Guardbanding and the World of ISO Guide 25
Is There Only One Way?

David Deaver
Presenter & Author
Fluke Corporation
PO 9090   M/S 169G
Everett, WA, USA 98206
TEL (425)356-6434 FAX (425)356-5649

ABSTRACT

Revisions of ISO Guide 25, drafts 5 and 6, currently proposed to be released as ISO 17025, are fairly explicit about claiming an in-tolerance condition:

> "When ... parameter(s) are claimed to be within specified tolerances, the measurement value(s), extended by the estimated uncertainty of measurement, shall fall within the appropriate specification limit."

Does this exactly mean that the guardband always equals the UUT specification minus the measurement uncertainty or should a broader interpretation of "extended" be allowed?

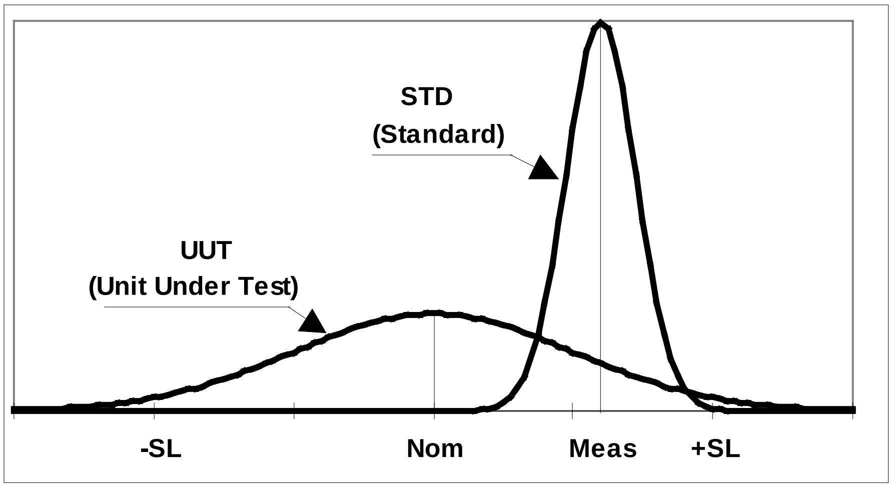
From a practical standpoint, guardbanding by the draft proposal can be unnecessarily expensive and counter productive. Can other guardbanding philosophies co-exist in the world of Guide 25 and accreditation?

INTRODUCTION

The author has struggled with this issue for the past 5 years. This paper reviews the concept of guardbanding and various strategies used. It then compares some of these strategies with ISO Guide 25 and describes the perspective of each approach. Why does Guide 25 propose what it does? How and when should it be used? Likewise, when are some other guardbanding approaches applicable? Finally, the paper will suggest a compromise alternative which may help to resolve this issue.

THE MEASUREMENT PROBLEM

Let us examine the typical measurement problem shown in Figure 1 below; the determination of the true value of the output of a unit under test (UUT) when a given (Nom) output is selected.



UUT: The distribution of possible values for the unit under test when "Nom" is requested.
Nom: The nominal (selected) value for the UUT
-SL: Lower specification limit for the UUT
+SL: Upper specification limit for the UUT
STD: The distribution of possible true values of the UUT output which will produce the "Meas" value from the Standard.
Meas: The value reported by the STD for the actual value of the UUT

Figure 1: Nomenclature

THE STANDARDS LAB PERSPECTIVE, the narrowest view of the problem

National laboratories and standards laboratories are generally called upon to provide an estimate of the true value of a parameter for a given UUT. This is given as a measured value and an uncertainty. National labs, standards bodies, and the academic community have published guidelines on the handling of measurement results and the calculation of their associated uncertainties. Three such guides are listed in the bibliography [1-3]. The ones cited here are quite detailed and are each 60-100 pages in length.

A standards lab solution to our measurement problem, then, would be to report the value, "Meas", and the uncertainty of the measurement; usually a single number which would represent the dispersion of the probability density function of the STD. "Unless otherwise indicated, one may assume that a **normal distribution** was used to calculate the quoted uncertainty, ... " [2].

A standards lab will typically be reluctant to make a declaration as to whether the UUT is in tolerance and equally reluctant to make adjustments to the UUT to make the "Nom" UUT output equal to the "Meas" value of the STD. Therefore, most standards labs do not have much comment regarding this area of the proposed ISO Guide 25 changes.

