidence that are population of sampled load cells was better than 0.05 % from year-to-year

Reliability Constraints	
Reliability Target =	95.0 %
Confidence Target =	95.0 %
Calculated Sample Size =	59

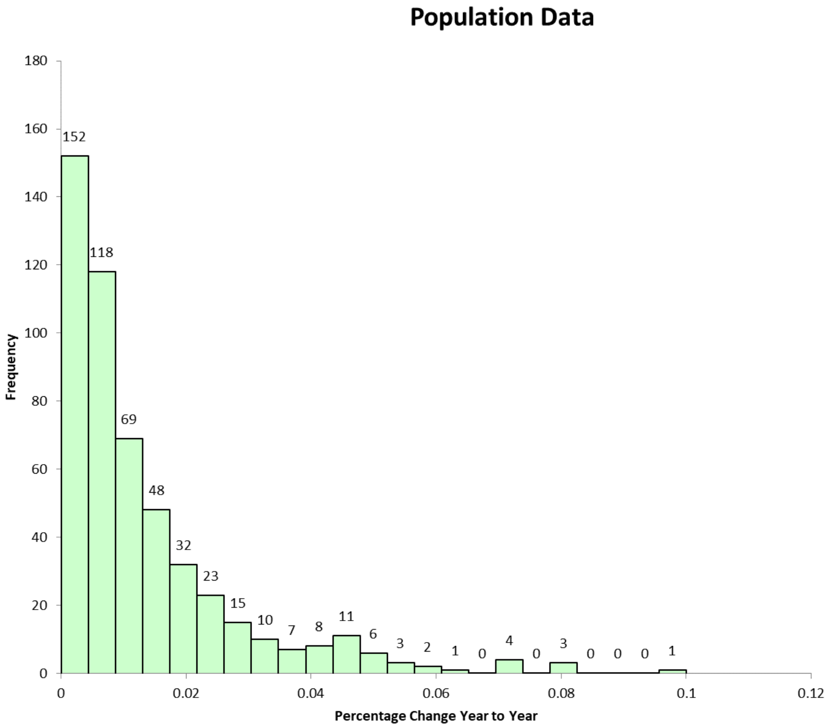
(Used to establish initial sample size)

☐ Correct to "True" EOPR?

Actual Samples 513

Calibration History Results	
Calibrations or Sample Size, n =	513
Failures/Rejects, c =	14
Confidence Level =	95%
Failure Rate =	2.73%
Unreliability (worst case) =	4.23%
R(t) =	95.77%

Upper Confidence Limit = 98.50%
Lower Confidence Limit = 95.77%



Descriptive Statistics
Mean=0.01
Standard Error=0.000442
Median=0.008
Standard Deviation=0.01
Variance=0
Sum=6.73199999999998
Count=513
Maximum=0.1
Minimum=0
Range=0.1
Kurtosis=5.8858
Skewness=2.184

Basics – The Quality of the Measurement System

A method for characterizing the quality of a measurement system relative to a specific tolerance to be measured is defined in three documents.

JCGM 106:2012 defines the *measurement capability index* (C_m)

ASME B89.7.4.1-2005 uses the same definition for C_m

ANSI/NCSL Z540.3-2006 redefines *the test uncertainty ratio* (TUR) with respect to the GUM

$$C_m = \frac{T_U - T_L}{4 \cdot u_m} = \frac{T_U - T_L}{2 \cdot U_{95}} \qquad TUR = \frac{L_{upper} - L_{lower}}{2 \cdot U_{95}} \qquad \begin{array}{l} U_{95} = k \cdot u_m \\ k = 1.96 \approx 2 \end{array}$$

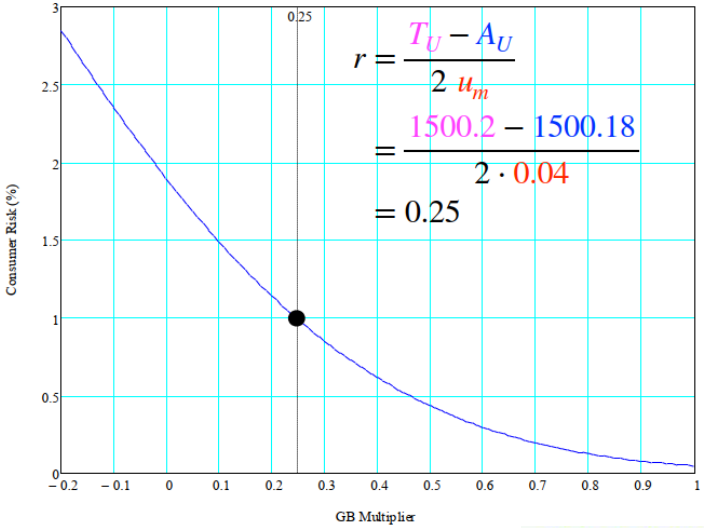
The C_m and *TUR* are useful indicators of the quality of the measurement system (i.e., large C_m *or TUR* is indicative of low measurement uncertainty compared with the tolerance).

As defined in the references, C_m and *TUR* are mathematically identical; however, there is a possibility that *TUR* could be confused with an older, less rigorous definition.

Therefore, we favor the *measurement capability index* (C_m).

Setting acceptance limits for Global Risk is more complex, requiring iterative numerical integration to find the acceptance limits for a desired level of risk. Many times, graphical solutions can be used to set the acceptance limits.

Guard Band Multiplier, **r = Norm.s.inv(0.6827)/2 = 0.2376**



$$r = \frac{T_U - A_U}{2\,u_m} = \frac{1500.2 - 1500.18}{2 \cdot 0.04} = 0.25$$

$$A_U = T_U - r \cdot 2\,u_m = 1500.18$$

$$C_m = \frac{0.4}{4 \cdot 0.04} = 2.5$$

$$C_p = \frac{0.4}{6 \cdot 0.12} = 0.556$$

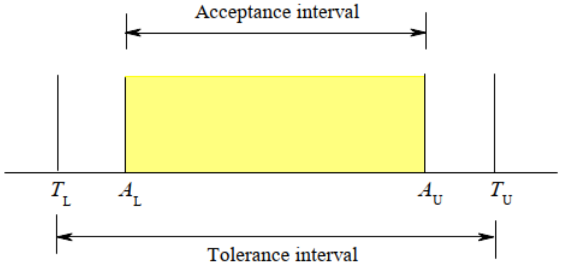
Conformance Probability, P_C	Guard Band Multiplier, r
0.80	0.42
0.85	0.52
0.90	0.64
0.95	0.82
0.977	1.00
0.99	1.16
0.999	1.55

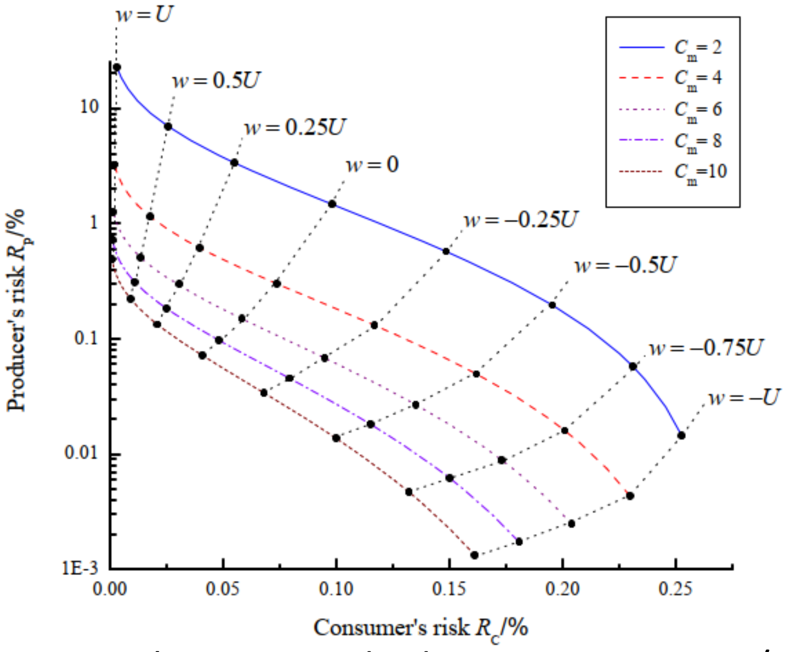
ASME B89.7.4.1-2005 numerical example. T=0.4 mm, TU=1500.2 mm, TL=1499.8 mm, um=0.04 mm, xm=1500.16 mm

$$P_c(0.4, 1500.2, 1499.8, 0.04, 1500.16) = 84.134\%$$

$$1 - P_c(0.4, 1500.2, 1499.8, 0.04, 1500.16) = 15.866\%$$

Global Risk





Global risks RP versus RC for a binary conformity assessment with prior standard uncertainty u0 = T/6. The five curves correspond to values of the measurement capability index Cm = T/(4um) in an interval from 2 to 10. The solid points locate Guard Bands with length parameters from w = −U to w = U, with U = 2u. Positive values of w correspond to guarded acceptance, with acceptance limits inside the tolerance limits as shown left.

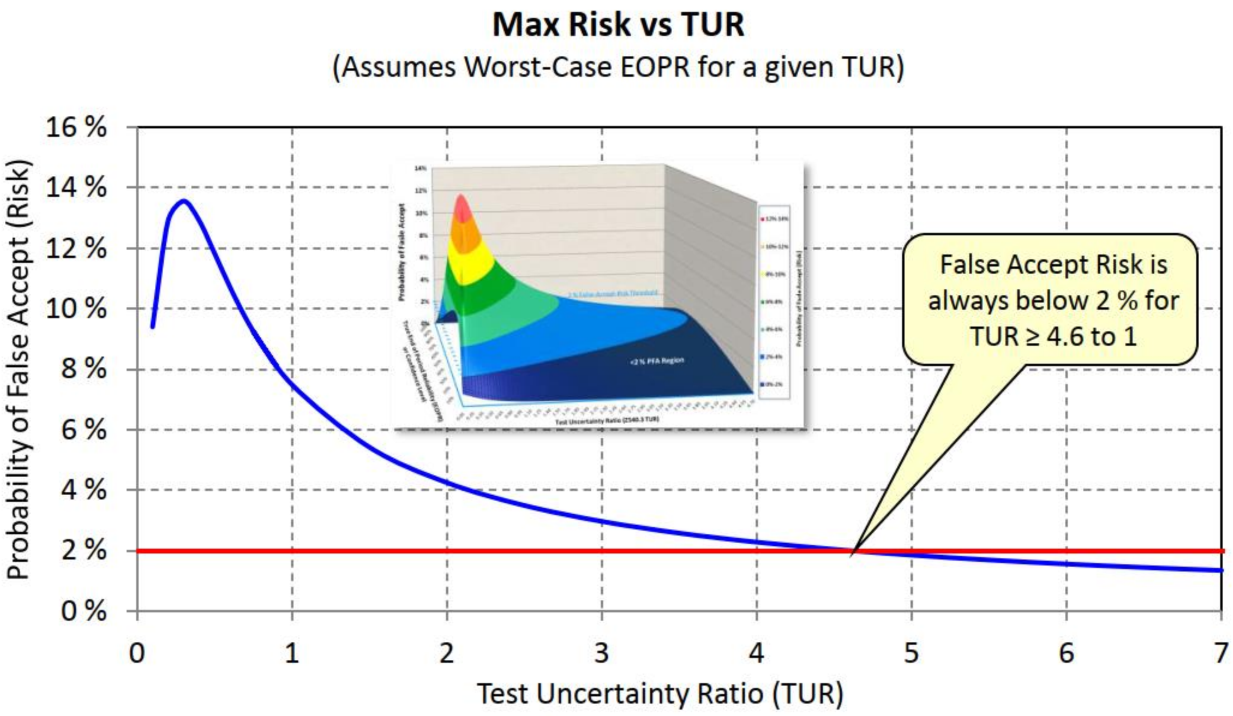
Global Risk

Scott Mimbs wrote a paper on EOPR at 89 % and how the 2 % PFA rule could be met by analyzing a population of instruments with years of history.

If one cannot gather all of the information, then further analysis would be needed, and TUR must be determined at each test point. If the analysis reveals the TUR is greater than 4.6:1, then the PFA will be less than 2 %. (From Risk Mitigation Strategies for Compliance Testing)

If neither the EOPR nor TUR threshold is met, one could choose to use Specific Risk methods or use another method.

Global Risk



The image is taken from Implementing Strategies for Risk Mitigation In the Modern Calibration Laboratory

Global Risk

We want to build an army of clones to defeat the Rebellion.

The optimum height is 70 inches ± 2 inches to fit our clones with the same gear and maximize cloning efficiencies.

Our measurement system has a TUR of 8:1 meaning our Calibration Process Uncertainty is 0.25 inches.

The question becomes what is the probability of saying a clone conforms to the specification when it does not?

Global Risk

Parameters | Notes

Mode: Full

Calculation: Integral

Lower Specification Limit: 68

Upper Specification Limit: 72

☑ Process Distribution:

	Parameter	Value
1	Distribution	normal
2	median	70
3	std	0.53148265198...

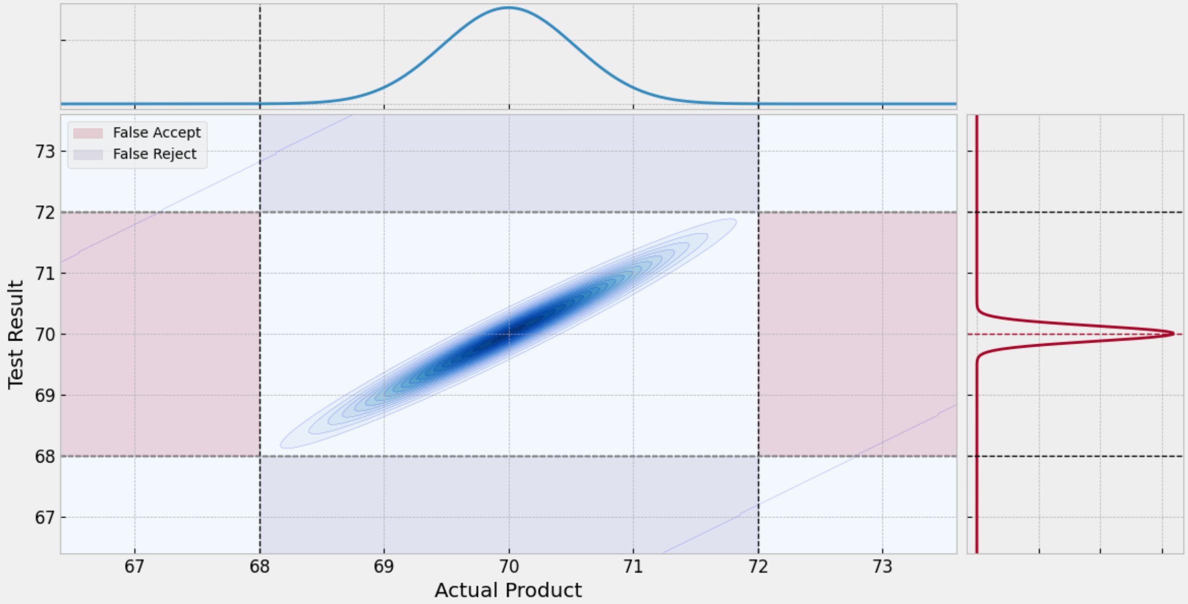
☑ Test Measurement:

	Parameter	Value
1	Distribution	normal
2	measurement	70
3	std	0.125
4	bias	0

☐ Guardband

Lower Guardband (relative): 0

Upper Guardband (relative): 0



Process Risk	Specific Measurement Risk	Global Risk
Process Risk: 0.017%	TUR: 8.0	Total PFA: 0.0039%
Upper limit risk: 0.0084%	Measured value: 70	Total PFR: 0.012%
Lower limit risk: 0.0084%	Result: ACCEPT	-
Process capability index (Cpk): 1.3	Specific FA Risk: 6.4e-56%	-

Global Risk

Global vs Specific Risk Example

A company has hired us to measure the speed of cars on a stretch of a single-lane road.

The customer has indicated they are okay with 57 -63 miles per hour (MPH) speeds.

Thus, our specification limit is based on 60 MPH ± 3 MPH. The posted speed limit is 60 MPH.

After much discussion, we decided to set up two radar guns at points A and B for the first day and report the results. (Example of Specific Risk is based on measuring individual speeds at point A or point B)

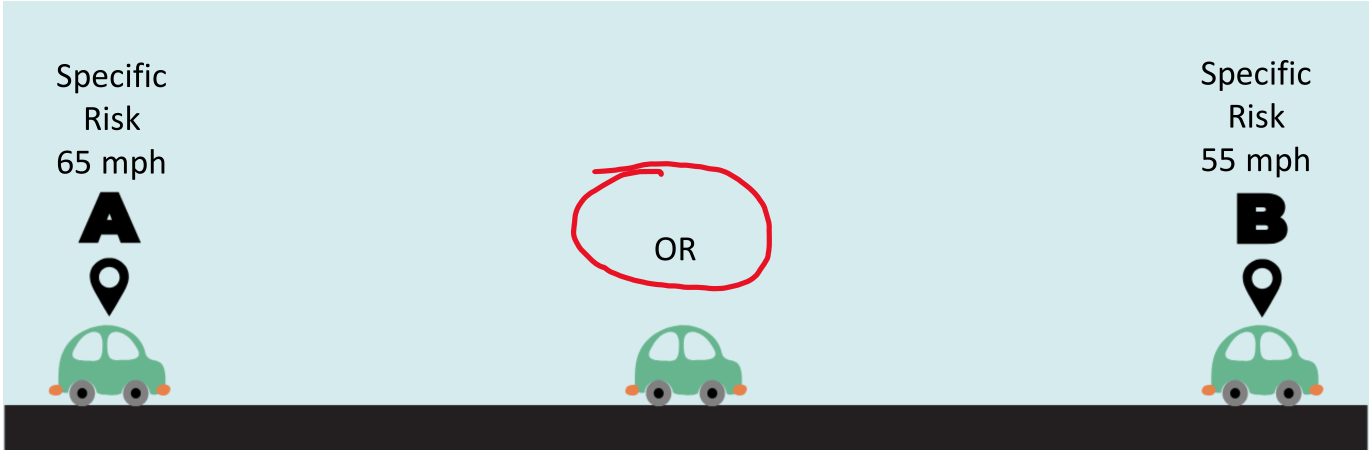
Global Risk vs Specific Risk Example

A company has hired us to measure the speed of cars on a stretch of a single-lane road.

