



Start With the End in Mind: Working Backwards to Size Your Measurement Equipment and Calibration Provider

A practical, risk-informed method for selecting equipment, decision rules, and calibration providers

Start With the End in Mind

		
Tolerances	Your Standard	The Lab
What do you need to measure?	Aim for at least 4.6:1 TUR.	Must be more accurate than your needs.

Abstract

Begin with the decision you must defend—product conformance with acceptable risk—and work backwards to derive the measurement capability required of both your equipment and your calibration provider. Set explicit risk targets (e.g., unconditional Probability of False Accept (PFA) $\leq 2\%$), choose a decision rule, and translate those targets into required Measurement Capability Index (C_m), also known as Test Uncertainty Ratio (TUR), and/or End-of-Period Reliability (EOPR). From there, pick equipment and a calibration lab that can deliver the necessary uncertainty, traceability, and decision rule

Morehouse Load Cell Reliability – Technical Paper
Author: Henry Zumbrun, Morehouse Instrument Company

implementation. This paper provides the steps, sizing rules of thumb, checklists, and purchase-order language to guide the process.



1. Start with the End in Mind

Ask two questions before any measurement or purchase:

1. How good does the measurement need to be?

Business lens: What happens if we're wrong—safety, regulatory exposure, recalls, warranty, lost uptime, scrap/rework? Which features are critical? What false-accept (consumer) and false-reject (producer) risks are acceptable?

2. How good can the measurement be made?

Technical lens: What can our process + equipment + provider actually deliver? Consider the uncertainty budget (reference standard, method, environment, operator, resolution, stability/drift) and how it behaves across the points where we'll decide conformance.

Why this matters: Answering these two questions up front prevents buying the wrong tool, asking a lab to do the impossible, or making conformity decisions that don't meet your risk tolerance.

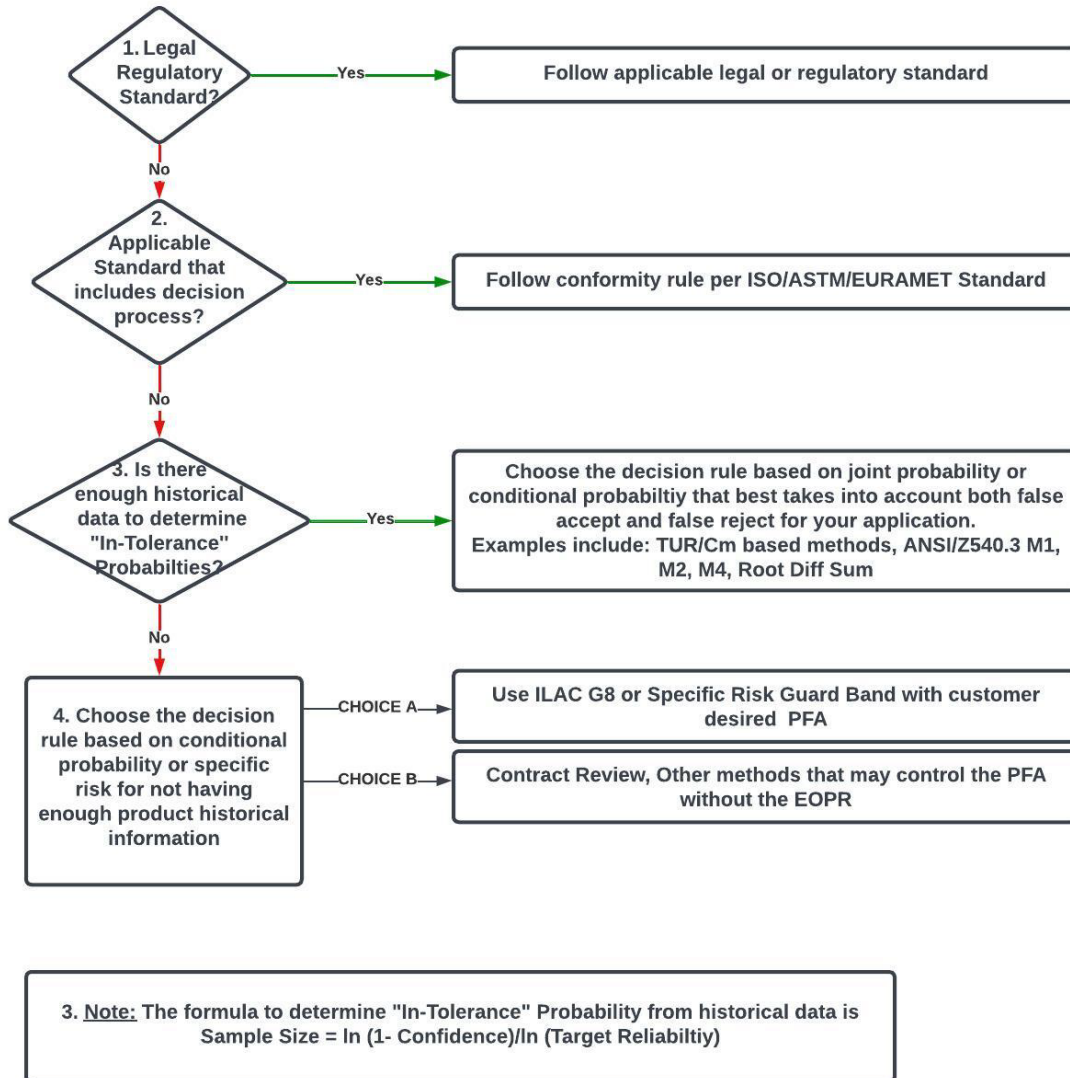
Clarifying these up front prevents costly misalignment later.