THE USER'S PERSPECTIVE, the fuzziest view of the problem

The user of a piece of test equipment which has been sent for calibration wants a very simple question answered. "Is my UUT good?" Often, the unit is returned without a very simple answer. From a standards lab and many calibration labs, a set of measurement results and uncertainties of measurement would be returned to the user who would have to determine the suitability of the instrument for his application. The user probably has made a determination sometime in the past that, if the UUT meets the manufacturer's specifications, it is suitable for the application. Therefore the user will rephrase the question, "Is my UUT within manufacturer's specified tolerances?" The reply to that question is often "Most likely", "Probably", or "None of the measured values were outside manufacturer's specs."

Why such vague answers? Referring again to Figure 1, we can see that for measured values near the "Nom" value, we can say with a very high probability that the true value of the UUT lies within the specification limits. However, for "Meas" values near the specification limits, no such statement can be made because of the uncertainty in the STD. In fact a "Meas" value just barely inside the specification limit has nearly as much probability of being out of tolerance as being in tolerance. Here is where the changes to ISO Guide 25 are being made to try to help the situation. The claim is that there is a region close to the specification limits where a measurement cannot be made with enough precision to determine with any degree of confidence whether the true value is inside or outside the specification.

The current drafts being considered for ISO Guide 25 specify that for a parameter to be declared "In Tolerance", the measured value has to be inside the specified limit by the measurement uncertainty of the standard. Though not explicitly stated, the implication would be that a value could not be declared to be out of tolerance unless the measured value exceeded the specification limit by a like amount.

Figure 2 illustrates the indeterminate regions. If the STD has small uncertainty (high TUR), these regions will be quite small. However, it the TUR is low, the indeterminate regions can grow quite large. (TUR, test uncertainty ratio is the ratio of the uncertainty of the unit under test to the uncertainty of the standard.)

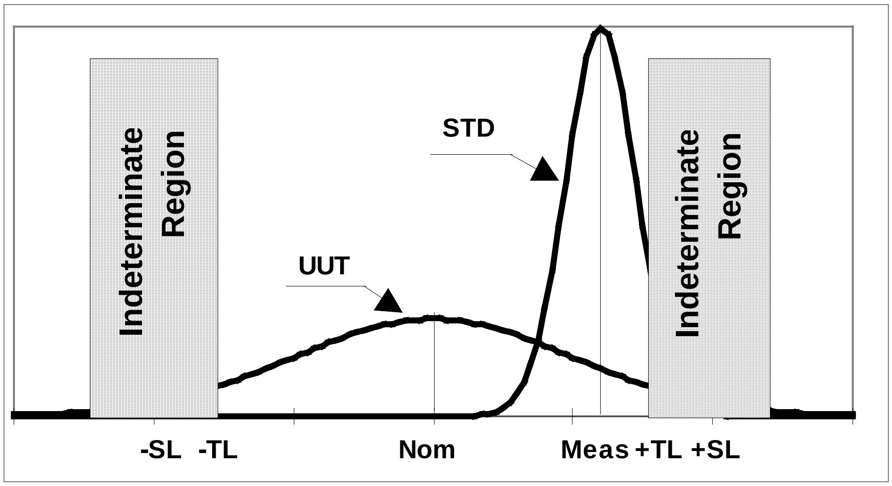
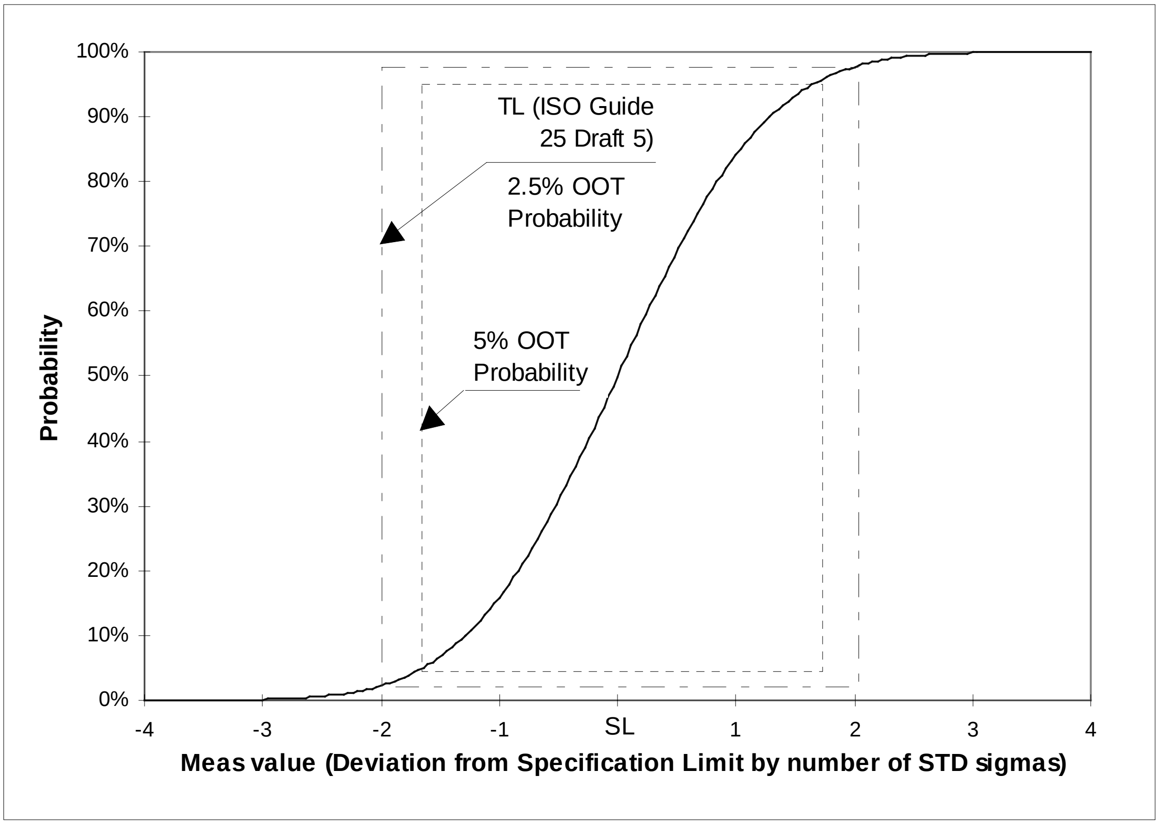