Our **specification limit** is based on **60 MPH ± 3 MPH**. The posted speed limit is **60 MPH**.

After much discussion, we decided to set up two radar guns at points A and B for the first day and report the results.

Specific Risk



If we wanted to look at the car's speed using Specific Risk, we might have a radar gun at either points A or B. In this example, the car is clocked at 65 mph at point A and 55 MPH at point B. Each point is 5 MPH below or above the speed limit.

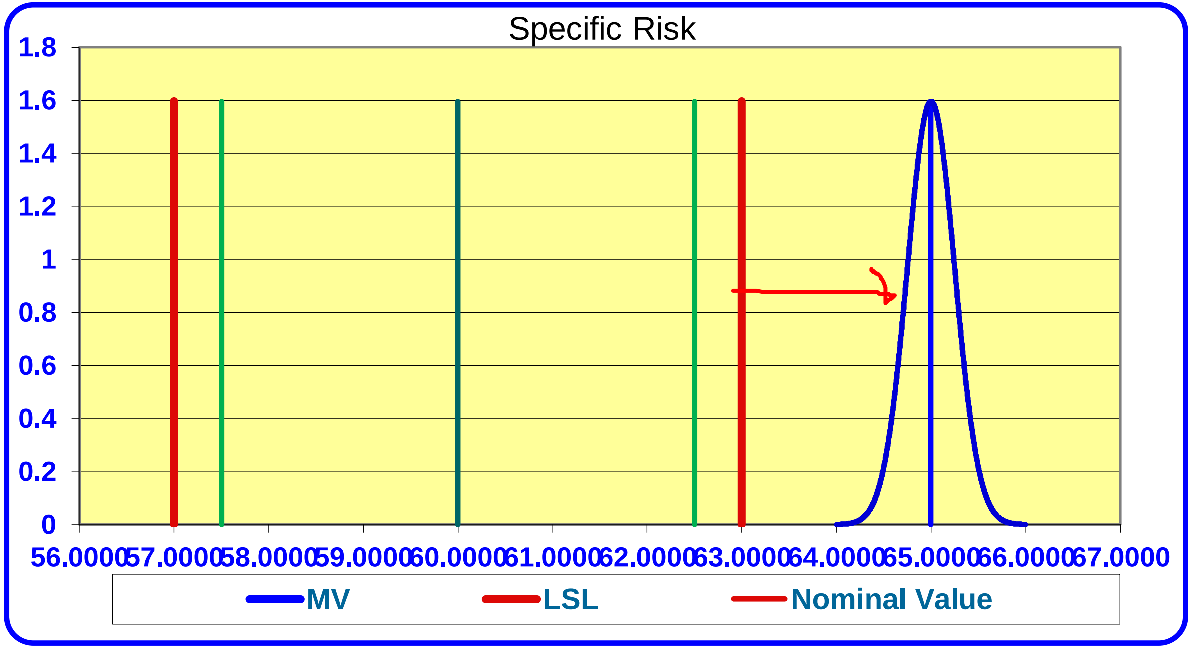
Specific Risk (Radar Gun at Point A)

Risk Calculator	
Upper Tolerance T_U	63.0000
Lower Tolerance T_L	57.0000
Nominal Value	60.0000
Measurement Unc um	0.2500
Measured Value xm	65.0000
Tolerance T	6.00

Probability of Conformance (p_c)	0.000%
Probability of NonConformance ($1 - p_c$)	100.000%

Setting the Guard Band Upper and Lower p_c	
Select Desired Conformance Probability	0.977
Guard Band Upper G_u	62.5000
Guard Band Lower G_L	57.5000

Specific Risk (Bench-Level)	
Conditional Probability False Accept	100.000%
Conditional Probability False Reject	0.000%



5 MPH Above (TUR 6:1)

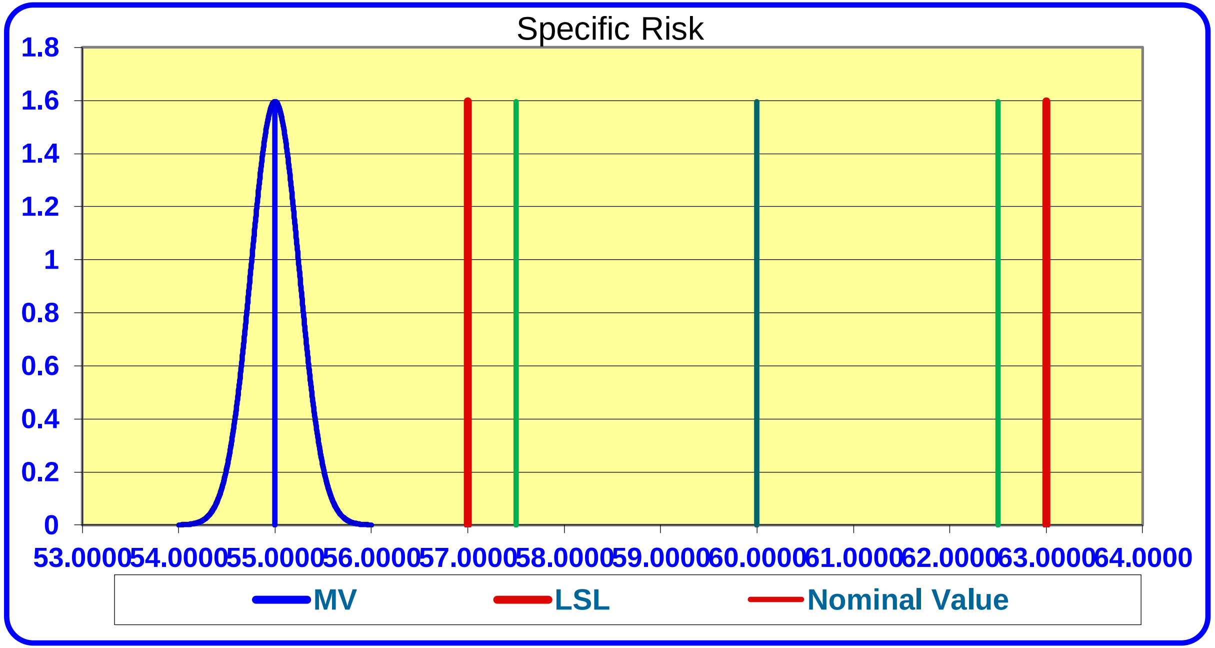
Specific Risk (Radar Gun at Point B)

Risk Calculator	
Upper Tolerance T_U	63.0000
Lower Tolerance T_L	57.0000
Nominal Value	60.0000
Measurement Unc um	0.2500
Measured Value xm	55.0000
Tolerance T	6.00

Probability of Conformance (p_c)	0.000%
Probability of NonConformance (1 - p_c)	100.000%

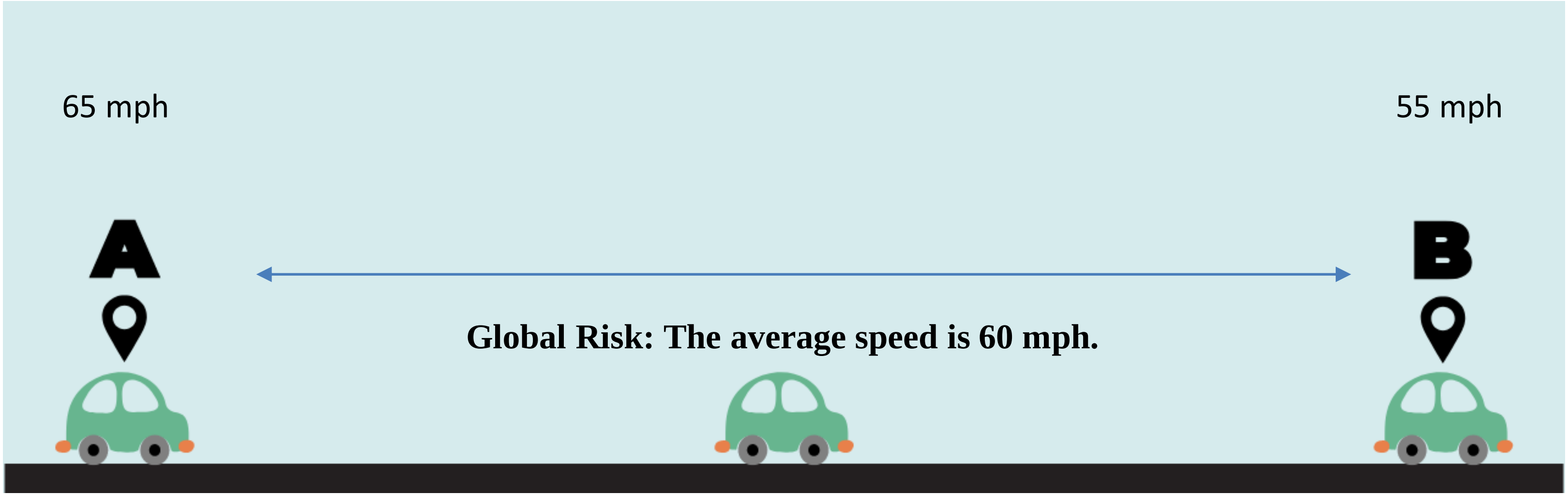
Setting the Guard Band Upper and Lower p_c	
Select Desired Conformance Probability	0.977
Guard Band Upper G_u	62.5000
Guard Band Lower G_L	57.5000

Specific Risk (Bench-Level)	
Conditional Probability False Accept	100.000%
Conditional Probability False Reject	0.000%



5 MPH Below (TUR 6:1)

Global Risk

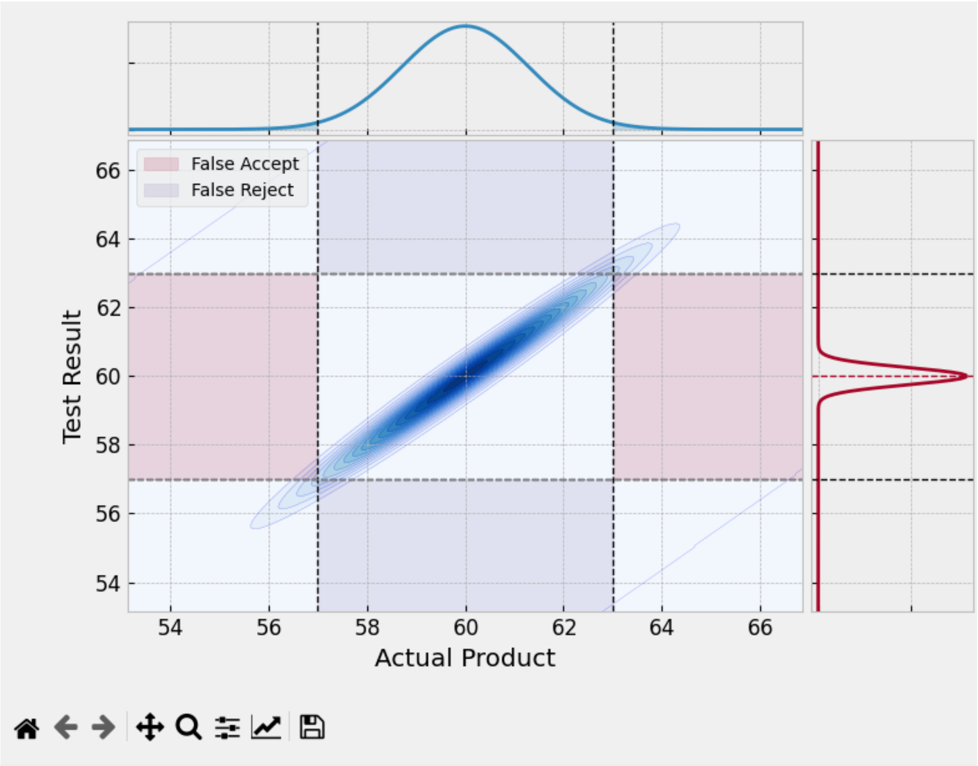


The car enters point A, traveling at 65 MPH, and then 0.5 miles into the drive, travels at 55 MPH. Global risk is based on measuring the average speed once a reliability target has been met (we took 10,000 data points and found 98 % to be good).

Global Risk Results

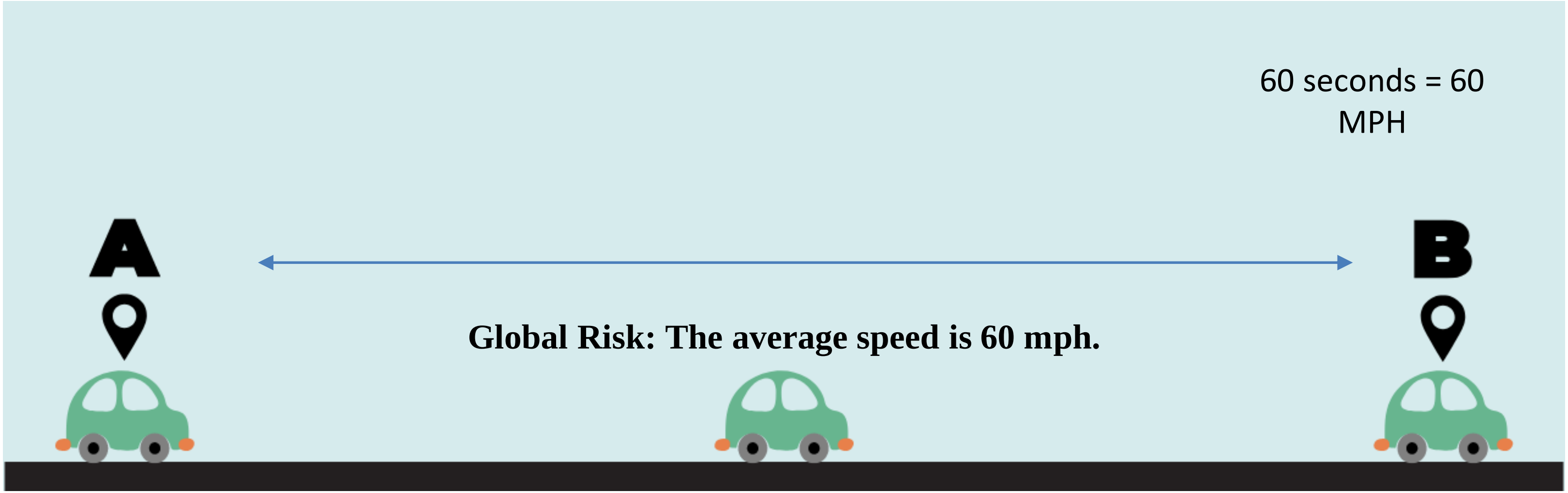
On Day 1, we record about 10,000 vehicles. Out of the 10,000 vehicles, 9,800, or 98 %, are observed to be driving between 57 - 63 MPH.

In our example, the TUR is 6:1 using 2 Radar Guns and we took the average speed.



Process Risk	Specific Measurement Risk	Global Risk
Process Risk: 2.0%	TUR: 6.0	Total PFA: 0.32%
Upper limit risk: 1.0%	Measured value: 60	Total PFR: 0.54%
Lower limit risk: 1.0%	Result: ACCEPT	-
Process capability index (Cpk): 0.78	Specific FA Risk: 1.8e-31%	-

Global Risk

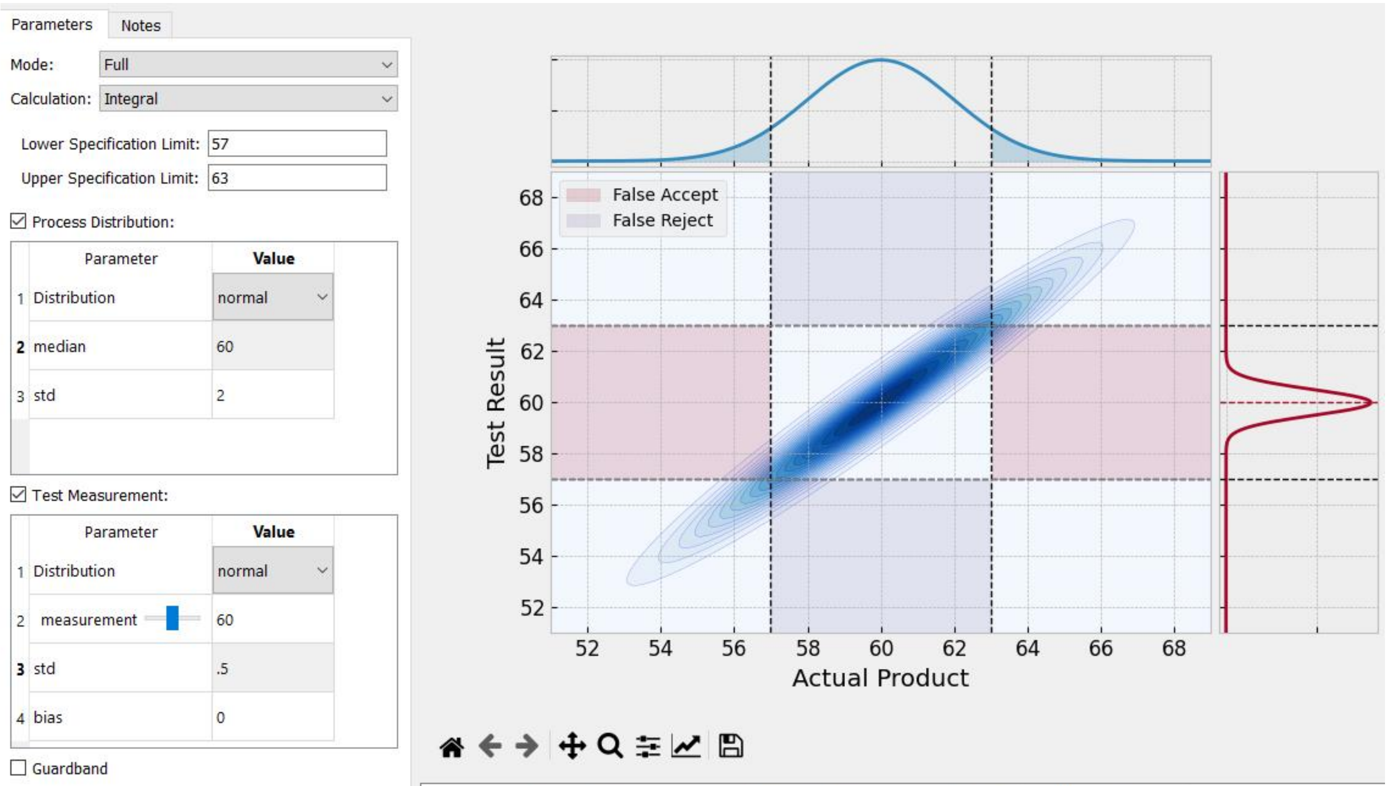


Since two radar guns are very good (high TUR), though expensive, maybe we consider a less expensive method, maybe an automated time-based method.