Define the end state in operational terms:

- Decision to be made: accept/reject/adjust.
- Target decision risk: e.g., PFA \leq 2 %, producer risk tolerable.
- Decision rule you may use: Subtracting U 95 % or another Specific Risk Decision Rule or a more Global Decision Rule like M6.
- Applicable tolerances & environments. For Example: % of reading, % of full scale, or absolute units.
- Required reporting: uncertainties ($k = 2$, or 3), U95 % or U99 % for what Confidence Interval and how it is calculated.

2. A Backward-Pass Method (from decision risk to capability)

Decision Rule Selection Guide for Calibration and testing Applications Requiring a Statement of Conformance



Step A — Set the risk target and decision rule.

- Depending on the risk level and application, select an unconditional (Global based on averages), or conditional (Bench-Level, Specific) PFA target (e.g., $\leq 2\%$). The decision tree above likely provides the appropriate guidance.

Morehouse Load Cell Reliability – Technical Paper
 Author: Henry Zumbrun, Morehouse Instrument Company

- Choose the appropriate decision rule: risk-limited guardband, subtracting 95 % of the MU, multi-state, Method 6, or other methods.
- State this in your purchase orders and calibration instructions.

Example: Calibrate XXX instrument, subtracting the 95 % Expanded Uncertainty. If PFA < 5 %, PASS; otherwise, FAIL.

Step B — Convert risk target to required capability.

- Rule of thumb: With shared-risk (no guardband), a TUR ≥ 4.6:1 constrains unconditional PFA to ≤ 2 %.

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

- Alternatively, if End-of-Period Reliability (EOPR) ≥ 89 % for like instruments, PFA ≤ 2 % regardless of measurement uncertainty.
- If neither condition holds, apply guardbands sized to the chosen risk limit and recompute required TUR.

Step C — Translate capability to equipment and provider requirements.

- Equipment: ensure the achievable expanded uncertainty U_{exp} satisfies $U_{exp} \leq \text{Tolerance} / 4.6$ when operating without guardbands at PFA 2 %.
- Provider: require a lab that can demonstrate CMCs, methods, and decision rules consistent with your targets and provide traceability to SI.

Step D — Validate with reliability data and control charts.

- Trend in-tolerance reliability by family of instruments; adjust intervals and guardbands accordingly.
- Maintain feedback loops when any out-of-tolerance condition is found.

3. Sizing Rules & Quick Math

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



Device Tolerances You Need to Calibrate (T)	Required U (k = 2)	Resulting TUR/C _m	Global Risk Requirements
0.01 %	0.0022 %	4.6:1	PFA ≤ 2 %
0.05 %	0.0109 %	4.6:1	PFA ≤ 2 %
0.10 %	0.0217 %	4.6:1	PFA ≤ 2 %
0.50 %	0.1087 %	4.6:1	PFA ≤ 2 %

Example: For a tolerance of 0.10 % and shared-risk decisions, targeting PFA ≤ 2 % implies $U_{exp} \leq 0.10 \% / 4.6 \approx 0.0217 \%$. If your current process delivers $U_{exp} = 0.033 \%$, you either (a) add guardbanding, (b) improve capability (equipment, method, environment), or (c) increase reliability (EOPR ≥ 89 %) and justify against the PFA model.