Figure 2 Indeterminate Regions (Guardbands)

These regions can be referred to as guardbands, setting test limits (±TL) different than the specification limits (±SL). ISO Guide 25 seems to be clarifying the picture a bit for the user by making sure that his UUT will be verified to test limits that will ensure to a high confidence that the UUT parameters are all within specified tolerances.

Figure 3 shows the guardbands in more detail in plots of the probability that measured values came from true values which exceed the specification limit of the UUT as a function of where the measured value. The horizontal axis is scaled in units of number of standard deviations **of the STD** that the measured value deviates from the specification limit (SL). The two plots are identical except one shows the probability with a linear vertical scale and the second with a log vertical scale. As one would expect, the probability of the true value being out of tolerance (OOT) when the measured value equal to the specification limit, is 50%. The ISO Guide 25, Draft 5 proposed guardband is shown which sets the test limit (TL) at the SL minus the uncertainty of the STD. If the STD is specified at the 2σ confidence level, this results in a 2.5% probability that the true value exceeds the SL when the measured value is equal to the TL.

It has been proposed [4] that a more appropriate setting of the guardband for the intent of ISO Guide 25 Draft 5 would be to set it at the 5% probability of an OOT true value producing a measured value at the TL. This would set the TL inside the SL by 1.6448σ or 82% of the 95% confidence specification of the STD. Before embracing this proposal, however, we should look at some other perspectives.