With a less accurate method our TUR might be 3:1 or half as good. What would this look like?

Global

Global Risk – Using a less accurate means of measuring (A process with a higher measurement uncertainty)



Process Risk	Specific Measurement Risk	Global Risk
Process Risk: 13%	TUR: 3.0	Total PFA: 2.0%
Upper limit risk: 6.7%	Measured value: 60	Total PFR: 3.3%
Lower limit risk: 6.7%	Result: ACCEPT	-
Process capability index (Cpk): 0.50	Specific FA Risk: 2.0e-07%	-

Global Versus Specific Risk Summary

Specific Risk is dependent on a single probability function and can be referred to as Probability of Conformance from the customer's point of view.

Global Risk is dependent on two probabilities, the second being the *a priori* knowledge, which could be taken as the process or instrument reliability.

Typically, when we talk about TUR, we are talking about Global Risk.

Though TUR is also a ratio that can be useful at the Specific Risk level as higher TURs increase our acceptance zone.

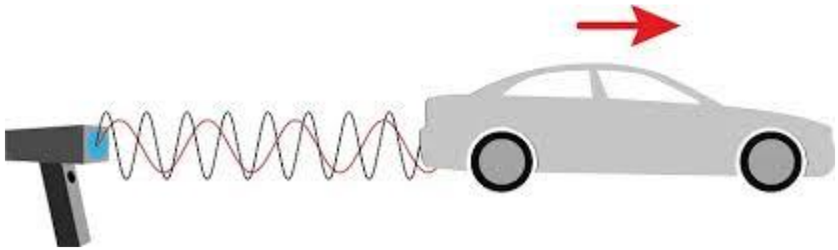
Example – Radar Gun

EXAMPLE 1 Speed limit enforcement

Highway law enforcement gets wind of 25 % of the cars speeding ☺

The speed of motorists is measured by police using devices such as radars and laser guns. A decision to issue a speeding ticket, which may potentially lead to an appearance in court, must be made with a high degree of Confidence that the speed limit has been exceeded.

If we know that we can only win a court case if there is a 99.9 % probability that our speed limit has been exceeded, when can we write a ticket?



Example – Radar Gun

Guarded Rejection	
Probability (%)	99.90%
Nominal Value	75.0000
Measurement Unc um	0.0200
99.9 % Acceptence Threshold	79.941
Probability of Making a Wrong Decision	0.10%

The Speed Limit is 60 mph with an
um of 2 %
um = 0.02
Probability = 99.9 %
TU = 60

What is Vmax or the speed someone has to be going to receive a speeding ticket with 99.9 % probability of actually speeding?

Vmax = TU / 1 – 0.02 z

ANSWER?

Vmax = 60 / 1 -0.02 (3.090) = 63.953 mph

What about 75 mph?

Vmax = 75 / 1 -0.02 (3.090) = 79.94 mph

$$P_c = \Phi(z) = 99.9\,\%$$
$$z = \Phi^{-1}(0.999) = 3.09$$

=NORMSINV(Probability)/1

Probability Table (single sided)	
Probability	Z-Value
0.5000	0.000
0.7000	0.524
0.7500	0.674
0.8500	1.036
0.9000	1.282
0.9500	1.645
0.9545	1.690
0.9773	2.000
0.9800	2.054
0.9900	2.326
0.9990	3.090
1.0000	5.998

Overall review with Examples

Global Risk

- Resistor Manufacturing – Based on **in-tolerance probability (*itp*),** within binomial confidence bound, utilizing appropriate reference standards
- As more items are manufactured, the binomial confidence bound reflects the uncertainty of the *itp*
- This method relies on the bivariate (2 variables) Gaussian joint probability distribution function, the UUT (in this case, a fixed resistor), and the measurement system is assumed to be a high-speed DMM

Specific Risk

- Resistor Measurement - anything failing the test tolerance in the manufacturing process may be segregated and subject to additional scrutiny
- If meager information is provided (such as a prototype), or the risk of passing a potentially bad item is high, a more conservative approach would utilize specific risk

Testing Cost Associated - specific risk versus global

Often, choosing the correct equipment for a specific task is an afterthought. Let's start with a simple example of verifying a manufactured fixed-value resistor.

The fixed value is 1,500 ohms ±0.2 ohms (±~0.013…3 %). The original idea is to manufacture these items as inexpensively as possible.

After some quick research, based on cost, a choice is made to purchase an Acme Low Buck DMM as it "seems" to fit the situation due to the exception resolution of 7 ½ digits, and it doesn't break the budget as the list cost is ~$4,405

The TUR being ~1.44:1 – the process owner felt this was adequate because a simple global GB tolerance of ±0.124 (in lieu of ±0.2) is in place for pass/fail criteria, and it was assumed the process would be centered around a nominal value of 1,500 ohms. This would give the PFA a global risk of ~1 %

With the process assumed centered around 1,500 ohms, this calculates to an *itp* of ~**99.93 % for specific risk (~1.2 % global risk** … "close enough" comes to mind)

But then, during the production process, something happens …

Testing Cost Associated - specific risk versus global

During preliminary production runs, it was discovered ~30 % of all products manufactured failed global GB tolerance testing of ±0.12Ω

An investigation is launched to verify that the rejected items are truly OOT. This is accomplished by sending failed samples to the metrology lab for verification.

The metrology lab retested the reportedly failed products and discovered less than 0.2 % were true failures

How can this happen?

Globally, the conformance tolerance may be Guard Banded at ±0.12Ω, yet for it to pass using specific risk, that same test tolerance must be reduced to ±0.039Ω leaving virtually no room for any deviations from the nominal value

The global risk tolerance must be reduced by ~40 %, while the specific risk tolerance must be reduced by >80 %!

Testing Cost Associated - specific risk versus global

The better solution is a more accurate DMM but at what cost?

Available DMM choices	Low Buck	Acme Standard	Acme Bronze	Acme Silver	Acme Gold
Estimated Acquisition Cost	$4,405.00	$6,000.00	$13,481.00	$14,315.00	$19,429.00
Measurement Uncertainty (1-sigma)	0.06939	0.04000	0.01038	0.00711	0.00677
UUT Specification	0.2	0.2	0.2	0.2	0.2
C_m	1.44	2.5	9.63	14.06	14.78

We already know the Acme Low Buck DMM isn't ideal, however the cost of an **Acme Bronze** >3x the expense but more accurate. Will it work?

Testing Cost Associated - specific risk versus global

Yes!, it will work. The expected false rejection rate is 1.035 % and the consumer risk is <1 % (process uncertainties excluded as this is exploratory)

Global Risk

New GB Limit =	100.00%
New GB Limit =	± 0.2
98.965%	0.970%
1.035%	99.030%
100.000%	100.000%

Given the cost of higher quality reference standards and the associated maintenance, is it cost-effective?

Testing Cost Associated - specific risk versus global

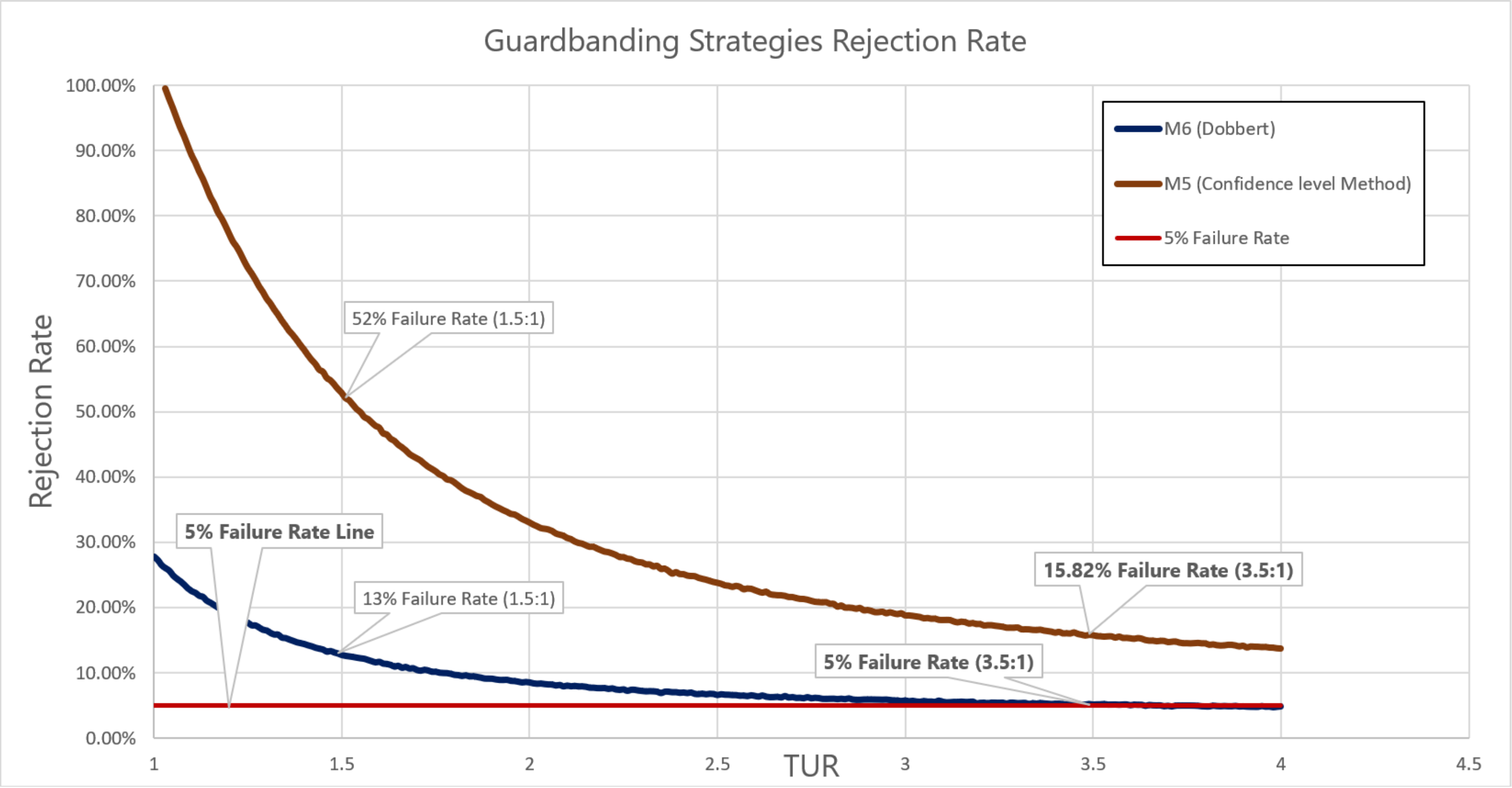
Returning to the question: “Given the cost of higher quality reference standards and the associated maintenance, is it cost-effective?”

Global Risk

New GB Limit =	100.00%
New GB Limit =	± 0.2
98.965%	0.970%
1.035%	99.030%
100.000%	100.000%

The answer is “maybe.” A maybe is a cautiously issued verdict because we don’t know the process uncertainty of our resistor batch. The previous process uncertainty (0.12) was determined using an incorrect reference meter, which was 60 % of the total tolerance. Therefore, additional testing/analysis is required to prove the concept.

Expected Failure Rates with GB Strategies

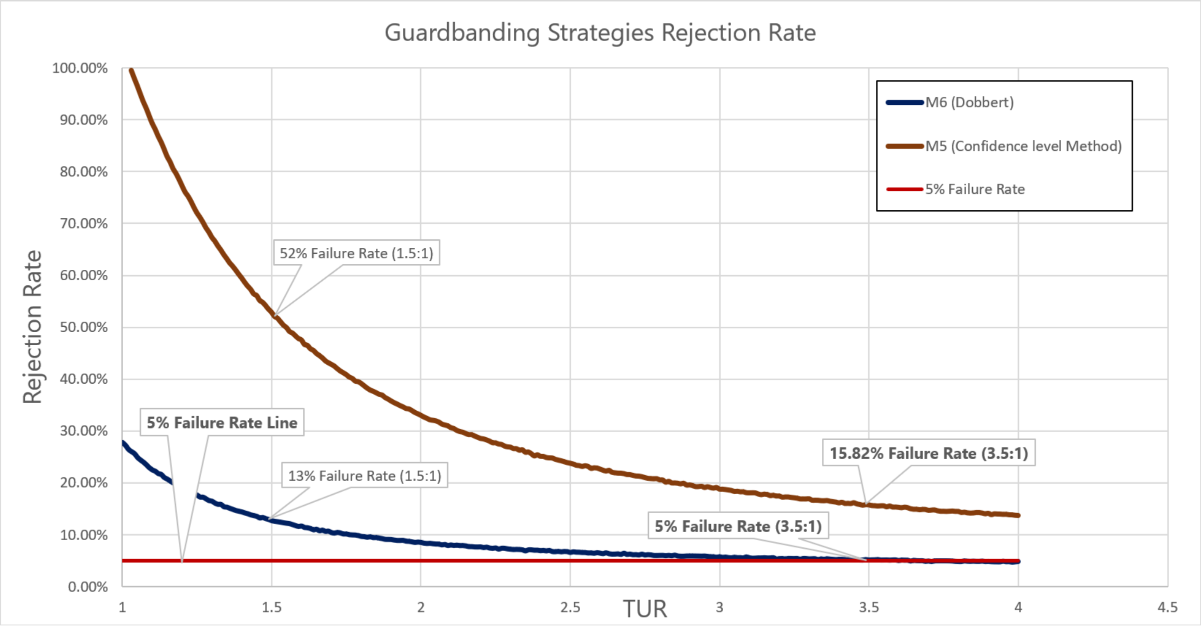


TUR's vs Total Cost due to false rejection & retest

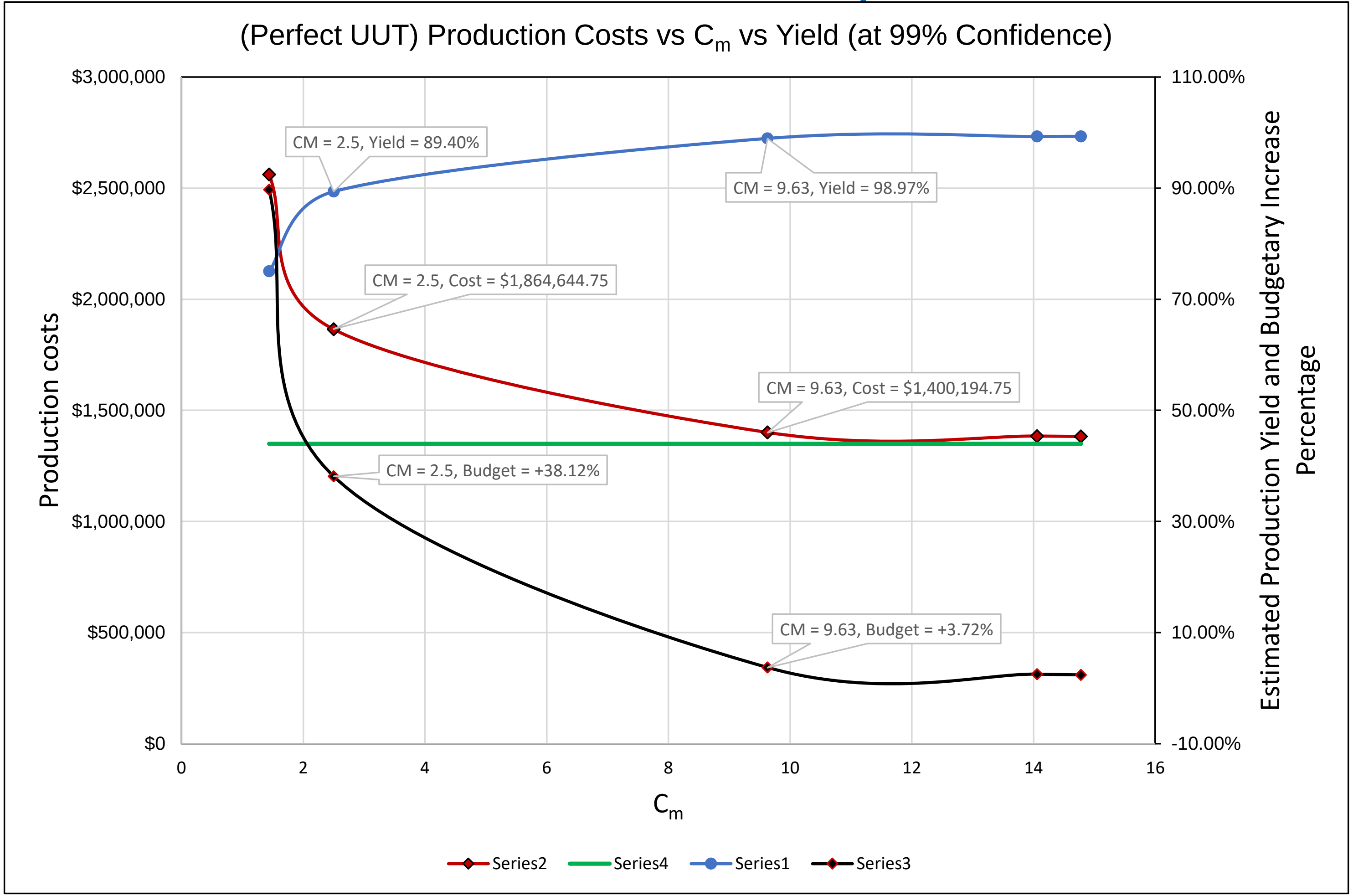
DMM Price	DMM	UUT Spec	1 sigma	TUR	PFA	PFR	Total Rejections	Falsely Rejected	Cost Due to FR	Cost to Retest	Likely Yield
$1,225	Acme Cheapest	±0.2	0.21038	0.475	-	-	-	-	-	-	-
$1,650	Acme Cheap	±0.2	0.11740	0.852	-	-	-	-	-	-	-
$4,405	Acme Low Buck	±0.2	0.06939	1.441	1.000%	24.970%	24,970	250	$337,099	$873,959	75.05%
$6,000	Acme Standard	±0.2	0.04000	2.500	1.000%	10.611%	10,611	106	$143,252	$371,393	89.40%
$13,481	Acme Bronze	±0.2	0.01038	9.630	0.970%	1.035%	1,035	10	$13,972	$36,223	98.97%
$14,315	Acme Silver	±0.2	0.00711	14.064	0.671%	0.702%	702	5	$9,472	$24,558	99.30%
$19,429	Acme Gold	±0.2	0.00677	14.779	0.639%	0.667%	667	4	$9,005	$23,346	99.33%

Widets Produced =	100,000
Cost per Widget =	$13.50
Cost to Retest Individual Pieces =	$35.00
Production cost =	$1,350,000.00

Total Rejections	Falsely Rejected	Falsely Rej as %
-	-	
-	-	
24,970	250	0.250%
10,611	106	0.106%
1,035	10	0.010%
702	5	0.005%
667	4	0.004%



TUR’s vs Total Cost due to false rejection & retest



Sample Quiz

What is the Definition of Metrological Traceability?