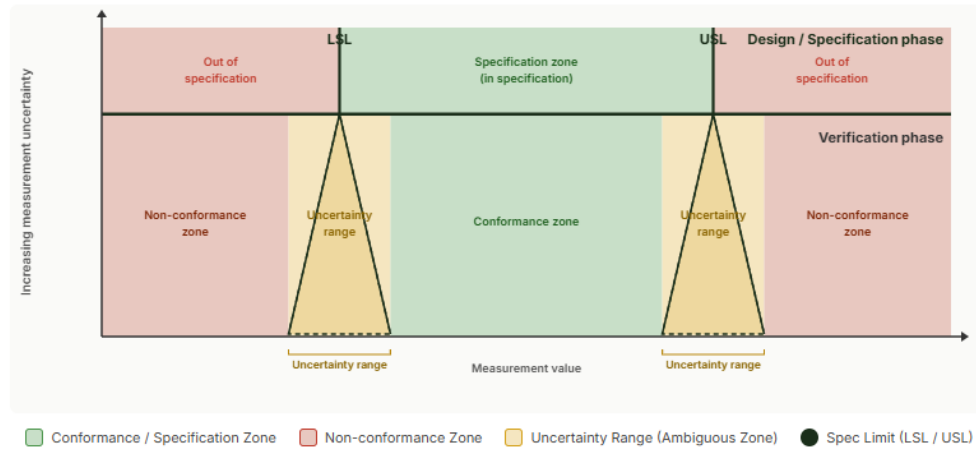
Note: You do not need to use 4.6:1, and in some cases, there will not be a standard that is 4.6 times better than what you wish to calibrate.



From Design to Verification: How Uncertainty Shrinks Your Usable Spec

At the design phase, the full specification range is available. During verification, measurement uncertainty creates zones of ambiguity at every tolerance boundary.

SPECIFICATION LIMITS VS. MEASUREMENT UNCERTAINTY ZONES



Every measurement we take comes with some level of doubt. This doubt is called **measurement uncertainty**. It's the range within which the "true value" probably lies.

The larger the TUR/C_m ratio, the more room you will have to issue a "Pass" statement of conformity.

4. Equipment Selection: Uncertainty, Throughput, and Cost of Ownership

When comparing machines, look beyond sticker price to long-term costs: downtime for reference calibration, cycle time, risk of damage in shipping, and rework costs from false rejects. Higher capability reduces decision risk and often the total cost of ownership.

This example compares three Morehouse force calibration machines and their typical Calibration and Measurement Capability (CMC). At 10 000 lbf with a ± 0.1 % tolerance (±10 lbf), a secondary-standard calibration delivers U ≈ 4.03 lbf (TUR 2.5:1, PFA ≈ 16 % at the boundary) while Morehouse primary deadweight standards deliver U ≈ 0.41 lbf (TUR 24:1, PFA ≈ 0 %). The cost example below uses a hypothetical cost of \$50 per half hour and assumes 1 000 instruments are calibrated per year.

Option	Approx. TUR/C _m	Decision Correctness / PFR	Rework Cost (1,000/year @ \$50/0.5hr)

Morehouse Load Cell Reliability – Technical Paper
Author: Henry Zumbrun, Morehouse Instrument Company



Morehouse BCM	2:1	~90–92 % / ~9–10 %	\$4,500–\$5,000
Morehouse UCM	3:1	~95–96 % / ~4–5 %	\$2,000–\$2,500
Morehouse Deadweight	50:1	~100 % / ~0 %	\$0

Practical implication: paying more up front for greater capability reduces rework and downtime enough to break even over time. At TUR 2.5:1, a guard band sized for 2.5 % specific risk consumes nearly 40 % of the tolerance, leaving a usable acceptance zone of only ± 6.05 lbf out of ± 10 lbf. At TUR 24:1 (Morehouse deadweight), the guard band shrinks to 0.40 lbf, preserving 96 % of the specification range. Higher TUR does not just reduce risk — it recovers usable tolerance that lower-capability machines sacrifice to guard bands.