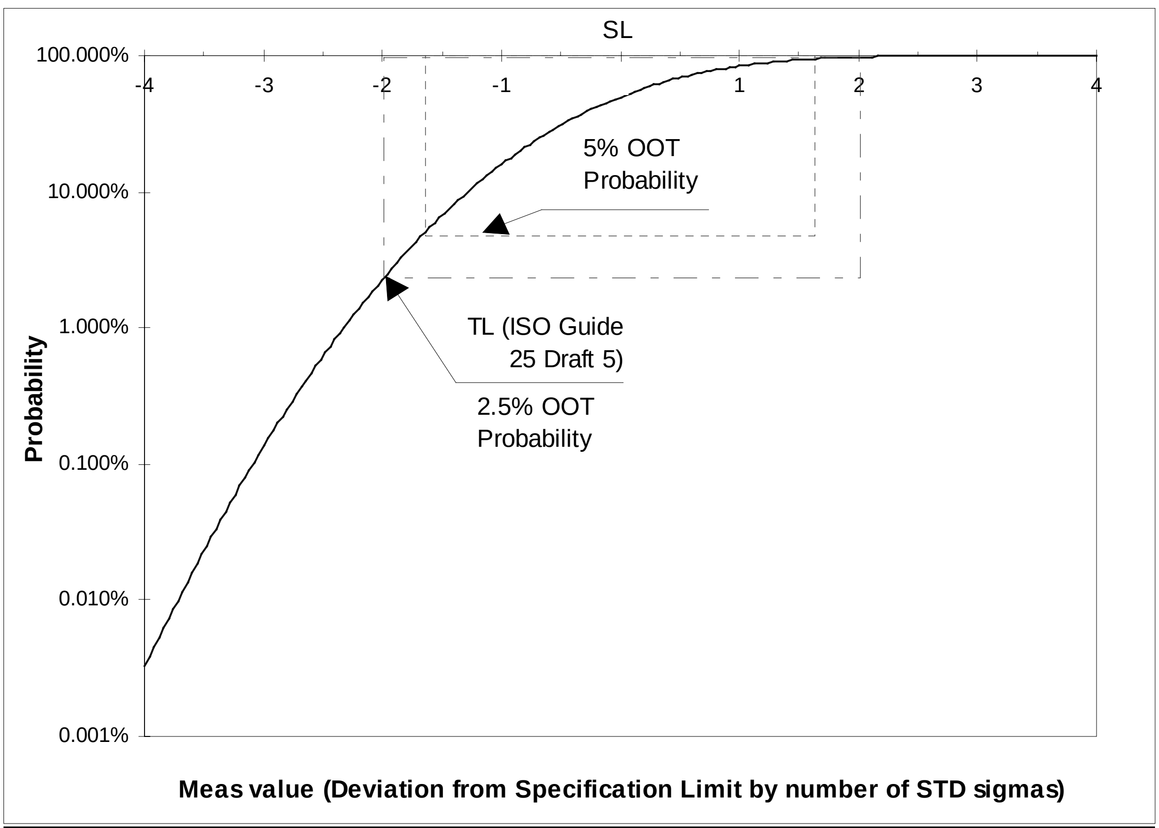


Figure 3 Probability that a Meas Value came from an OOT True Value

THE MANUFACTURER'S PERSPECTIVE, the broadest view of the problem

A manufacturer has a considerably different view of this measurement problem. To be successful, he must manage yield and out of tolerance risk. First, he must be able to consistently produce product which falls within the specification limits and secondly, eliminate (discard or rework) any product which falls outside the specification limits. The problem can be best understood and attacked by using statistical and process tools.

Statistical quality control has taught us that the highest emphasis for the manufacturer should be to reduce the variation in his product, the UUT. All the other aspects of the problem become much easier if the production processes are under control. This is, of course, a multi-disciplinary task; the specifications must be set realistically and with respect to workload, the product must be designed for minimum sensitivity to component variations and environmental stresses for it's intended life, and it must be fabricated and tested using well controlled process.

Testing of the product is used both to eliminate defective products and to verify the manufacturing process. Other measures of the quality of the process are in-tolerance "as found" rates, warranty failures, mean time between failure (MTBF) and environmental audits.

This manufacturer's perspective has resulted in the most passionate responses to the proposed changes to portion of ISO Guide 25 quoted in the abstract. Even if a manufacturer does have his processes well under control, the measurement requirements in the standards can cause substantial yield reductions.

Consider the following illustration using calculations from [5]. A single parameter product has it's specifications set at 2 sigma, about 95% confidence level. Therefore, nearly 5% of the products can be out of tolerance. Assume the STD used to test product has a fourth of the uncertainty of the UUT (TUR of 4) and we accept parts that measure in tolerance and reject those that measure greater than the specifications. 16% of the bad parts will be accepted and 84% of the out of tolerance parts will be rejected, reducing the number of bad parts to 0.8%. However, we will also reject 1.5% of the good parts. So far, in this example, we let the test limits equal the specification limits. Now, if we test to the guardband recommended by ISO Guide 25 Draft 5, 75% of the specification limit in this case, the false accepts are a mere 0.02%. But what a price we will pay to achieve that low rate! We could expect to reject 10% of our good products. In [5-6], equations, graphs, and tables are provided to assess the balance between false accepts and false rejects. This balance can be adjusted by judicious placement of guardbands. The guardbanding recommended in this revision of Draft 5 may be appropriate for the aerospace industry where the additional cost of testing (rejecting good product) is offset by higher mission reliability, but for many other situations, the dramatic cost increase resulting from rejecting many more good products is hardly justified.