Metrological Traceability: Property of a measurement result whereby the result can be related to a reference through a ***documented unbroken chain of calibrations, each contributing to the measurement uncertainty.***

Define a Decision Rule?

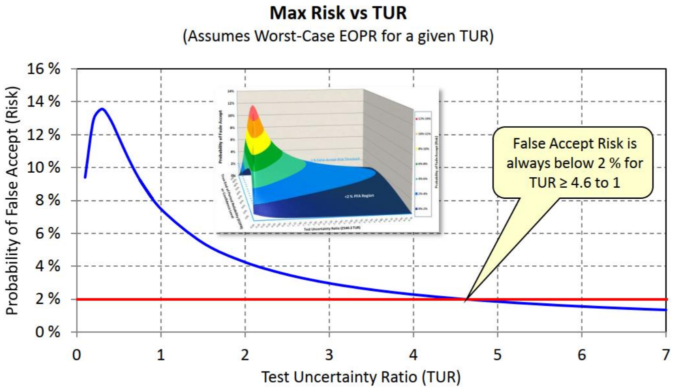
Decision Rule that describes how measurement uncertainty is accounted for when stating conformity with a specified requirement. *(ISO/IEC 17025:2017 3.7 a rule that describes how measurement uncertainty will be accounted for when stating conformity with a specified requirement).*

Sample Quiz

Describe the Difference Between Specific Risk and Global Risk

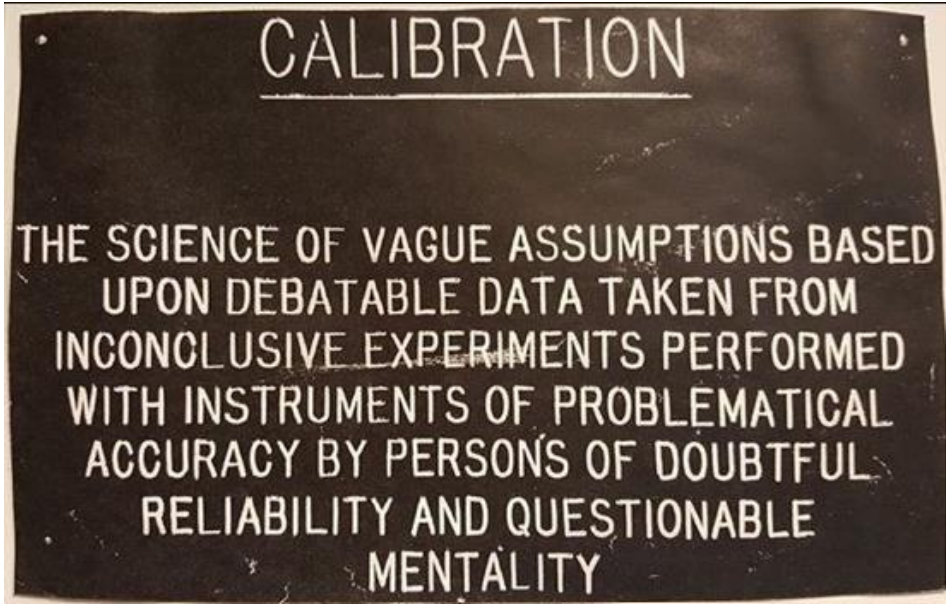
Specific (Bench-Level) Risk mitigation can be thought of as "controlling the quality of the workpieces," while **Program Level (Global) Risk** strategies are described as "controlling the average quality of workpieces."

What is the Measurement Capability Index (TUR) that will limit the Probability of Non-Conformance (FAR) risk to less than 2 %?



"Trust but Verify"

Or as Henry likes to say "Cut Twice, Measure Once" - ☺



What are Some of the Things We Can Control to Mitigate Our Risk?

We can raise our tolerance if we do not need what the manufacturer states.

We can decrease the time between calibrations

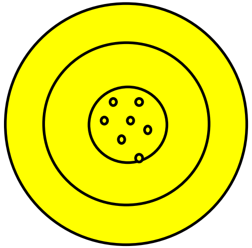
Maybe it is a matter of a too coarse resolution where a different indicator would help.

If we want to observe high in-tolerance probability, we will need to have one of these conditions met

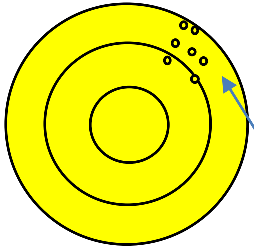
- An extremely good UUT, and an acceptable Reference Standard, providing a minuscule PFA (e.g., 0.01 %), or
- A relatively good UUT, and a good Reference Standard, providing an acceptable PFA (e.g. < 2.0 %)

- S. Mimbs Rule of 89 paper

Accuracy and Precision



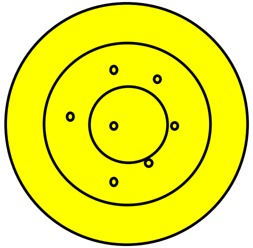
High Precision (Small Random Error)
High Accuracy (Low Bias)



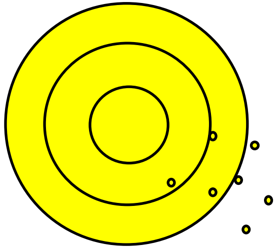
High Precision
Low Accuracy (High Bias)

This is what we see happening a lot and the reason for this discussion.

A precise instrument with a known Systematic Error



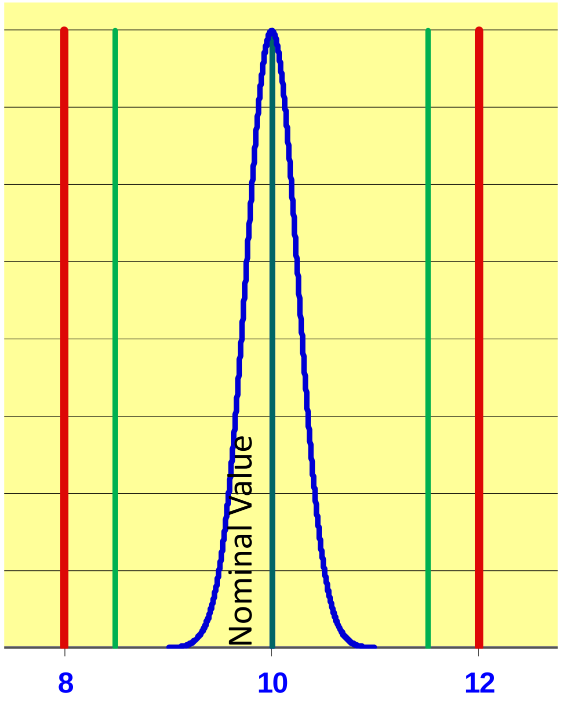
Low Precision (Large Random Error)
High Accuracy (Low Bias)



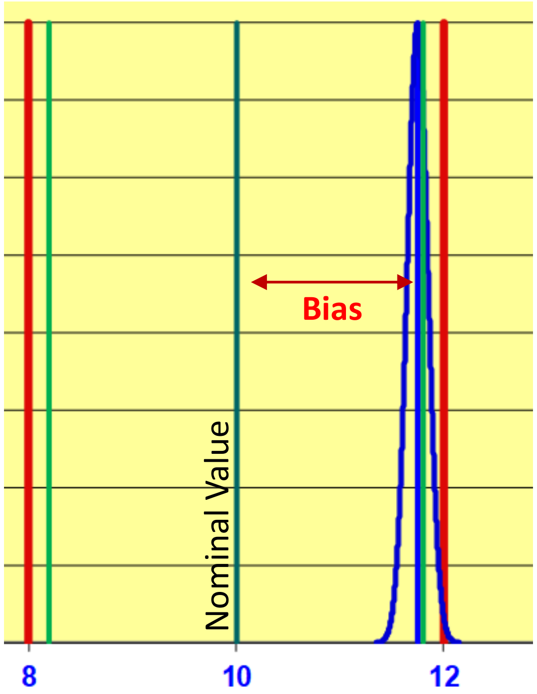
Low Precision
Low Accuracy (High Bias)

Instrument Measurement Uncertainty Guard Banding

Nominal Value of **10**
Measured Value of **10**, **No Bias**

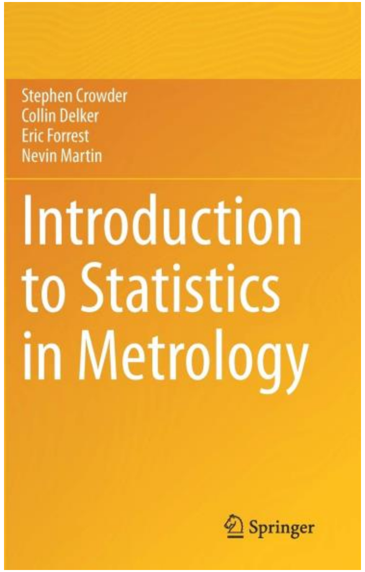


Nominal Value of **10**
Measured Value of **11.75**, **Bias**



Bias – Centered Measurement

- Page 92 Section 5.2 Introduction to Statistics in Metrology



5.2.1.5 Risk with Biased Measurements

While the 4:1 TUR requirement is commonly used to ensure a measurement is adequate for making an accept/reject determination, this metric assumes that the process distribution is centered between the specification limits, that is $\mu_p = (SL_U + SL_L)/2$. If this is not the case, TUR cannot be reliably used as an indicator of risk, however, the PFA and PFR equations are still valid assuming the correct μ_p is used.

The measurement uncertainty distribution is also assumed to be centered about the actual value t when calculating TUR. The measurement process is said to be biased if it is not centered about t and systematically overstates or understates the true value of the measurement. Properly accounting for measurement bias provides a more accurate risk evaluation. If bias is ignored, the risk might be understated, perhaps significantly.

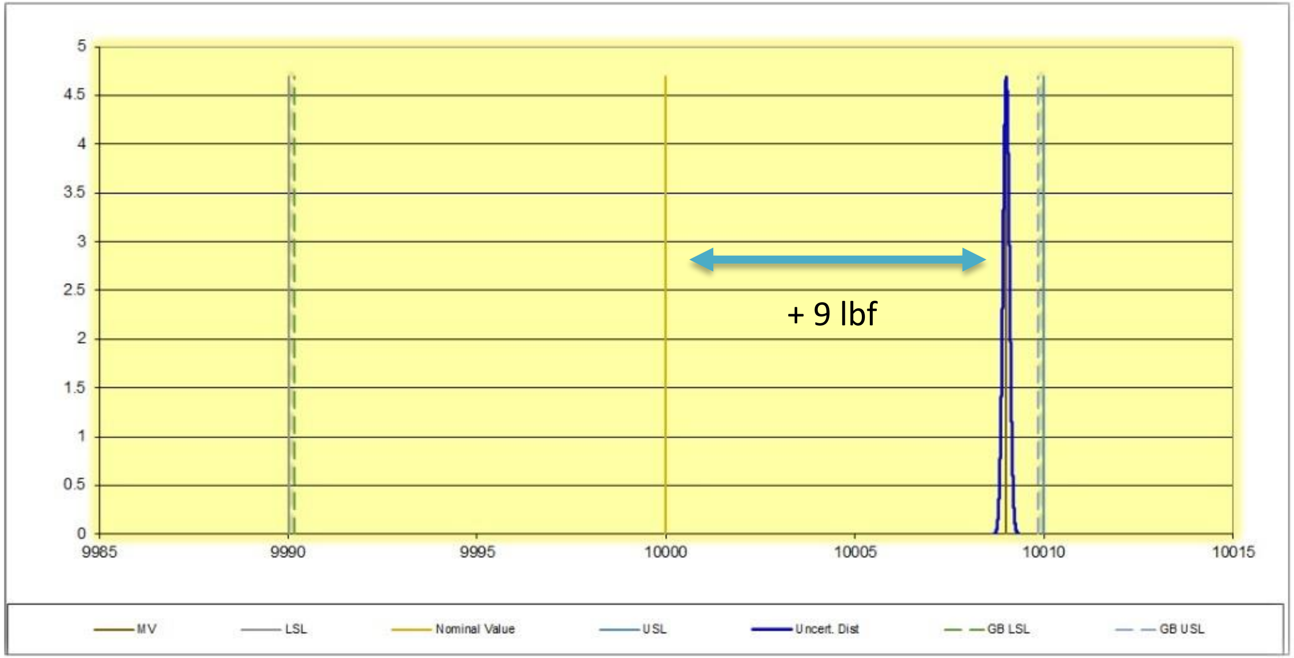
In the presence of bias, the distribution of the measurement y, given the actual value t, shifts from a $N\left(t, \sigma_m^2\right)$ distribution to a $N\left(t - b_m, \sigma_m^2\right)$ distribution, where b_m is the measurement bias.

With bias b_m, the expressions for the PFA and PFR (without guardbanding) become

$$\begin{aligned} \text{PFA} = &\int_{-\infty}^{SL_L} \left(\int_{SL_L}^{SL_U} \frac{1}{\sigma_m \sqrt{2\pi}} e^{-\frac{1}{2\sigma_m^2}(y-(t-b_m))^2} dy \right) \frac{1}{\sigma_p \sqrt{2\pi}} e^{-\frac{1}{2\sigma_p^2}(t-\mu_p)^2} dt \\ &+ \int_{SL_U}^{+\infty} \left(\int_{SL_L}^{SL_U} \frac{1}{\sigma_m \sqrt{2\pi}} e^{-\frac{1}{2\sigma_m^2}(y-(t-b_m))^2} dy \right) \frac{1}{\sigma_p \sqrt{2\pi}} e^{-\frac{1}{2\sigma_p^2}(t-\mu_p)^2} dt. \end{aligned} \qquad (5.18)$$

Instrument Measurement
+ 9 lbf Bias

Nominal Value	**10000.0**
Lower specification Limit	9990.0
Upper Specification Limit	10010.0
Measured Value	**10009.0**
Measurement Error	9.0
Std. Uncert. (k=1)	**0.085**
Total Risk	**0.00%**
Upper Limit Risk	0.000%
Lower Limit Risk	0.000%
TUR =	**58.78943644**
Cpk=	**5.999032319**
TAR=	**62.5**
Simple Guard Band (Subtract Uncertainty)	
Guard Band LSL	**9990.170**
Guard Band USL	**10009.8299**
Percent of Spec	**98.30%**
Guard Band Limits for Risk of	**2.500%**
Guard Band LSL	**9990.167**
Guard Band USL	**10009.833**
Percent of Spec	**98.33%**



Graph Showing 10 009.0 as the measured value with a 58.789:1 TUR, which is achieved by using a lab with low uncertainties (Morehouse actual example) There is a bias of + 9 lbf in this example.

Force Applied	Measurement Value	Offset, Bias ,Systemic Measurement Error
10 000.00	10 009.00	+ 9
10 000.00	10 009.00	+ 9

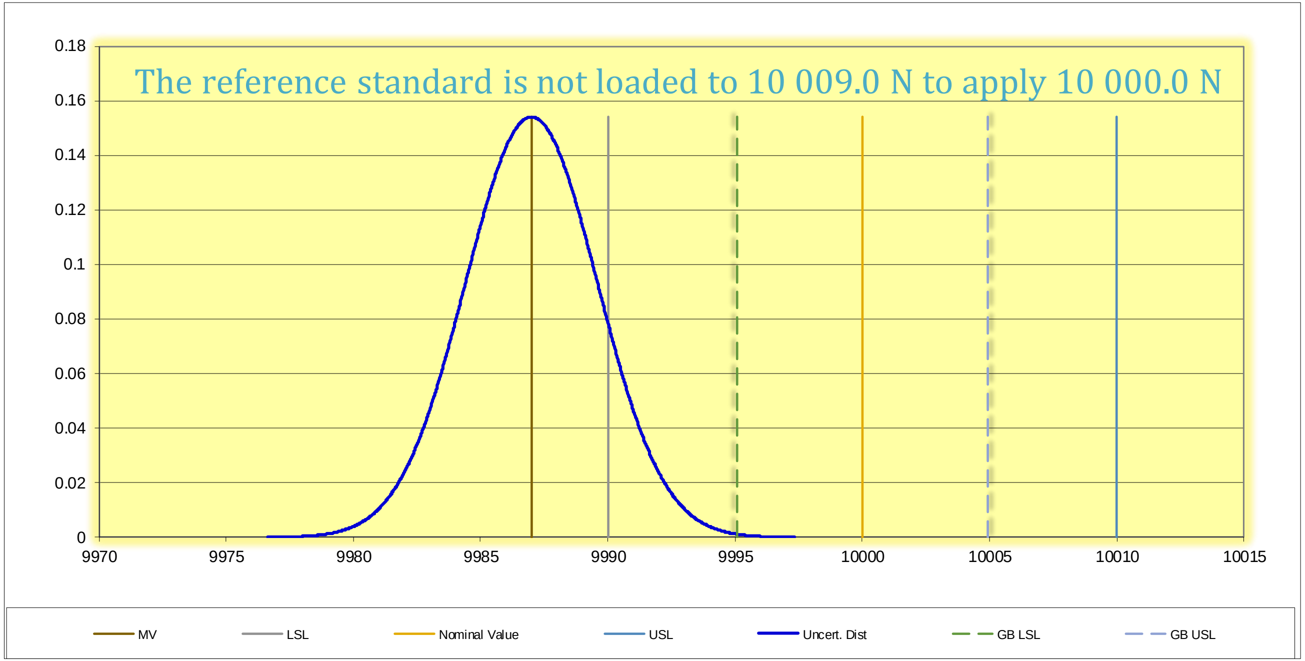
When you know the value to generate 10 000.0 N is 10 0009.0 N.