5. Calibration Provider Capability



Select providers that:

- Demonstrate traceability via a documented, unbroken chain to the SI, with uncertainties at each tier.
- Publish CMCs appropriate to your needs and that can achieve the U_{exp} required by your target TUR.
- State the decision rule used and can apply guardbands sized to your risk target.
- Provide EOPR trends or incorporate your reliability data to justify risk compliance.
- Support feedback analysis when their standards are found out-of-tolerance.

Morehouse Instrument Company meets all five criteria. With primary deadweight standards to 120 000 lbf, dual accreditation (NVLAP Lab Code 600259 and A2LA Cert. 1398.01), and CMCs 10–50 \times lower than typical secondary-standard labs, Morehouse delivers the measurement capability needed to keep TUR

Morehouse Load Cell Reliability – Technical Paper
Author: Henry Zumbrun, Morehouse Instrument Company

high and decision risk low. For detailed guidance on selecting and documenting decision rules, see the companion document: Decision Rules & Statements of Conformity — A Simplified Guidance Document for Calibration Laboratories.

6. Lifecycle & Reliability Considerations

Risk exists across the lifecycle—research, design, test, and product assurance. Control measurement risk early to prevent expensive downstream consequences. Use reliability trending to set intervals and keep decision risk controlled without over-calibrating.

7. Implementation Checklist (Purchase Order / SOW Language)

- 1) Decision Rule & Risk: “Calibrate using a decision rule that constrains unconditional PFA to $\leq 2\%$. If shared-risk, ensure $TUR \geq 4.6$; otherwise apply guard bands sized to the target risk.”
- 2) Uncertainty & TUR: “Report expanded uncertainty ($k = 2$). Required $U_{exp} \leq T/4.6$ unless guard bands are used.”
- 3) Traceability: “Provide metrological traceability statement showing an unbroken chain to the SI with uncertainties at each stage.”
- 4) Reliability: “Provide or accept EOPR data; adjust intervals to maintain target risk.”
- 5) Feedback: “Notify and perform impact analysis if any reference standard is found out-of-tolerance.”

Key Takeaways

- “ISO/IEC 17025 requires clear decision rules.”
- “Shared risk can mean up to 50 % false accepts.”
- “Aim for $\geq 4.6:1$ TUR when using Method 6 (global risk). At this threshold, guard bands drop to zero and global PFA remains below 2 %. For specific risk, the benefit of higher TUR is wider acceptance limits.”
- “High Measurement Uncertainty in Relationship to the Tolerance = High Risk!”

Method 6 — Dobbert Managed Guard Band

In late 2006, ANSI/NCSL Z540.3 added a Global consumer risk requirement of 2 % or less for calibrations requiring a conformance decision.

- Many calibration labs did not have the reliability data (EOPR) needed for **Global Risk** models.

In response, Michael Dobbert developed a **Managed Risk Guardband** that does not require EOPR.

Paul Reese and John Harben wrote a paper titled Risk Mitigation Strategies for Compliance Testing, which proved that for the worst-case EOPR, if the TUR was 4.6:1 or greater, the false accept risk would always be below 2 %. The formula is shown in Appendix B.

References (source summaries)

- Measurement Quality Assurance for End-Item Users (NASA): Begin-with-the-end approach; traceability and uncertainty fundamentals; two key questions on ‘how good does it need to be’ and ‘how good can it be made.’
- NASA HDBK 8739.19-4 Annex 4: Definitions and methods for PFA/FRR (UFAR/CFAR), bench-level vs process-level risk, guard bands, and compensating measures.
- Using Reliability to Meet Z540.3’s 2 % Rule: Engineering basis that $TUR \geq 4.6$ or $EOPR \geq 89\%$ are sufficient for $PFA \leq 2\%$.
- NASA Risk Management Handbook: RIDM/CRM framework—set risk tolerance, make risk-informed choices, then manage residual risk.
- Metrology and the Consequences of Bad Measurements & Metrology’s Influence Across the Product Lifecycle: Direct/indirect risk, lifecycle impacts, and the importance of integrating metrology early.