The manufacturer would like to be held accountable for distributions, confidence levels, and uncertainties without the imposition of specific testing methods which may dramatically affect yields.

The manufacturer may be able to demonstrate that very few defective products make it into a customer's hands, but the customer would still like to be able to determine with some assurance that the product in his hands isn't one of them.

THE CAL LAB'S PERSPECTIVE, a confused view of the problem

We have made tremendous strides in standardizing how uncertainties are stated, what should be included in calibration reports, and many other aspects of the calibration lab. Clearly, however, calibration labs are confused about how they are to test and report results and compliance to specifications, especially when faced with low TURs. There are many historical methods such as MIL-45662A, no longer an active US military standard but which still lives on in hundreds of standards labs throughout the country. There are new standards which many labs have adopted. There are draft versions of standards not yet adopted which are already being implemented in some labs. And there are methods in the literature which are not necessarily recognized by any standard which are being adopted by labs. There is definitely a need to have a standard which can allow a determination of suitability or compliance to be met for a UUT when tested with a standard.

The cal lab's perspective is different from the manufacturer's perspective. The manufacturer may say, "There is a relatively small probability of getting measured values very, very close to the specification limit." The calibration lab's problem, however, is one of conditional probability, "I don't get very many readings close to the specification limit, but when I do, I don't know if I can claim the unit is in tolerance."

THE HISTORICAL PERSPECTIVE, an overview of the problem

This is not a new issue. It has been a struggle for decades [7-12]. Here is review of some of the methods in use today, most of which are analyzed in more detail in [6].

- Report measured values and the UUT and let the user make the determination; the standards lab view.
- Use TL=SL when the TURs are sufficiently high. Historically, for electrical work and still for mechanical, 10:1 TURs are recommended. MIL-45662A recommended 4:1 and ISO 100012 allowed that TURs as low as 3:1 might be tolerated. Lower than these, the uncertainties are to be stated and the user has to make a determination but only on the low TUR parameters. The argument for this approach is that, if the indeterminate zone is relatively small, even if the true value lies outside the specification limits, it probably isn't outside by far.

- $TL = (1 - \frac{1}{TUR}) \times SL$ ISO Guide 25 Draft 5 proposal.
- $TL = (1 - \frac{0.8224}{TUR}) \times SL$ Argues that Draft 5 should have considered only one side of the STD's distribution when setting the TL. [6]
- $TL = (1.25 - \frac{1}{TUR}) \times SL$ NCSL Recommended practice RP-10
- Keep same risk as a 4:1 TUR with lower TURs by guardbanding.
- Keep same risk as a 3:1 TUR with lower TURs by guardbanding.
- RSS method: $TL = (1 - \frac{1}{TUR^2}) \times SL$ Approximates the equal risk methods

A NEW PERSPECTIVE, one compromise view of the problem

The author presents for consideration (hopefully not for additional confusion) another means of looking at the calibration problem shown in Figures 1 & 2 which allows a priori knowledge or assumptions about the UUT to be applied to help reduce the size of the indeterminate zone. Without the application of at least some of these assumptions, the complaint of the producer is that over-aggressively attacking the problem of falsely accepting defective units causes an inordinate number of good units to be rejected.

In Figures 1-3, we considered the probability density function of true values which could produce a given measured value without regard to the distribution of the UUT. However, from these figures we can see, if the distribution of UUT true values are normally distributed and reasonably centered on the nominal value, then it is more likely a measured value came from a true value which is closer to the nominal value than the measured value.

Let's quantify how much taking into account the shape of the distribution of the UUT affects the test limit. Equation 1 is the expression for P(UUT), the probability density function of the true values, x, for the UUT, assuming a normal distribution with the mean centered on the nominal output. The expression is simplified by normalizing to a zero mean and standard deviation of 1.