The right thing for the end-user to do is to load the device to 10 009.0 N to apply 10 000.0 N of force.

What Happens When We Do Not Correct **the** Bias?

Let us assume they do not do that and use this device to calibrate another 10,000 N instrument.

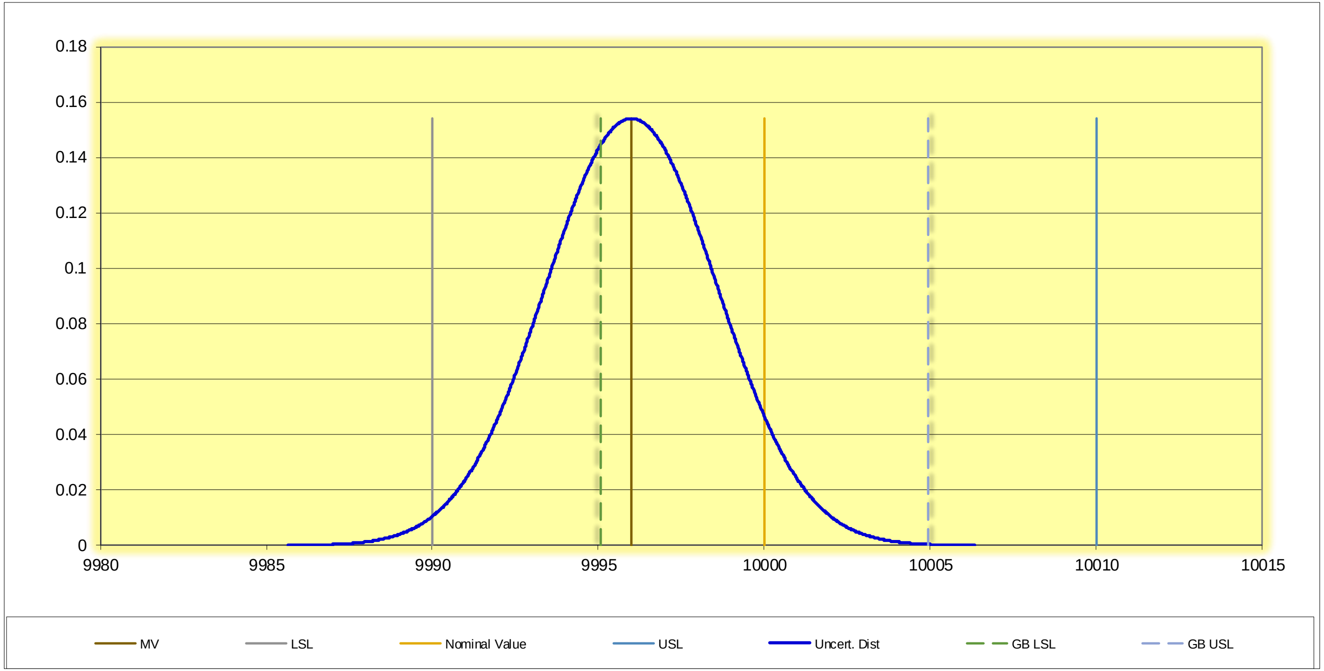
Nominal Value	**10000.0**
Lower specification Limit	9990.0
Upper Specification Limit	10010.0
Measured Value	**9987.0**
Measurement Error	-13.0
Std. Uncert. (k=1)	**2.589**
Total Risk	**87.67%**
Upper Limit Risk	0.000%
Lower Limit Risk	87.672%
TUR =	**1.931223436**
Cpk=	**-0.59120171**
TAR=	**3.99840064**
Simple Guard Band (Subtract Uncertainty)	
Guard Band LSL	**9995.178**
Guard Band USL	**10004.8219**
Percent of Spec	**48.22%**
Guard Band Limits for Risk of	**2.500%**
Guard Band LSL	**9995.074**
Guard Band USL	**10004.926**
Percent of Spec	**49.26%**



What Happens When We Correct the Bias?

The right thing for the end-user to do is to load the device to 10 009.0 N to apply 10 000.0 N of force. When this practice is followed, the DUT is now in specification.

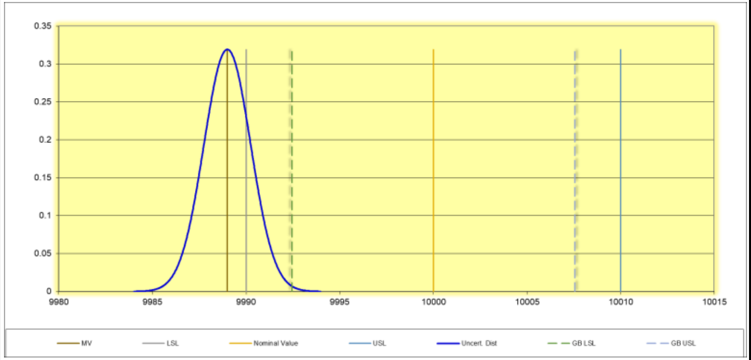
Nominal Value	**10000.0**
Lower specification Limit	9990.0
Upper Specification Limit	10010.0
Measured Value	**9996.0**
Measurement Error	-4.0
Std. Uncert. (k=1)	**2.589**
Total Risk	**1.02%**
Upper Limit Risk	0.000%
Lower Limit Risk	1.024%
TUR =	**1.931223436**
Cpk=	**1.182403422**
TAR=	**3.99840064**
Simple Guard Band (Subtract Uncertainty)	
Guard Band LSL	**9995.178**
Guard Band USL	**10004.8219**
Percent of Spec	**48.22%**
Guard Band Limits for Risk of	**2.500%**
Guard Band LSL	**9995.074**
Guard Band USL	**10004.926**
Percent of Spec	**49.26%**



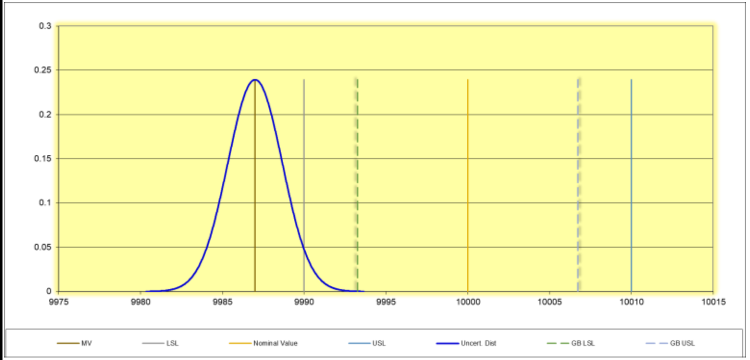
Not Correcting for Bias

Measurement Bias Not Corrected

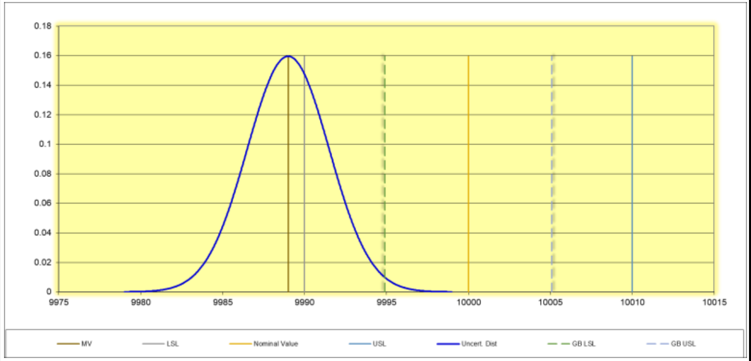
Type	TUR	Bias	Total Risk
Reference	4:1	-9	78.81%



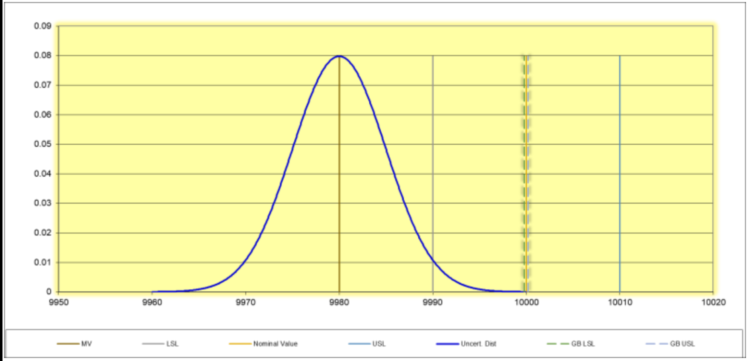
Type	TUR	Bias	Total Risk
Working	3:1	-13	96.41%



Type	TUR	Bias	Total Risk
General	2:1	-11	65.54%



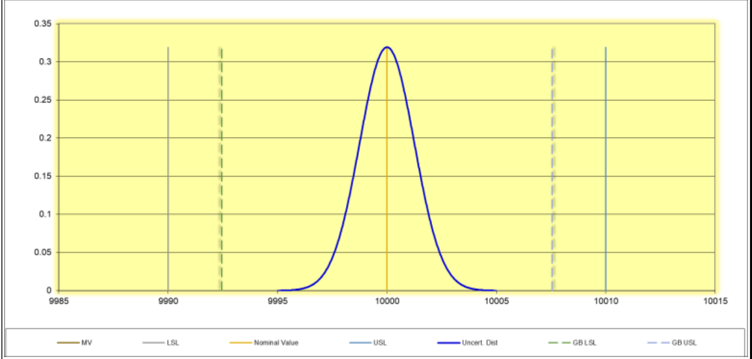
Type	TUR	Bias	Total Risk
Process	1:1	-20	97.25%



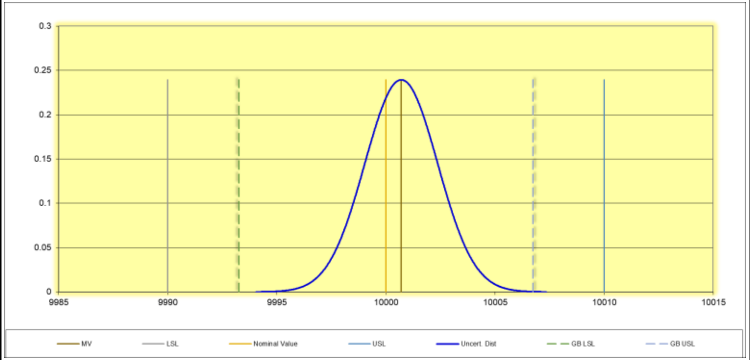
Correcting for Bias

Measurement Bias Corrected

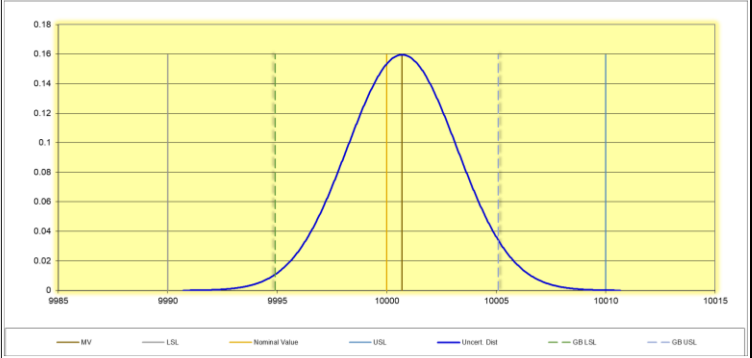
Type	TUR	Bias	Total Risk
Reference	**4:1**	**0**	**0.00%**



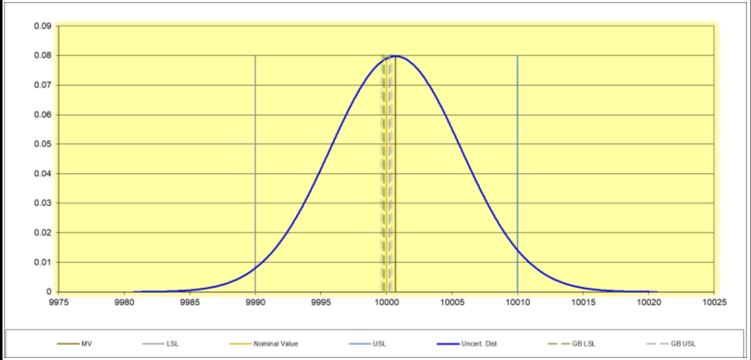
Type	TUR	Bias	Total Risk
Working	**3:1**	**0.7**	**0.00%**



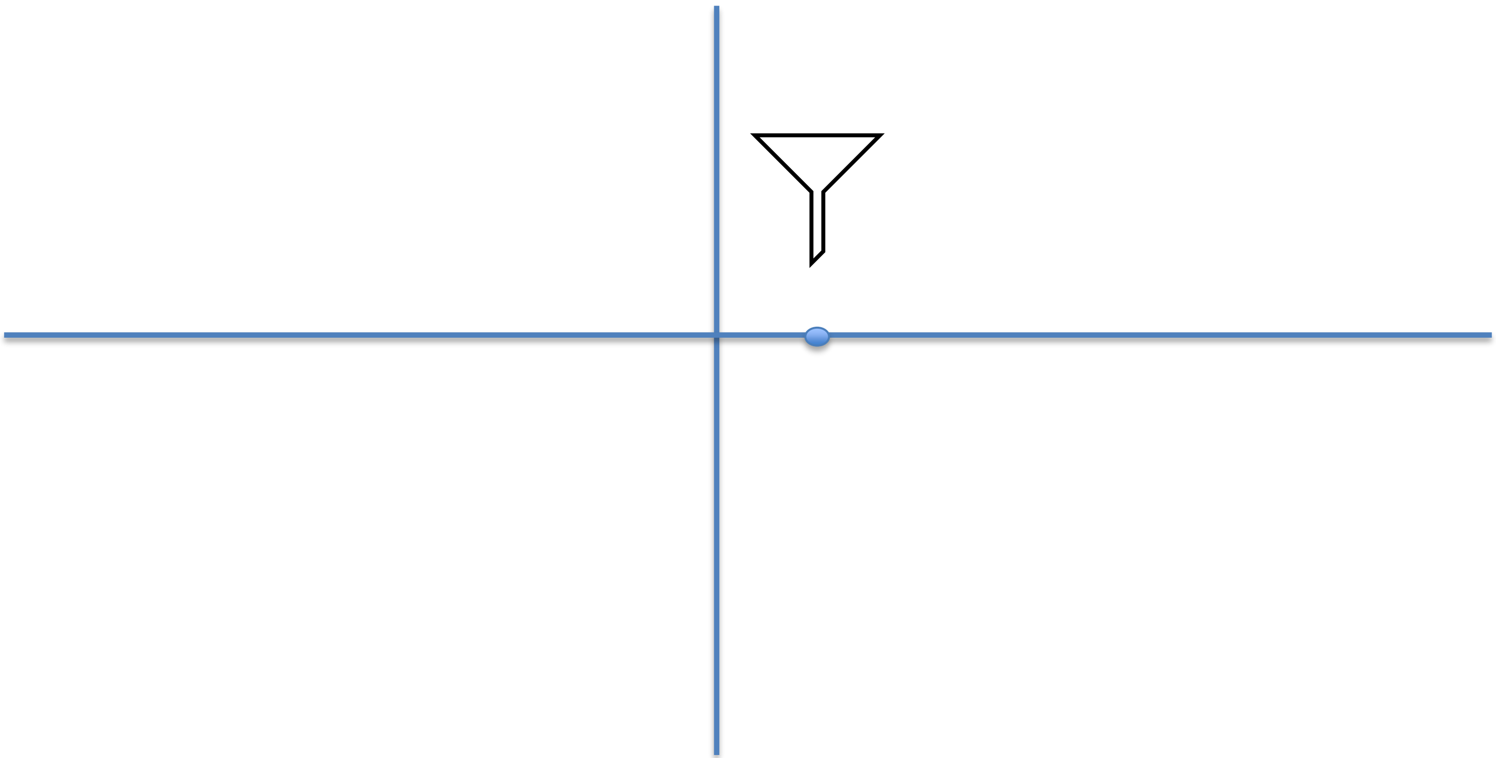
Type	TUR	Bias	Total Risk
General	**2:1**	**0.5**	**0.00%**