Appendix A — Bench-Level vs. Global Risk Further Explained

NASA HDBK 8739.19-4 Annex 4 distinguishes between two levels of risk assessment:

- Bench-level (conditional) risk: the false accept/reject risk calculated for a single measurement decision at the bench.
- Global (unconditional) risk: the false accept risk across a population or fleet of instruments. This is the basis for Z540.3’s 2 % requirement.

Bench-level view (conditional risk, specific risk)

Imagine you are calibrating a **single instrument** on a bench. Based on the tolerance and measurement uncertainty, your analysis shows a **Probability of False Accept (PFA) = 4 %**. That means, if you only looked at this one decision in isolation, there’s a 4 % chance you would wrongly accept an out-of-tolerance item.

This is called **conditional risk** because it conditions on this one calibration event and the measurement result. Annex 4 of NASA HDBK 8739.19-4 often labels this **CFAR** (conditional false accept risk).

Fleet-level view (unconditional risk, global risk)

Morehouse Load Cell Reliability – Technical Paper
Author: Henry Zumbrun, Morehouse Instrument Company



Now zoom out to the **whole fleet** of similar instruments. Say you track them with a reliability analysis and find that **89 % of instruments remain in tolerance until the end of their calibration cycle**. That percentage is the **End-of-Period Reliability (EOPR)**.

This fleet-level reliability directly limits how often an out-of-tolerance instrument even arrives at the bench. In statistical terms, it bounds the **unconditional probability of false accept (UFAR)** across the fleet. NASA and Z540.3 analyses show that when **EOPR ≥ 89 %**, the UFAR is guaranteed to be **≤ 2 %**, even if some individual bench-level calculations show higher conditional risks.

Why the numbers differ

- **At the bench:** You only see that specific test, so the calculated PFA can be higher (e.g. 4 %).
- **Across the fleet:** Because most instruments are still good at the end of the interval (≥89 %), the average chance of a false accept across all instruments is much lower. That’s why the **unconditional risk drops to ≤2 % overall**.

Appendix B — Guard banding Methods and Example

Guard bands shrink the acceptance region to cap risk. Annex 4 describes several methods; Method 6 is commonly applied:

$$A = T - U95 \% \times M$$

- **A = acceptance limit**
- **T = tolerance limit**
- **U95 % = calibration process 95 % expanded uncertainty**
- **M = multiplier: the fraction of the 95 % expanded uncertainty for which the acceptance limits provide the desired false-accept risk.**

$$M_{2 \%} = 1.04 - e^{(0.38 \ln(C_m) - 0.54)}$$

For example, with T = 0.10 %, Cm = 2.5, and U95 = 0.033 %, Method 6 produces M ≈ 0.74, so the guardband reduces the acceptance region by 0.024 %. This ensures unconditional PFA ≤ 2 %.

Labs must report when guard bands are applied, and show guard banded acceptance limits on the calibration certificate.

Appendix C — Dual Dominance Conditions

Two conditions are sufficient to guarantee unconditional $PFA \leq 2\%$:

- Bench-level dominance: $TUR \geq 4.6$ (uncertainty small enough).
- Global dominance: $EOPR \geq 89\%$ (fleet reliability high enough).

Either condition alone suffices. In practice, organizations often aim to satisfy both for extra assurance.

Appendix D — Reliability Trending Example

End-of-Period Reliability (EOPR) is the proportion of instruments remaining in tolerance at the end of the calibration interval.

Example: A lab finds 85 % of torque wrenches in-tolerance at 12 months. This yields $EOPR = 85\%$, which is below the 89 % threshold. By shortening the interval to 9 months, trending shows 91 % in-tolerance, restoring compliance. This demonstrates how interval adjustment controls risk when $TUR \geq 4.6$ cannot be achieved.

Appendix E — Linking to RIDM and CRM

The NASA Risk Management Handbook describes two complementary disciplines:

- Risk-Informed Decision Making (RIDM): set risk tolerance, define alternatives, and make measurement system choices consistent with mission objectives.
- Continuous Risk Management (CRM): monitor reliability data, adjust calibration intervals, and apply guard bands to manage residual risk.

Backward-sizing is an example of RIDM, while reliability trending and interval control are CRM practices.