Equation 1
$$P(UUT) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

The distribution of measured values (m) resulting from true values (x) is given by:

Equation 2
$$P(STD) = \frac{TUR}{\sqrt{2\pi}} e^{-\frac{TUR^2 (m-x)^2}{2}}$$

The probability of measuring a value m is the product of the probability density functions of the UUT and the STD integrated over all values of x.

Equation 3
$$P(m) = \frac{TUR}{2\pi} \int_{-\infty}^{\infty} e^{-\frac{x^2 + TUR^2 (m-x)^2}{2}} dx$$

We would like to know the worst case risk of accepting a defective unit. This occurs for a measured value at the test limit which can also be expressed as a guardband factor times the specification limit (TL=K x SL). To obtain the probability that a measured value of SL came from a true value greater than the specification limit, Equation 3 can be integrated for values of x greater than SL and normalized by dividing by the probability of obtaining a reading of SL.

Equation 4
$$P(x>SL \mid TL) = \frac{\int_{SL}^{\infty} e^{-\frac{x^2 + TUR^2 (K \bullet SL - x)^2}{2}} dx}{\int_{-\infty}^{\infty} e^{-\frac{x^2 + TUR^2 (K \bullet SL - x)^2}{2}} dx}$$

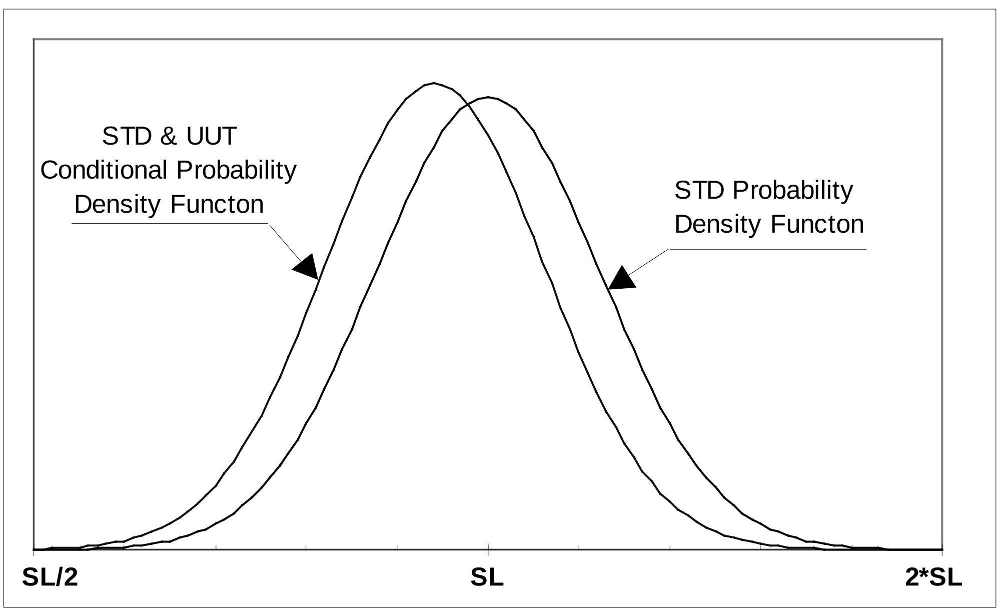


Figure 4 Probability Density Function Shift With UUT Distribution

Figure 4 shows the resulting shift in the mean of the probability density function when the UUT and STD are both normally distributed and the TUR is 4:1. Note that the mean is shifted about 6% and the standard deviation is somewhat smaller. The shift is more pronounced at lower TURs. This shift doesn't seem huge, but it is sufficient to indicate that, if the STD measured right at the specification limit, instead of a 50% chance that the true value is greater than the SL, we could claim only a 31% chance the true value exceeds SL. Figure 5, below, shows the probability a measured value at the test limit, TL=K x SL, came from a true value greater than the SL.

Figure 5, on the next page, shows plots of the probability a measured value comes from a true value greater than the SL for test limits from 50% to 100% of the specification limit. The test limit is assumed to be at 2 sigma of the UUT. The plots show that for lower TURs there is even greater advantage from the assumption that the UUT is normally distributed.