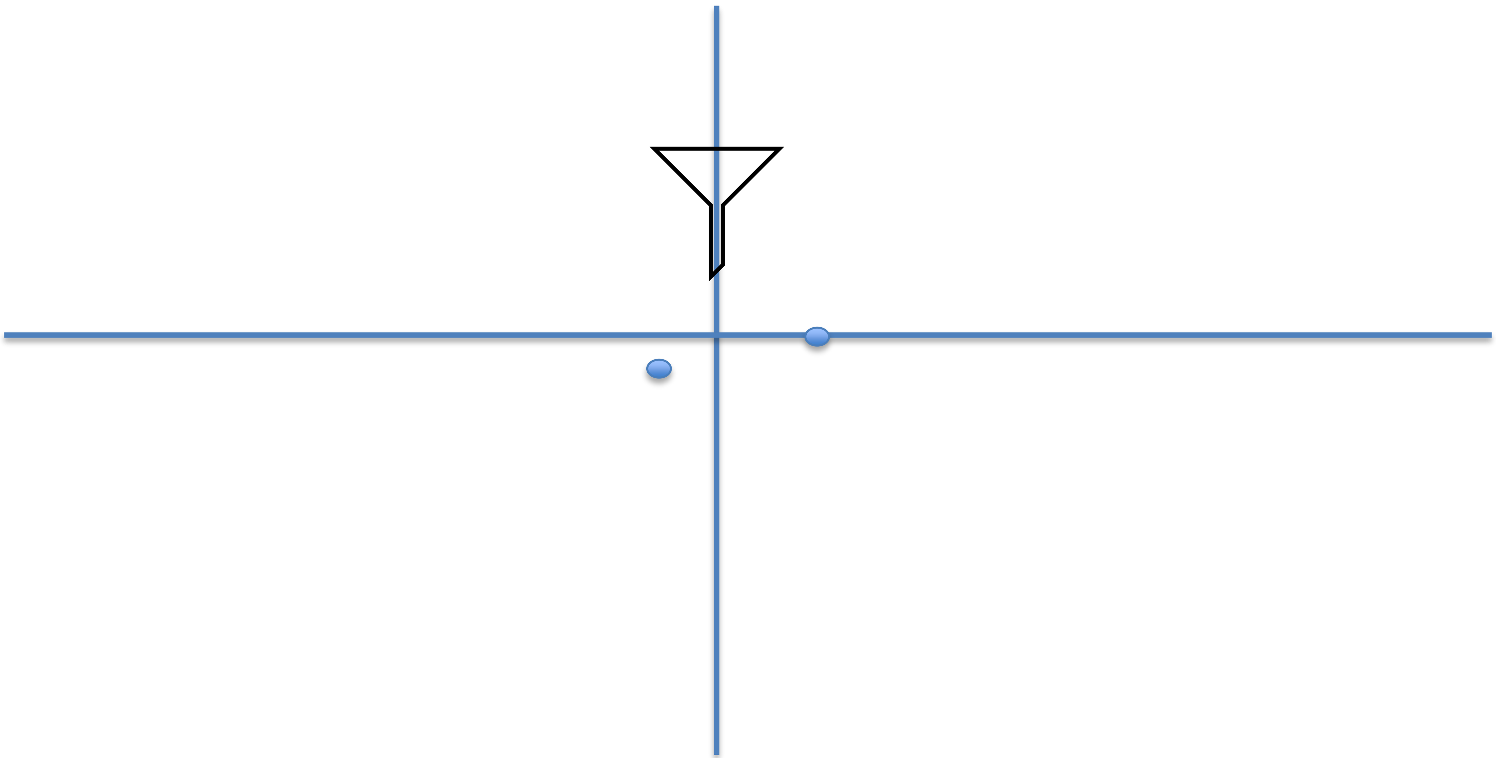
Type	TUR	Bias	Total Risk
Process	**1:1**	**0.6**	**4.76%**



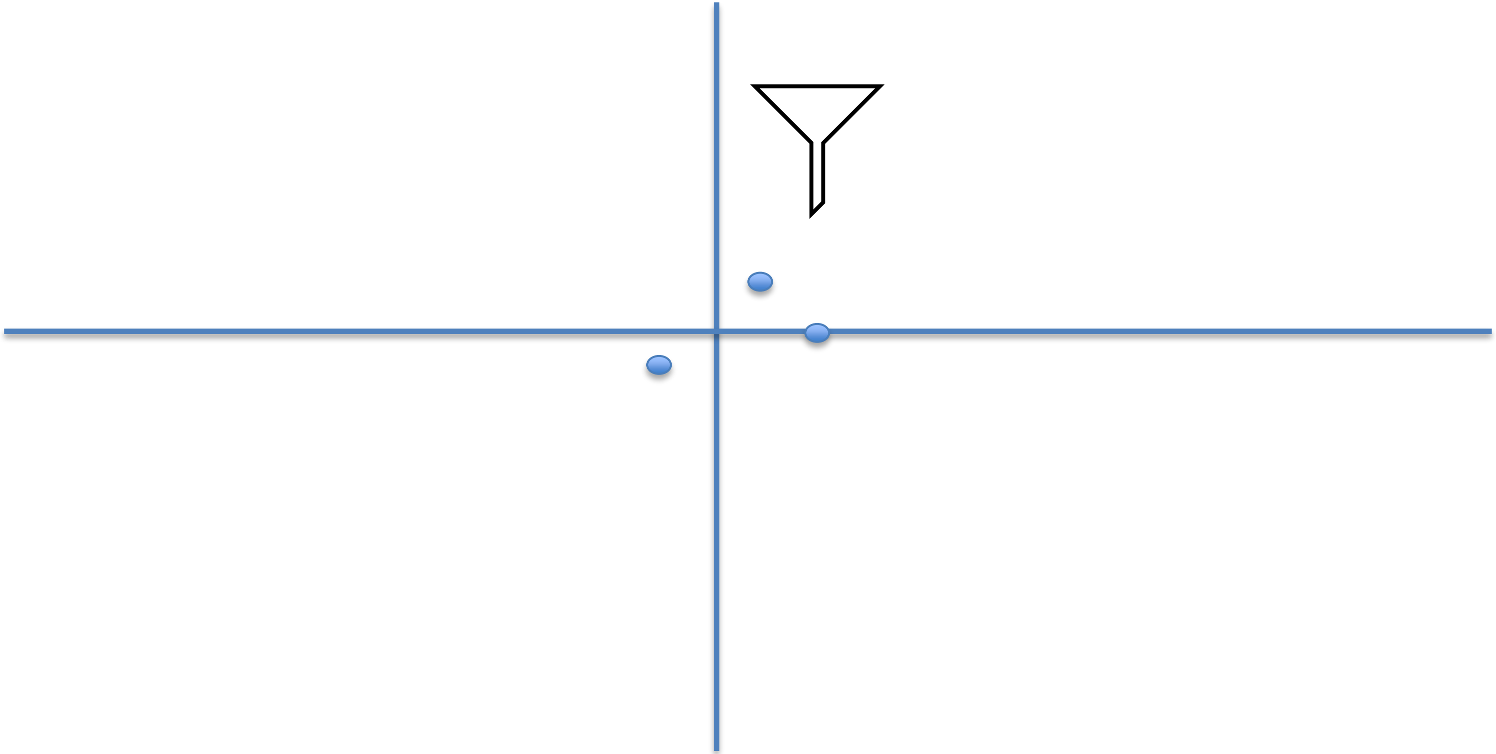
Deming Funnel – Adjust to try and correct any bias



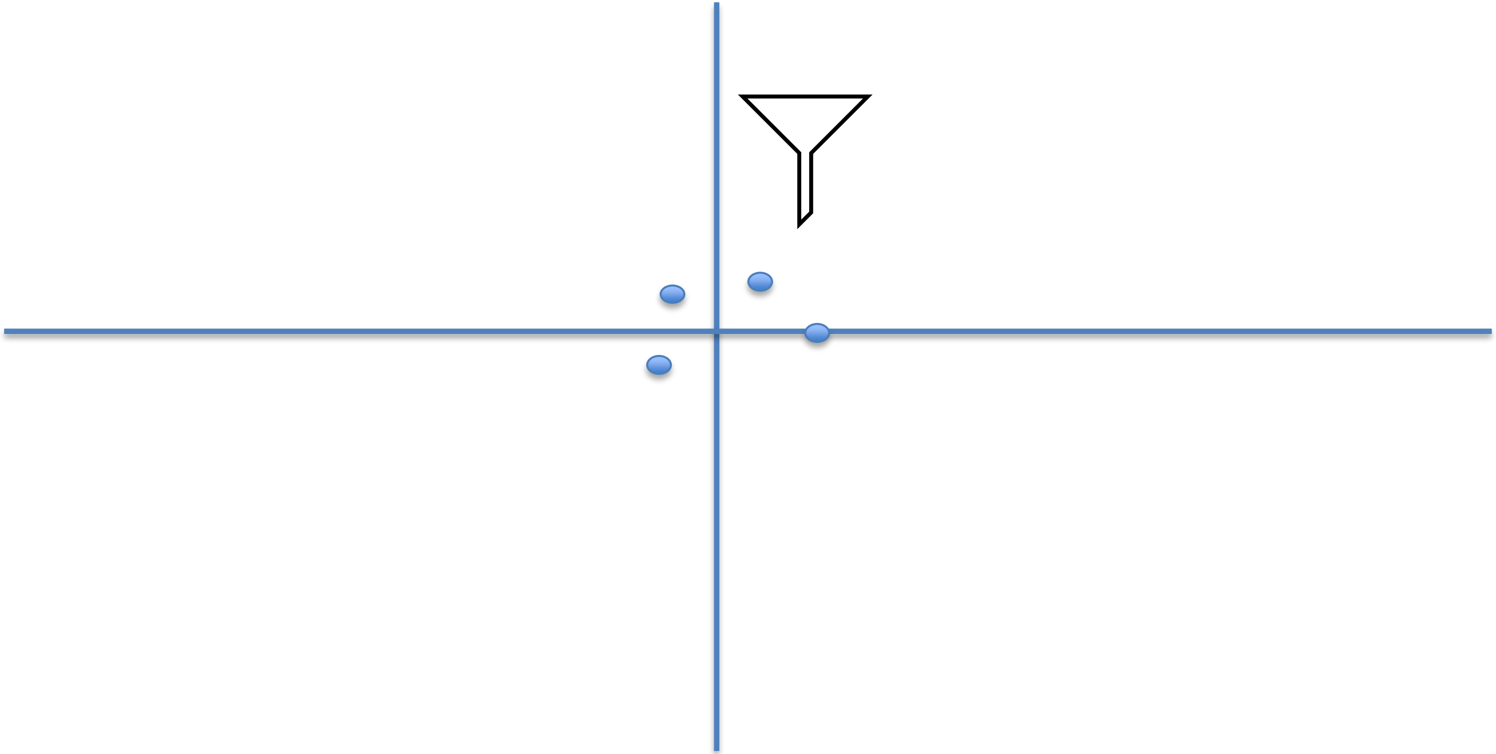
Deming Funnel – Adjust to the Measurement Point



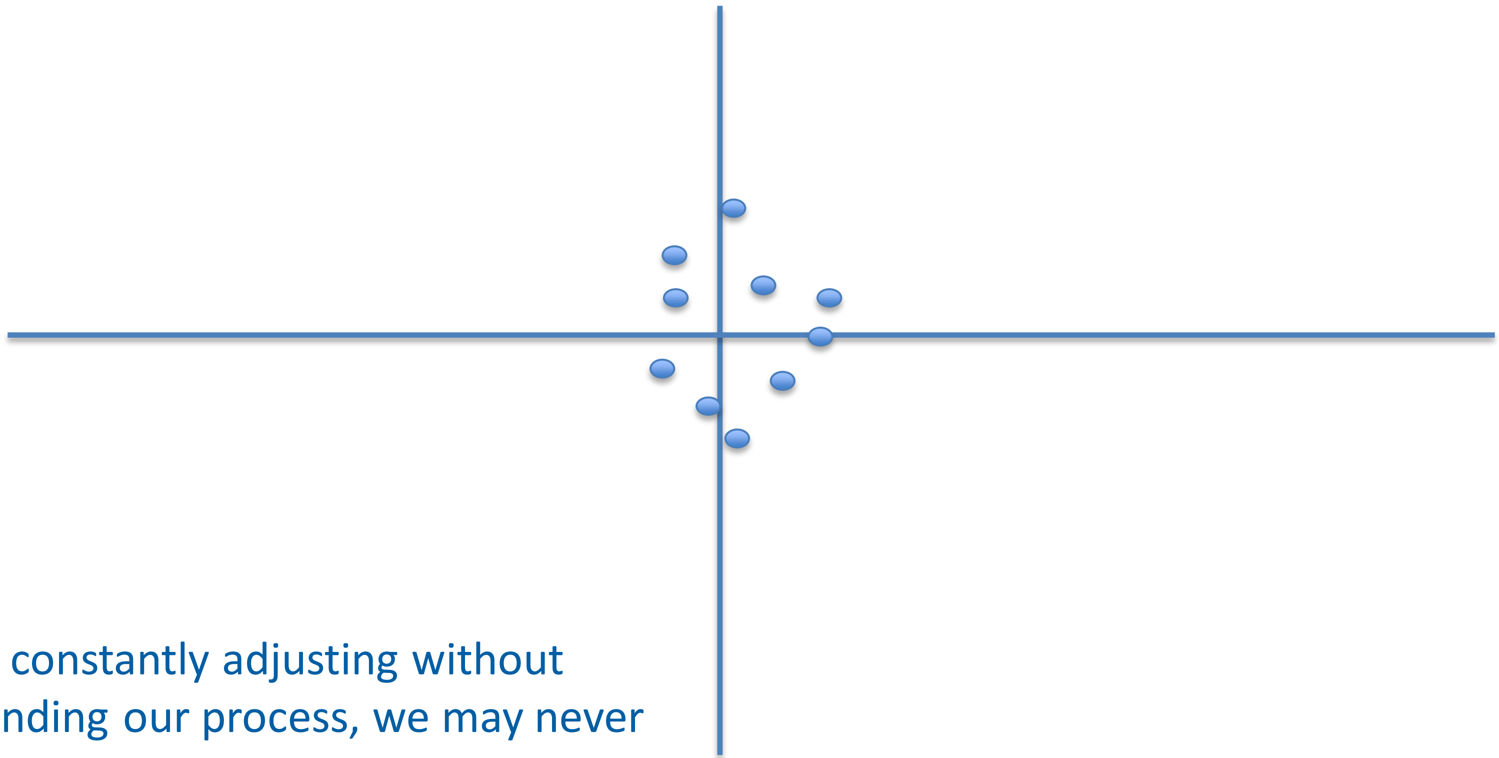
Deming Funnel – Adjust to the Measurement Point



Deming Funnel – Adjust to the Measurement Point

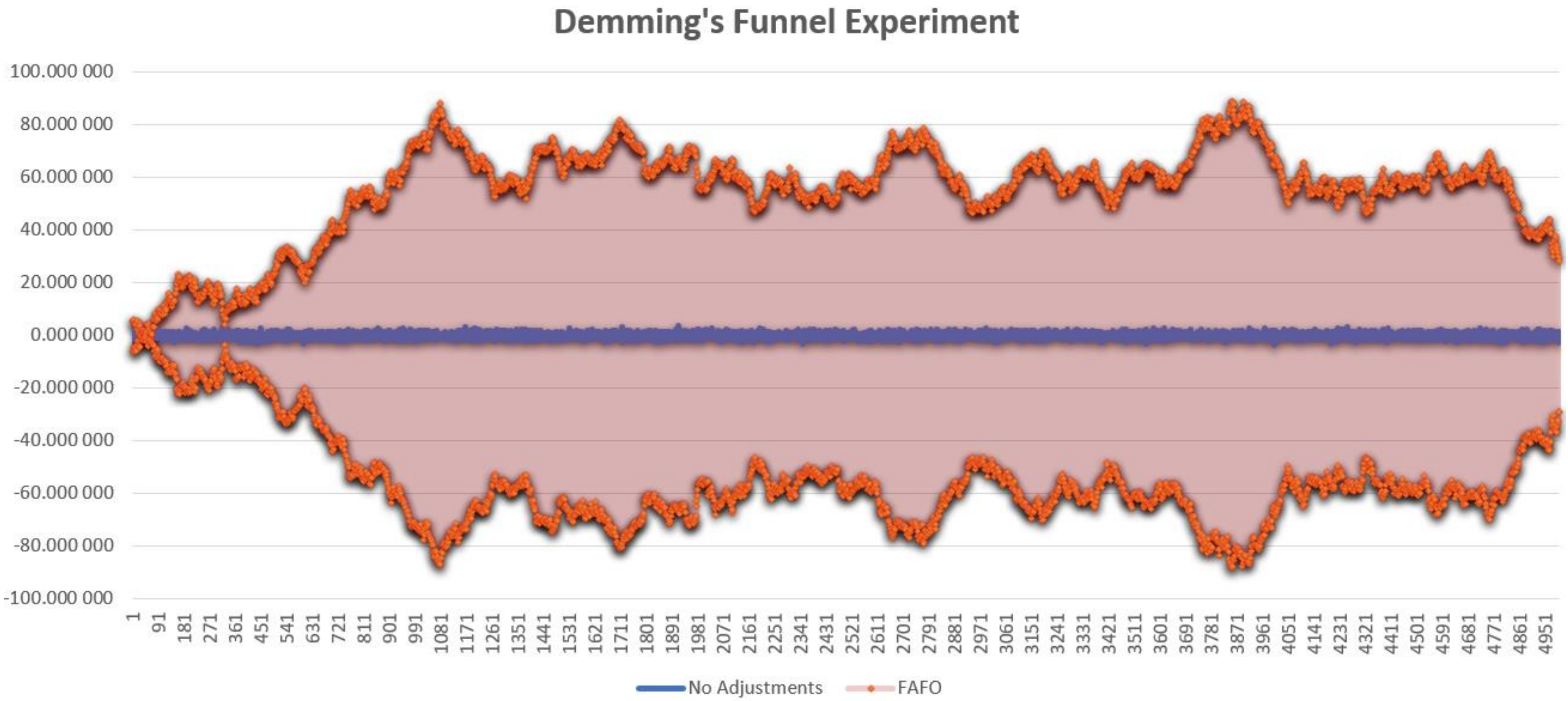


Deming Funnel – Adjust to the Measurement Point



If we are constantly adjusting without understanding our process, we may never hit the target.

Deming Funnel – Adjust to the Measurement Point



If we are constantly adjusting without understanding our process, we may never hit the target.