Appendix F — Lifecycle and Cost Impacts

Here's what that looks like in practice:

TUR	Correct Decision	PFR (False Reject)	Avg. Cost per Decision @ \$50 ½ hr.
2:1	~90–92 %	~9–10 %	\$4.50–\$5.00
3:1	~95–96 %	~4–5 %	\$2.00–\$2.50

Morehouse Load Cell Reliability – Technical Paper
Author: Henry Zumbrun, Morehouse Instrument Company



TUR	Correct Decision	PFR (False Reject)	Avg. Cost per Decision @ \$50 ½ hr.
50:1	~100 %	≈0 %	≈ \$0.00

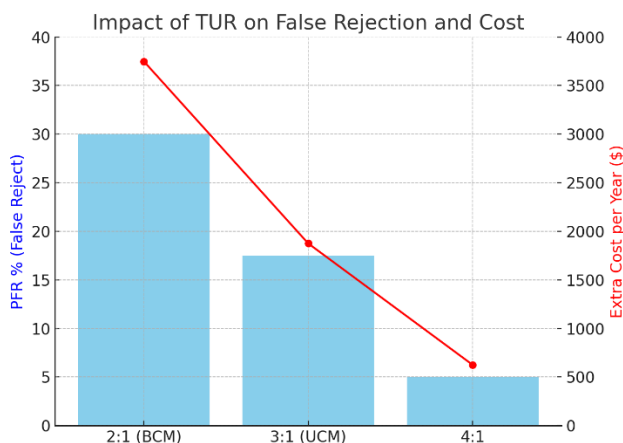
If you do 1,000 calibrations a year:

- The **Morehouse BCM (2:1)** costs you about **\$5,000 annually** in hidden rework.
- The **Morehouse UCM (3:1)** cuts that to about **\$2,500 annually**.
- The **Morehouse Deadweight (50:1)** avoids that cost entirely.

So, if a Morehouse UCM costs \$15,000 more than a Morehouse BCM, rework savings alone cover that difference in ~6 years.

And if your devices require tighter tolerances (say 0.05 %):

- At **1:1 TUR**, hidden rework costs jump to **\$17.50 per decision**.
- At **1.5:1 TUR**, that drops to ~\$8.
- At **25:1 TUR**, it's essentially \$0.



Metrology choices impact lifecycle costs with Morehouse Machines:

- A Morehouse BCM (TUR ~2:1) may appear cheaper, but leads to higher false reject rates (~9–10 %), costing ~\$5 000 per 1 000 calibrations in rework. At this TUR, guard bands consume ~40 % of the tolerance, severely limiting the usable acceptance zone.
- A Morehouse UCM (TUR ~3:1) cuts false rejects in half (~4–5 %), halving rework cost. Guard bands shrink, recovering more usable tolerance.

Morehouse Load Cell Reliability – Technical Paper
Author: Henry Zumbrun, Morehouse Instrument Company



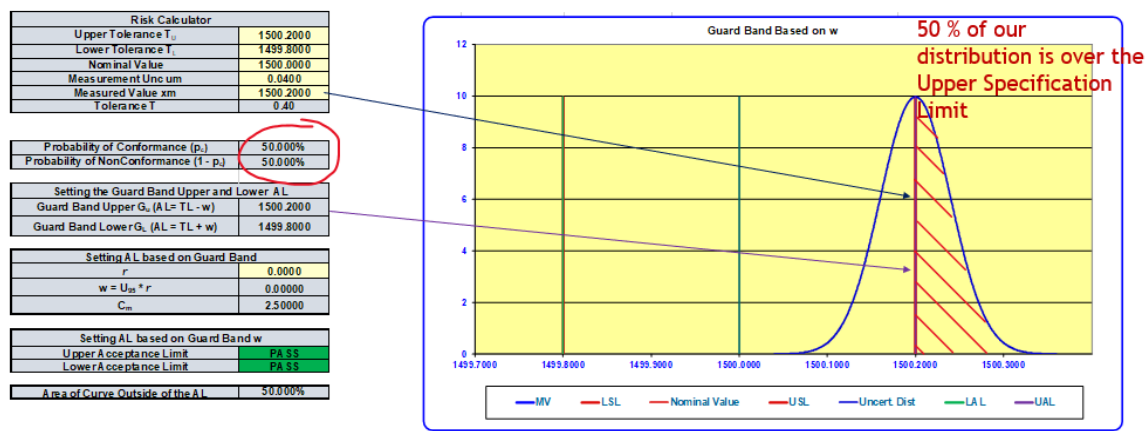
- Morehouse Deadweight standards (TUR ~50:1) virtually eliminate false rejects, preserving 96 %+ of the specification range and saving long-term costs despite a higher upfront price. At 10 000 lbf, the Morehouse deadweight CMC of 0.16 lbf produces $U \approx 0.41$ lbf for the measurement process — compared to $U \approx 4.03$ lbf from a typical secondary-standard supplier (CMC 4.00 lbf).