How could these results be simplified to be a practical tool in a working cal lab? From the plots in Figure 5 it can be seen that the guardband factor does not change dramatically with TUR:

TUR	TL (% of SL)
4:1	85%
3:1	82%
2:1	79%
1.5:1	79%

Table 1: Guardband Factors for 5% Worst Case False Accept Rate

A DIFFERENT PERSPECTIVE, a sanity check of the result

If the measured value is equal to the specification limit, is it reasonable to think there is higher probability the true value is in tolerance than out? Consider a rain gauge which consists of a cylinder of negligible mass which collects rain water. Each day, my hypothetical lab must report how much water was in the gauge by weighing it with the uncertainty of measurement. No water spills or evaporates and the only measurement error is the weighing operation whose errors are normally distributed about the true value of the mass being measured and independent of the mass being measured. The capacity of the cylinder is 1 liter which weighs 1 kg. Two days ago, I measured 0.4 kg with a 2σ uncertainty of ±0.01 kg and concluded there is an equal chance the true mass of the rainwater was a little more or a little less than 0.4 kg. Yesterday, however, the cylinder was completely full. I carefully weighed it and got a measured value of 1.005 kg. Can I conclude, after looking at my tables of normal curves that there is an 84% chance there was more than 1 kg of water in the cylinder. Today it didn't rain and I measured 0.001 kg ±0.01 kg!