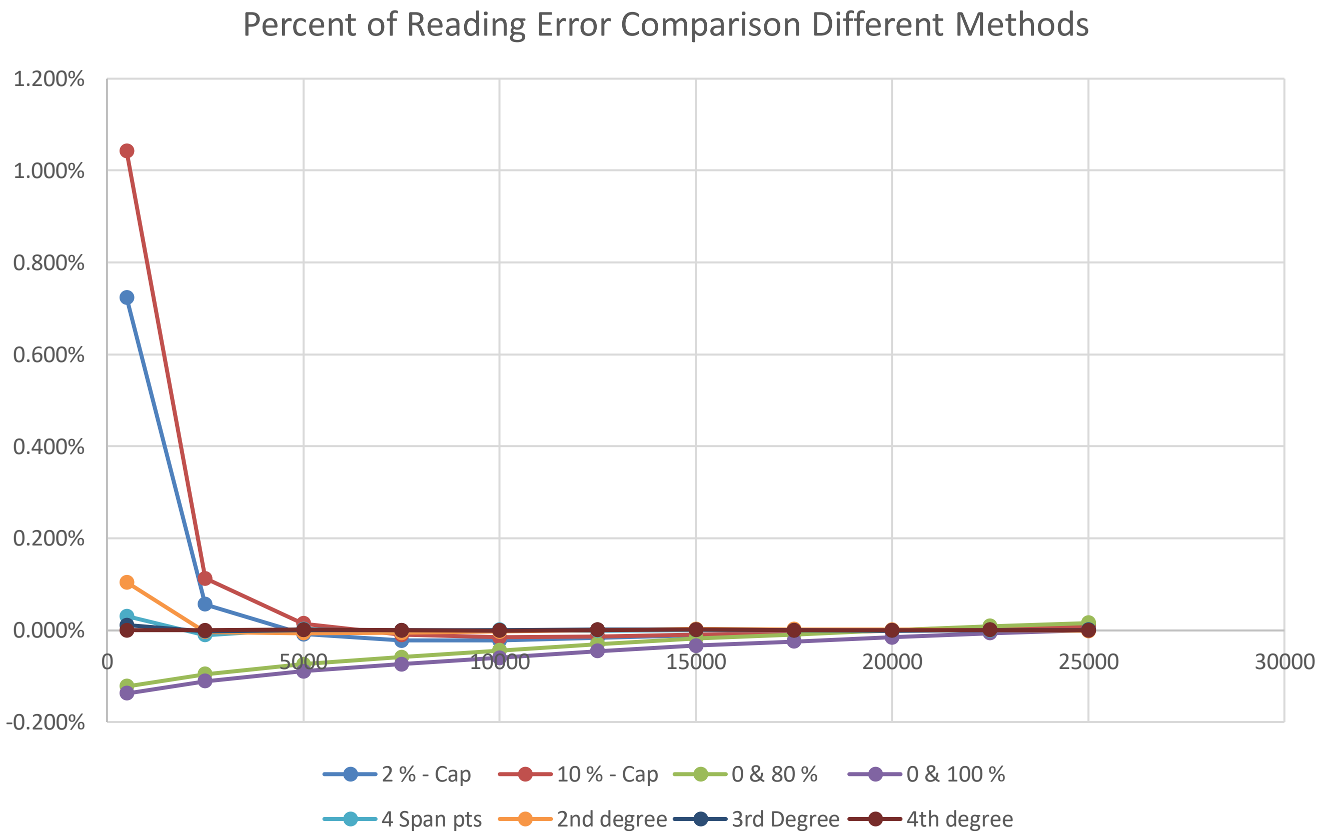
Solution for Force Measurements

Morehouse has many options with our force calibrations systems that use coefficients generated at the time of calibration. Our 4215 plus and C705P use coefficients that are programmed into the indicator to help correct and minimize measurement bias.

Solution for Force Measurements

		Indicator with 2-pt adjustments			Using Coefficient Conversion			
Applied Force lbf	Actual Readings (mV/V)	Programmed Points	Calculated Values 2 pt span	Error	Calculated Values polynomial	Error	Diff in Errors	% difference
200	0.08279		199.6	0.4	199.9	0.1	0.25	189%
1000	0.41415	0.41415	1000.0	0.0	999.9	0.1	-0.11	116%
2000	0.82851		1997.6	2.4	1999.9	0.1	2.26	1846%
3000	1.24302		2997.0	3.0	2999.9	0.1	2.82	2109%
4000	1.65767		3996.8	3.2	3999.9	0.1	3.06	2413%
5000	2.07242		4996.8	3.2	4999.9	0.1	3.05	2180%
6000	2.48726		5997.0	3.0	5999.9	0.1	2.83	2060%
7000	2.90216		6997.4	2.6	6999.9	0.1	2.47	1856%
8000	3.31709		7997.8	2.2	7999.9	0.1	2.02	1446%
9000	3.73203		8998.3	1.7	8999.9	0.1	1.56	1055%
10000	4.14696	4.14696	9998.7	1.3	9999.9	0.1	1.12	776%

Solution for Force Measurements



Bias Conclusion

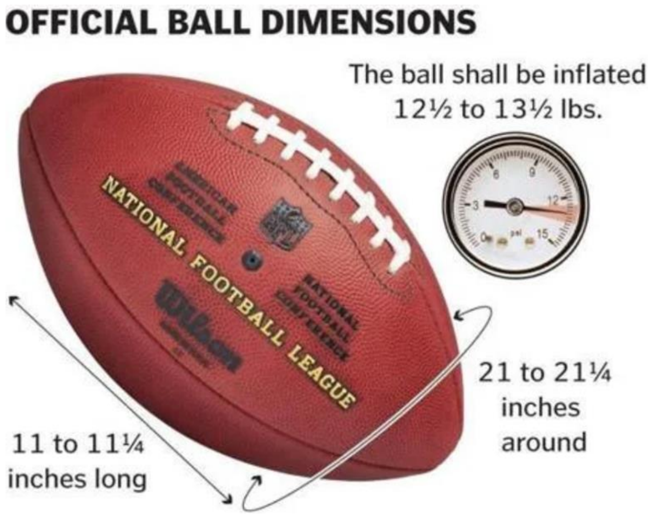
- Not correcting for bias seems to be a problem many in the calibration deal with, and their unsuspecting customers are likely getting calibrations that carry too much overall Measurement Risk.
- The habit of insisting on a 4:1 TUR assumes the measurement process is centered (measurement bias is corrected).
- When bias is not corrected, the risk of making a measurement that does not properly account for bias can result in an underestimation of measurement uncertainty and therefore disagrees with the metrologically traceability definition and undermines measurement confidence.

Case Study– "Deflate Gate"

- Deflate gate suggested that the New England Patriots used an illegal process for lowering the inflation of game footballs at the behest of quarterback Tom Brady

NFL Rulebook (Goodell 2014) states "The ball *shall* be made up of an inflated (12.5 to 13.5 pounds) urethane bladder enclosed in a pebble grained, leather case (natural tan color) without corrugations of any kind. It *shall* have the form of a prolate spheroid, and the size and weight *shall* be:

Long axis = 11 to 11.25"
Long circumference = 28 to 28.5"
Short circumference = 21 to 21.25"
Weight = 14 to 15 oz.



Case Study– “Deflate Gate”

- The NFL Chose to use the following gauges - One “no name” the other model CJ-01 manufactured for Wilson by Jiao Hsiung Industry Corp. (Exponent findings 2015)
- The process: two measurements were taken on each game ball (11 balls in total) at halftime, with a different gauge and operator used for each. Degrees of freedom = 1
- Although both gauges likely produced by Jiao Hsiung Industry Corp (JHIC), Wilson has no stated accuracy. The display reads ±0.05 PSIG (the last digit is either 0 or 5)
- Similar gauges have a stated accuracy of ±1% of Full Scale (FS) which equates to ±0.2 PSIG where FS = 20 PSIG – <u>we will assume this is the accuracy of the game gauges</u>
- Neither gauge used in the game had a traceable calibration, which makes the specification difficult to prove and therefore the true accuracy is likely worse

Case Study– “Deflate Gate”

- At best, that gauge can provide ±3.3 PSIG (~0.817 x 4) uncertainty (assuming a 4:1 TUR desired) – it’s 6.6x less accurate than the NFL requirement of ±0.5 PSIG

		u(Accuracy)	u(StDev)	u(Resolution)		
		0.11547	0.15000	0.01443		
	df	100000.00	1.00	100000.00		
Measurement Equation Inputs	**Value**					
Accuracy	0.20000	0.31547	0.20000	0.20000		
StDev	0.15000	0.15000	0.30000	0.15000		
Resolution	0.05000	0.05000	0.05000	0.06443		
Result	0.40000	0.51547	0.55000	0.41443		
		0.11547	0.15000	0.01443	$c_i u_i$	
u_c =	0.18985	0.01333	0.02250	0.00021	$(c_i u_i)^2$	
df =	2.57	37.0%	62.4%	0.6%	<--rel $(c_i u_i)^2$	100.0%
k =	4.303	1.00000	1.00000	1.00000	c_i	Σrel $(c_i u_i)^2$
U =	0.81684		204.21%	$U_{relative}$, %		

"Deflate Gate"

Risk Calculator	
Upper Tolerance T_U	13.5
Lower Tolerance T_L	12.5
Nominal Value (NV)	13
Measured Value xm	13.0000
Measurement Unc um	0.4084
Maximum Allowable Risk (PFA)	2.0000%
Tolerance T	1.00

In-Tol Probability with given U_{c95} (as is)	*77.91%*
Probability of non-conformance	*22.09%*

Probability of Conformance (p_c)	77.913%
Probability of NonConformance (1 - p_c)	22.087%

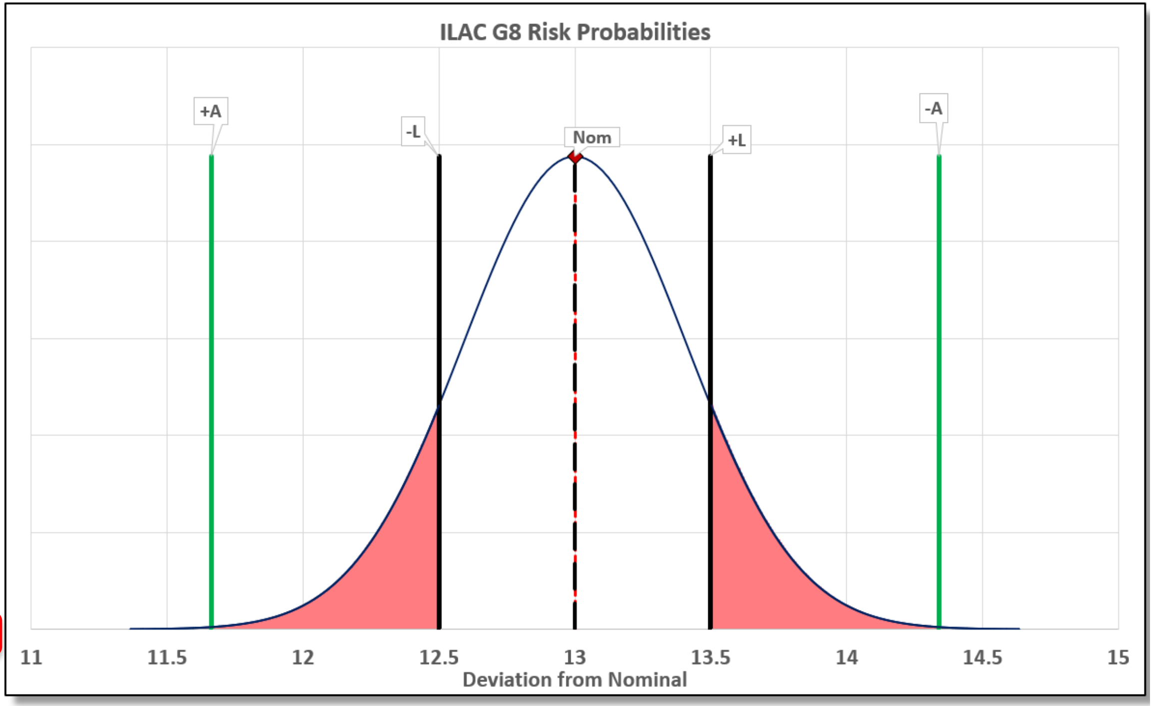
Setting the Guard Band Upper and Lower AL	
Guard Band Upper G_u (AL= TL - w)	13.5000
Guard Band Lower G_L (AL = TL + w)	12.5000
Relaxed Upper Acceptance Limit	14.3388
Relaxed Lower Acceptance Limit	11.6612

Setting AL based on Probability of Conformance	
Probability of Conformance (pc)	98.00%
r	1.0269
w = U_{95} * *r*	0.83879
C_m (TUR)	0.61211

Setting AL based on Guard Band w	
Upper Acceptance Limit	PASS
Lower Acceptance Limit	PASS

Area of Curve Outside of the AL	0.105%
Area of Curve Outside of the TL	22.087%

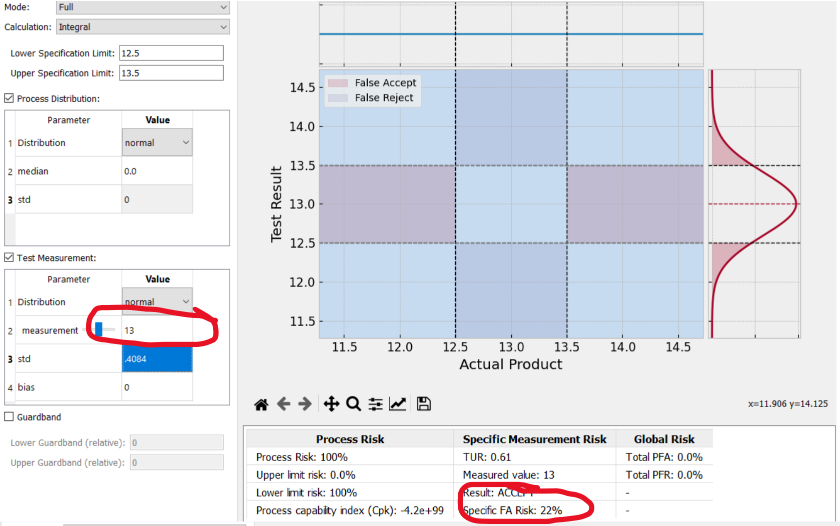
Data Inputs

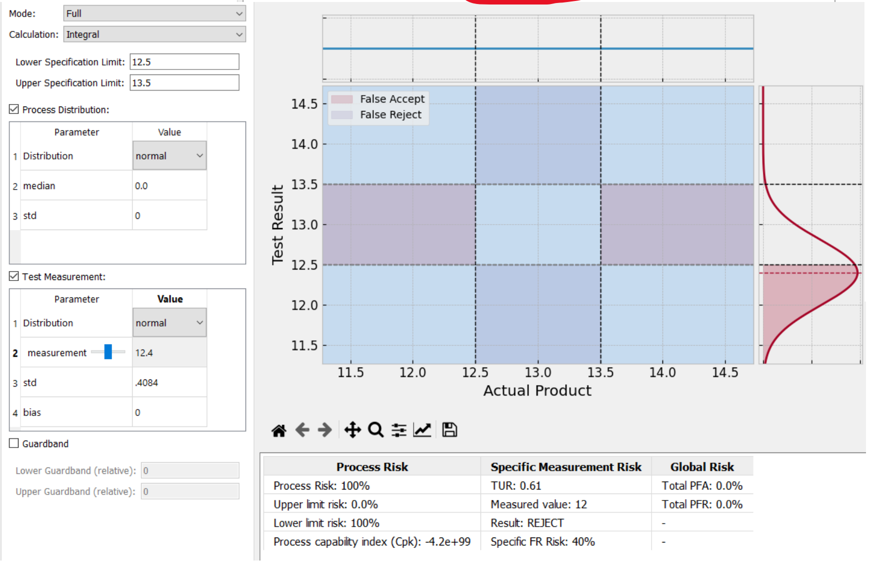


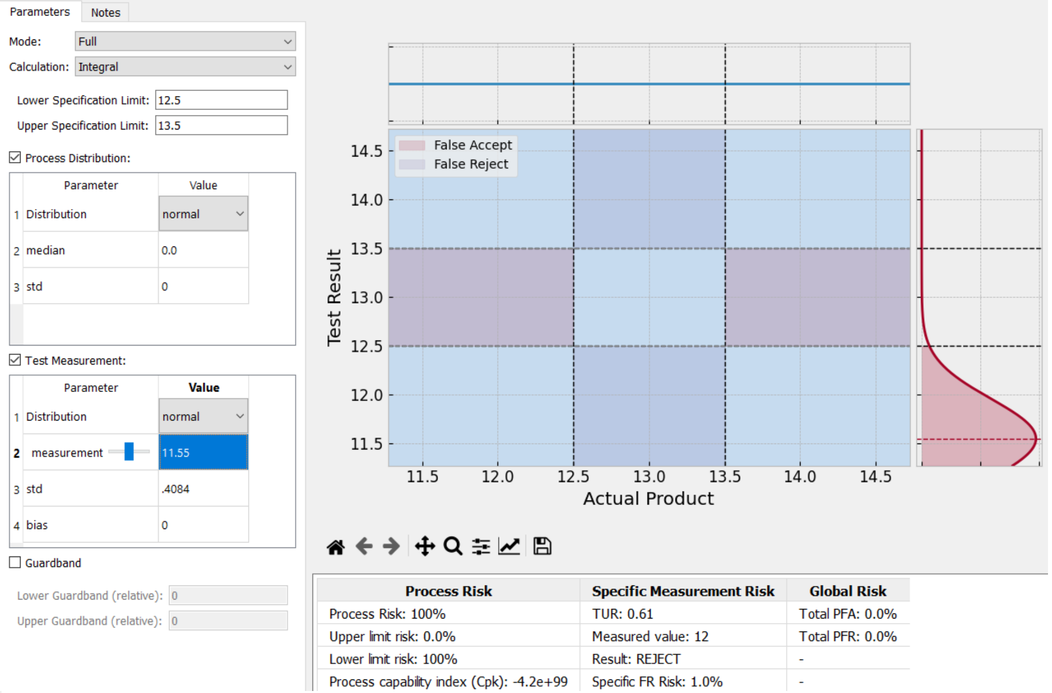
Case Study– "Deflate Gate" Conclusion

- The NFL used an inappropriate instrument to verify the pressure integrity of the game ball
- "Deflate gate" totaled more than $22.5M by end of investigation
- The Additel GP30, at ±0.05% FS (±0.015 psig) costs ~$714 (including an accredited calibration)
- The NFL used a $30 gauge which, at best, is good for measurements ±3.5 psig

What does this slide tell us?