Appendix G — Shared Risk

Several in the industry often take a "shared risk" approach, and some disguise this by stating a Guardband of $w=0$ has been applied. What that phrase means is that **NO Guardband** has been applied.

What "shared risk" means is as long as the measured value is less than or equal to the tolerance, the value is reported. It is agreed upon as being in conformance, though many are not applying a Guardband ($w=0$) and taking the approach that the end-user can decide if it is *in or out of conformance*.

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



ISO/IEC 17025:2017 Clause 7.1.3 dictates that the customer must agree upon the decision rule when a statement of conformity is to be made. The customer shall specify the acceptance criteria for conformity. The decision rule(s) requirement is unnecessary if no conformance criteria is specified.

Simply put, regarding decision rules, if you choose not to decide, you still have made a choice (Shared Risk).

Under this scenario, the consumer accepts the risk of up to 50 % Probability of False Accept (PFA). If the UUT also carries uncorrected systematic bias, the effective PFA can exceed 50 % because the measurement distribution is shifted further outside the tolerance boundary.

Morehouse Load Cell Reliability – Technical Paper
Author: Henry Zumbrun, Morehouse Instrument Company



Thus, as a customer, you could have a much higher level of Risk than you wanted.

TRUE COST OF MEASUREMENT DECISIONS Executive Summary Report <small>Morehouse Instrument Company — ISO/IEC 17025 Accredited — NVLAP 600259 — AZLA 1398.01</small>								
Measurement Risk Scenario Selected					Specific Risk (PFA ≤ 2%)			
KEY PARAMETERS								
Tolerance	EOPR	Labor Rate	Rework Cost	Cals/Year	Price/Cal	Revenue/yr	Swap Req?	
0.100%	89%	\$200/hr	\$116.67	500	\$300	\$150,000	No	
RECOMMENDATION								
Best Machine:	UCM Automated		Annual Profit:	\$50,584		Payback:	2.0 years	
Worst Available:	Competitor 1		Annual Profit:	-\$3,917		Spread:	\$54,501 /yr	
<small>At 0.100% tolerance, 89% EOPR, 500 cals/yr at \$200/hr. UCM Automated is the best machine at \$50,584/yr profit, 2.0 years payback. Worst available: Competitor 1 at \$3,917/yr loss. Best option: UCM Automated at \$50,584/yr profit, 2.0 years payback. Saves \$54,501/yr vs worst (Competitor 1). Switching from Competitor 1 (\$3,917 loss) to UCM Automated (\$50,584 profit) saves \$54,501/yr, payback in 2.0 years.</small>								
MACHINE COST & PROFITABILITY COMPARISON (Specific Risk (PFA ≤ 2%) Decision Rule) — Applied Force: 2,000 lbf								
Machine	TUR	Labor/yr	Rework/yr	Cal Cost/yr	Swap/yr	Total Cost/yr	Profit/yr	Payback
BCM	2.0	\$90,000	\$19,822	\$2,000	\$0	\$117,870	\$32,130	0.8 yr
PCM	3.0	\$90,000	\$10,865	\$2,000	\$0	\$102,865	\$47,135	0.7 yr
UCM Manual	3.3	\$100,000	\$9,509	\$2,000	\$0	\$111,509	\$38,491	1.3 yr
UCM Automated	5.0	\$90,000	\$5,533	\$2,000	\$0	\$98,416	\$50,584	2.0 yr
Deadweight	50.0	\$120,000	\$531	\$0	\$0	\$120,531	\$29,469	11.9 yr
Competitor 1	2.0	\$100,000	\$19,822	\$2,000	\$0	\$153,917	-\$3,917	Never
Competitor 2	1.7	\$100,000	\$25,857	\$2,000	\$0	\$153,917	-\$3,917	Never

Need help figuring out your true ownership costs? Download our [True Cost Spreadsheet](#).