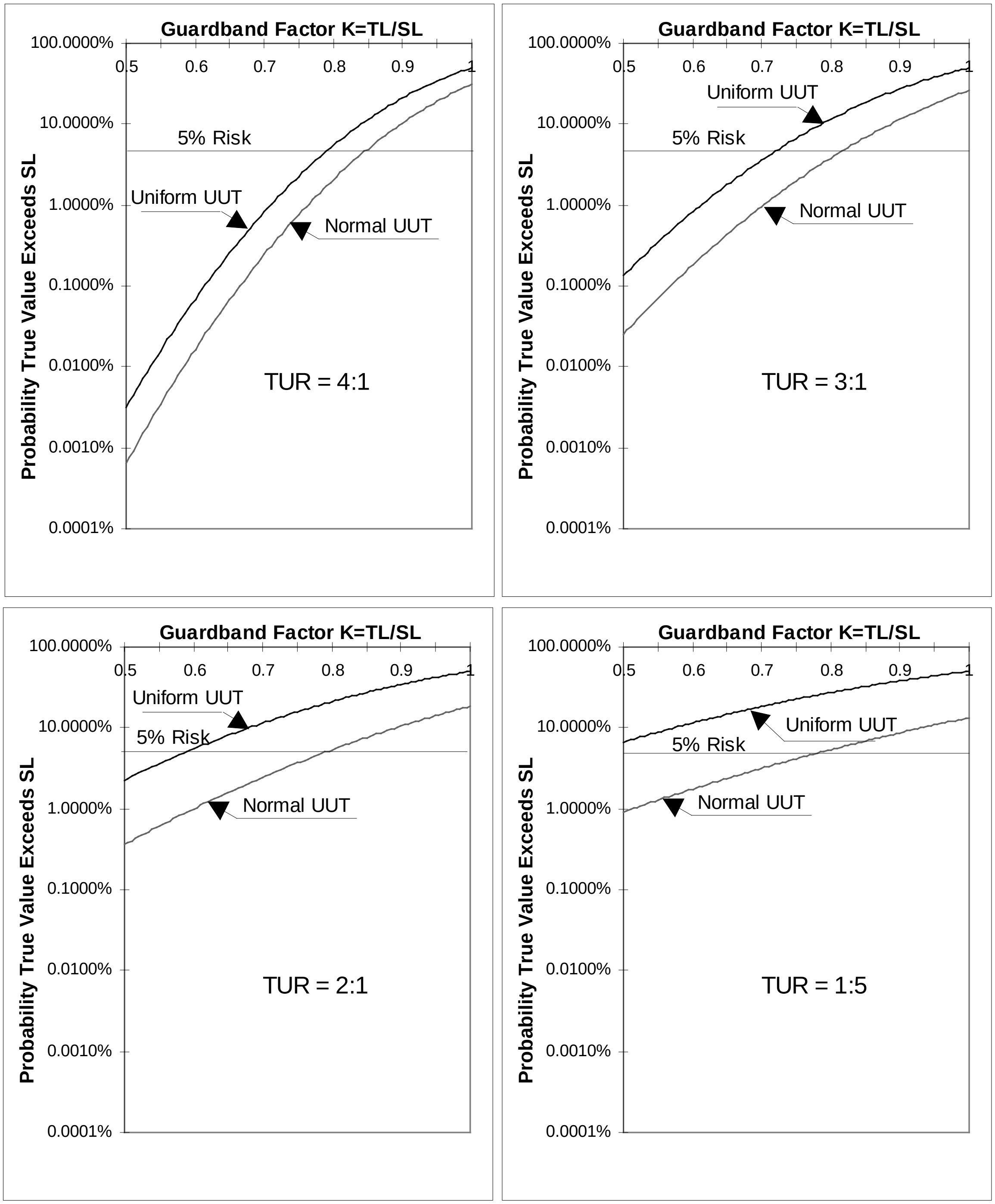


Figure 5 Probability of True Value > SL for K=0.5 to 1 and TUR 1.5:1 to 4:1

CONCLUSION

It is the opinion of the author that the guardbanding approach proposed in Drafts 5 & 6 of ISO Guide 25 overagressively controls the false accept test risk while ignoring the calamitous results of a tremendously increased false reject test risk.

Here is a proposal for consideration as a compromise position for the measurement problem discussed in this paper:

For the measurement of UUTs with STDs that can be determined or presumed to normally distributed with small bias:

- For reasonably high TUR, say greater than 4:1, allow TL=SL and make the declaration that the point was found “not to be out of tolerance” rather than “within specified tolerances”. This takes into account that any parameters which are accepted under this criteria which are actually out of tolerance, are not likely to be very far out of tolerance.

- For TURs between 1.5:1 and 4:1, parameters with measured values which are less than 80% of the specification limits may be declared in tolerance

The author readily admits to being mathematically challenged and encourage comments and criticism of this proposal with the goal of finding agreement on a method of determining compliance to specifications which strikes an equitable balance between false accepts and false rejects.

ACKNOWLEDGMENTS

The author wishes to recognize Ray Kletke of the Fluke Corporation for suggesting the concept of the conditional probability presented in this paper and his encouragement and critiques during its creation.

As in previous papers, MathCAD® was used to do any of the really hard work.

REFERENCES:

[1] ISO/TAG 4/WG 3: June 1992, “Guide to the Expression of Uncertainty in Measurement.

[2] ANSI/NCSL Z540-1-1994, “Calibration Laboratories and Measuring and Test Equipment - General Requirements”, American National Standards Institute, July, 1994.

[3] Singapore Institute of Standards and Industrial Research, “Guidelines on the Evaluation and Expression of Measurement Uncertainty”, July, 1995

[4] Agy, Dave and Wong, Warren, “A Statistical Approach to Calibrate High End Digital Multimeters”, 1997 Measurement Science Conference.

[5] Deaver, David K., “Maintaining Your Confidence (In a World of Declining Test Uncertainty Ratios)”, 1993 NCSL Workshop & Symposium, pp. 133-153

[6] Deaver, David K., “Guardbanding with Confidence”, 1994 NCSL Workshop & Symposium, pp. 383-394

[7] Capell, Frank, "How Good is Your TUR?", Evaluation Engineering, January, 1991, pp. 80-84

[8] Castrup, Howard, “Evaluation of Customer and Manufacturer Risk vs. Acceptance Test Instrument In-Tolerance Level”, NCSL Workshop & Symposium, 1980.

[9] Eagle, Alan R., “A Method for Handling Errors in Testing and Measuring”, Industrial Quality Control, March, 1954, pp. 10-15

[10] Ferling, John A., “The Role of Accuracy Ratios in Test and Measurement Processes”, Measurement Science Conference, 1984

[11] Grubbs, Frank E. and Coons, Helen J., “On setting Test Limits Relative to Specification Limits”, Industrial Quality Control, March, 1954, pp. 15-20

[12] Hayes, Jerry L., “Calibration and Maintenance of Test and Measuring Equipment”, Encyclopedia of Applied Physics, Vol 3, VCH Publishers, Inc., 1-56081-062-9/92, 1992

[13] Hutchinson, Bill, “Setting Guardband Test Limits to Satisfy MIL-STD-45662A Requirements”, 1991 NCSL Workshop & Symposium, pp. 305-309

[14] Schumacher, Rolf B. F., “Decision Making and Customizing Decision Limits in Measurement Assurance”, 1992 NCSL Workshop & Symposium, pp. 301-310