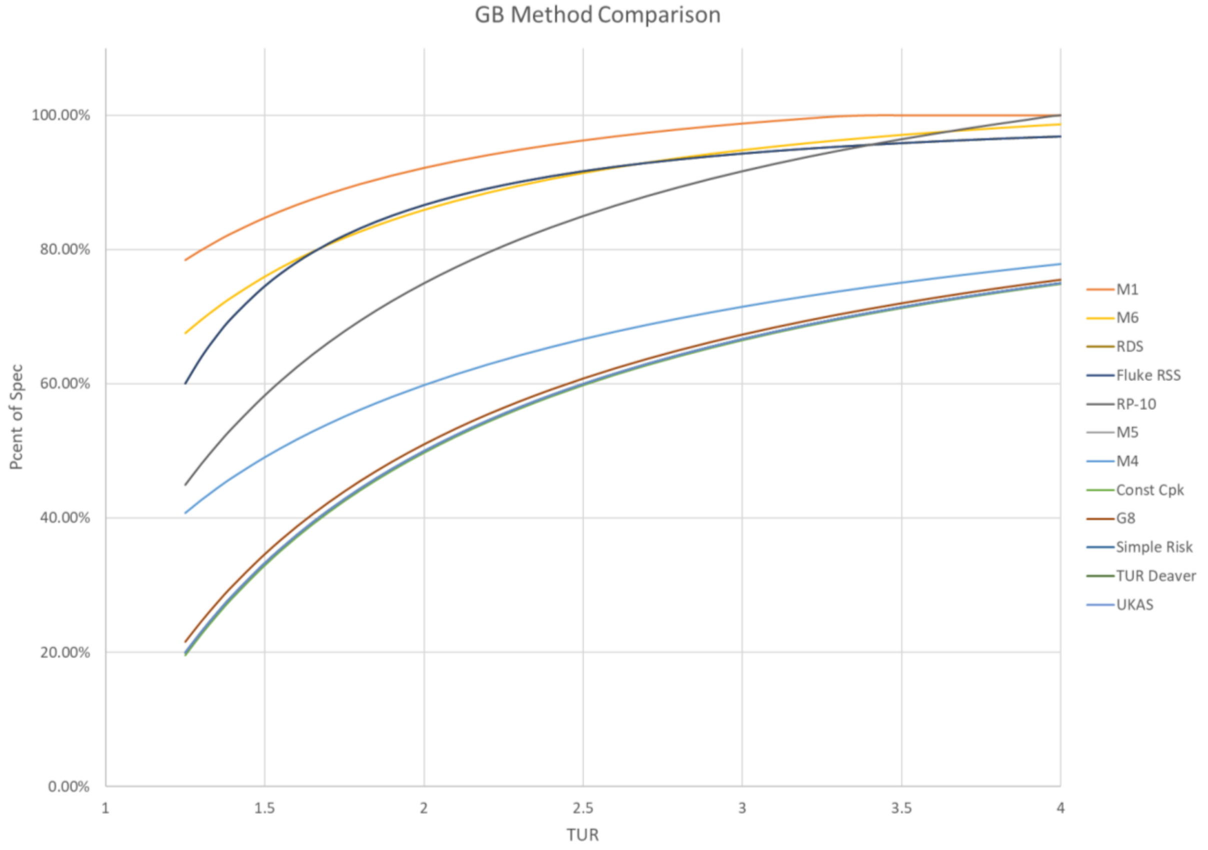


When the measured value is centered (13) the FA risk is 22 %.

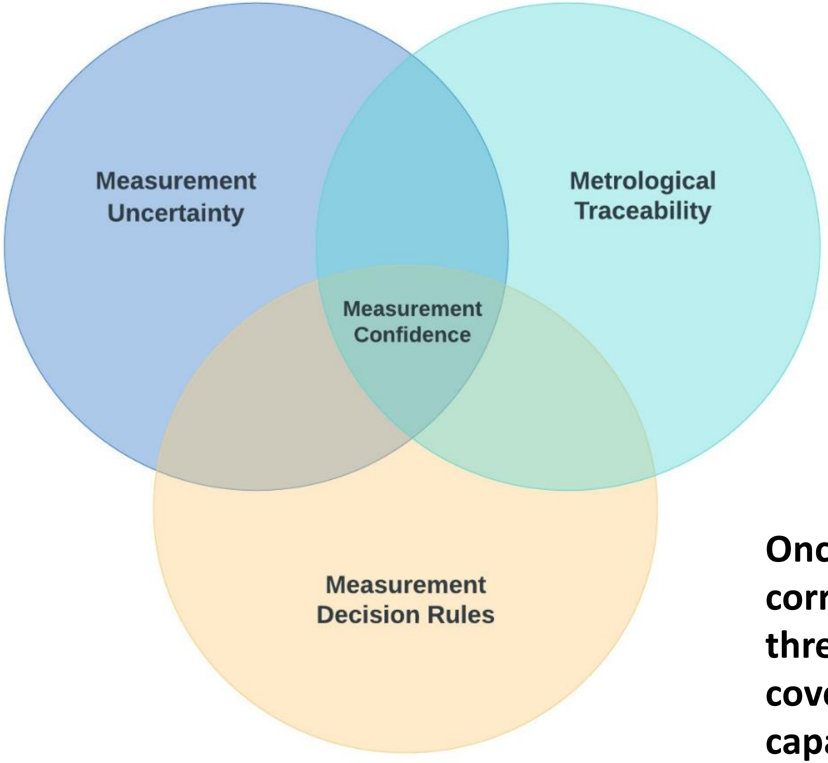
When the measured value is at 12.4 the FR risk is 40 %

At a measured value of 12, the FR risk is 1 %

Selecting the proper Guard Banding method



Measurement Confidence



Once known biases are corrected, and we have these three pillars of measurement covered, we need to prove our capability.

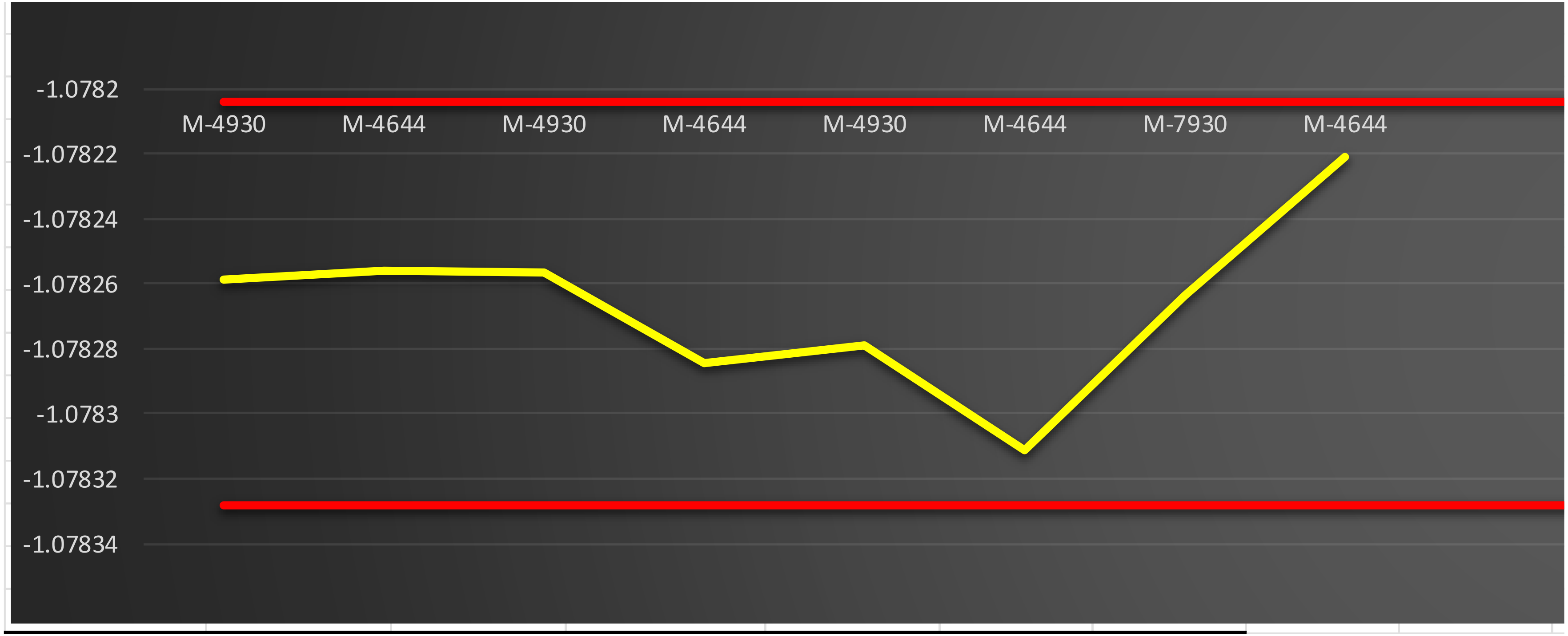
Proficiency Testing Services

https://mhforce.com/calibration/force-ilc-and-pt/

Measurement Assurance
SPC DATA

	2021		2021		2022		March-2023-CK	
1K Load Cell	**M-4930**	**M-4644**	**M-4930**	**M-4644**	**M-4930**	**M-4644**	**M-7930**	**M-4644**
1	-1.07826	-1.07832	-1.07827	-1.07833	-1.07826	-1.07828	-1.07826	-1.07822
2	-1.07828	-1.07827	-1.07827	-1.07832	-1.07824	-1.0783	-1.07824	-1.07821
3	-1.07822	-1.07827	-1.07827	-1.07832	-1.07826	-1.07828	-1.07826	-1.07822
4	-1.07823	-1.07827	-1.07826	-1.07831	-1.07828	-1.07832	-1.07826	-1.07823
5	-1.07828	-1.07826	-1.07823	-1.0783	-1.07828	-1.07833	-1.07826	-1.07822
6	-1.07826	-1.07826	-1.07824	-1.07829	-1.07827	-1.07832	-1.07827	-1.07823
7	-1.07833	-1.07826	-1.07825	-1.07829	-1.07826	-1.07828	-1.07827	-1.07823
8	-1.07824	-1.07824	-1.07823	-1.07828	-1.07828	-1.0783	-1.07826	-1.07822
9	-1.07822	-1.07825	-1.07825	-1.07828	-1.0783	-1.07828	-1.07827	-1.07821
10	-1.07830	-1.07825	-1.07824	-1.07827	-1.0783	-1.07832	-1.07825	-1.07821
11	-1.07822	-1.07825	-1.07824	-1.07826	-1.07829	-1.07833	-1.07827	-1.07823
12	-1.07822	-1.07824	-1.07829	-1.07826	-1.07828	-1.07834	-1.07825	-1.07821
13	-1.07828	-1.07824	-1.07828	-1.07826	-1.07831	-1.07832	-1.07827	-1.07823
14	-1.07825	-1.07823	-1.07826	-1.07825	-1.07828	-1.07834	-1.07825	-1.07822
15	-1.07829	-1.07823	-1.07827	-1.07825	-1.07829	-1.07833	-1.07826	-1.07822
Ending Zero	-0.00008	-0.00008	-0.00006	-0.00006	-0.00006	-0.00004	0	-0.000014
Range	0.00011	9E-05	6E-05	8E-05	7E-05	6E-05	3E-05	2E-05
Std. Dev.	3.4198E-05	2.2297E-05	1.83874E-05	2.669E-05	1.85E-05	2.26E-05	9.26E-06	7.99E-06
Average	-1.07825867	-1.078256	-1.078256667	-1.0782847	-1.078279	-1.078311	-1.07826	-1.078221
X-Double Bar	-1.07826583	-1.0782587	-1.078258667	-1.0782587	-1.078259	-1.078259	-1.078259	-1.078259
UCL	-1.07820362	-1.0782036	-1.078203623	-1.0782036	-1.078204	-1.078204	-1.078204	-1.078204
LCL	-1.07832804	-1.078328	-1.078328043	-1.078328	-1.078328	-1.078328	-1.078328	-1.078328
Z-Score	0.047201816		0.621147887		0.794795305		2.280684097	
Note: Anything beyond spe								
EN Ratio	0.042865463		0.450087363		0.525101923		0.632265581	

SPC DATA GRAPHS



Common Issues with Laboratories Performing Measurements

1. CMC (Measurement Uncertainty) values that are unrealistic.
2. Lack of understanding of the standards.
3. Not properly evaluating Measurement Risk or Probability of False Accept (PFA).
4. The lab does not replicate how the instruments are used by using the right adapters.
5. Passing too much risk to you!

Decision Rules Conclusion

- Calculating Measurement Uncertainty correctly is essential to everything that comes after it including decision rules.
- Metrological Traceability relies on a documented unbroken chain of contributions, each contributing to the measurement uncertainty, linking them to an appropriate reference.
- A decision rule should take into account the measurement uncertainty.
- Using the manufacturer's accuracy specification and not correcting for bias can further increase Measurement Risk.

Sample Quiz (Open Discussion)

Was the NFL right to go after Tom B? Explain your Answer

It is quite possible Tom Brady knew the PSI per each football was low, however, the NFL clearly lacks the appropriate understanding of Measurement Uncertainty. If the NFL used Guarded Rejection, there would be a % of doubt that the balls were deflated.

What are some reasons to increase Tolerance Limits?

Reliability Targets are not being met
The Measurement Capability Index (TUR) Ratio is too low
The Manufacturer used Averages and the Tolerance Cannot be Achieved

Exercise

1. In two words, define risk.

2. How are measurement risk and product risk related?

3. Describe the amount of risk that could be passed to you if using Simple Acceptance (W = 0, No GB) and the measurement is on the tolerance line assuming a TUR of > 2:1

4. What are the elements that determine the probability of incorrect measurement decisions for:
Specific risk:
Global risk:

Exercise

1. In two words, define risk.
Likelihood
Consequence

2. How are measurement risk and product risk related?
Measurement risk is the risk of incorrect decisions
Product risk is the negative consequence from the measurement-based decision.

3. Describe the amount of risk that could be passed to you if using Simple Acceptance (W = 0, No GB)
When the measurement is on the Tolerance limit both Consumer Risk and Producer Risk is at 50 % (Shared Risk)

4. What are the elements that determines the probability of incorrect measurement decisions for:
Specific risk: amount of uncertainty in measurement process, measurement location relative to limits
Global risk: elements of Specific risk, plus a priori knowledge of product/process

Exercise

6. What is the difference between guarded acceptance and guarded rejection?

7. In your own words, describe why a decision rule is considered a business decision.

Exercise

6. What is the difference between guarded acceptance and guarded rejection?

Guarded acceptance decreases the acceptance zone and guarded rejection increases the acceptance zone.

7. In your own words, describe why a decision rule is considered a business decision.

A decision rule establishes where the accept/reject criteria lie with respect to the specification limits, thus setting the consumer and producer risk. The risk criteria should be established based on the consequences of a bad decision, whether it be safety, economic, or otherwise.

Takeaways

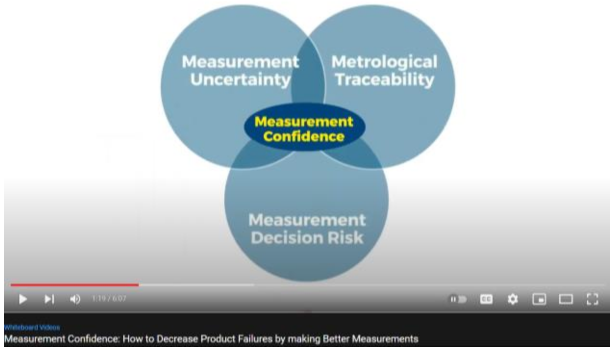
Recommended Reading - Guidance

- **ILAC G8:09/2019 *Guidelines on Decision Rules and Statements of Conformity***
- **JCGM 106:2012 *Evaluation of measurement data – The role of measurement uncertainty in conformity assessment***
- **UKAS LAB 48: *Decision Rules and Statements of Conformity***
- ISO/IEC 17025 2017 *General requirements for the competence of testing and calibration laboratories*
- Handbook for the Application of ANSI Z540.3-2006: *Requirements for the Calibration of Measuring and Test Equipment*
- The Metrology Handbook 3rd Edition Chapter 30
- NCSLI-RP18 Estimation and Evaluation of Measurement Decision Risk
- ASME B89.7.3.1-2001 Guidelines for Decision Rules: Considering Measurement Uncertainty in Determining Conformance to Specifications
- ASME B89.7.4.1-2005 Measurement Uncertainty and Conformance Testing: Risk Analysis *
- ISO 14253-5 Part 1: Decision rules for proving conformity or nonconformity with specifications
- WADA Technical Document – TD2017DK
- Decision Rules Guidance Document by Henry Z, Dilip S, Greg C and more

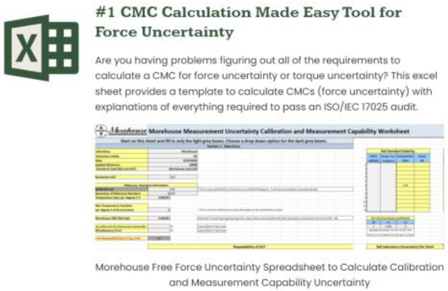
Recommended Reading - Papers

- Evaluation of Guard Banding Methods for Calibration and Product Acceptance – Colin J. Delker
- A STUDY OF AND RECOMMENDATIONS FOR APPLYING THE FALSE ACCEPTANCE RISK SPECIFICATION OF Z540.3 – D. Deaver J. Sompri
- Guard-banding Methods-An Overview – S. Rishi
- A Guard-Band Strategy for Managing False-Accept Risk- M. Dobbert
- **Risk Mitigation Strategies for Compliance Testing** – J. Harben & P. Reese
- Measurement Decision Risk – The Importance of Definitions – S. Mimbs
- Understanding Measurement Risk – M. Dobbert
- Conformance Testing: Measurement Decision Rules – S. Mimbs
- Using Reliability to Meet Z540.3's 2 % Rule – S. Mimbs
- Analytical Metrology SPC Methods for ATE Implementation – *H. Castrup*
- The Force of Decision Rules: Applying Specific and **Global Risk** to Star Wars – H. Zumbrun & G. Cenker
- Unraveling the Tom Brady Deflate Gate – G. Cenker, & H. Zumbrun
- The Definition of a Fool is a Drowning Man Who Tries to Keep It a Secret – G. Cenker, & H, Zumbrun
- Calibration in Regulated Industries: Federal Agency use of ANSI Z540.3 and ISO 17025 – P. Reese

Want More Information?



[Morehouse YouTube Videos]



Morehouse Free Downloads

IndySoft

IndySoft | Calibration Management Software

Thank you

greg.cenker@indysoft.com

hzumbrun@mhforce.com

SCAN ME