A STUDY OF AND RECOMMENTATIONS FOR APPLYING THE FALSE ACCEPTANCE RISK SPECIFICATION OF Z540.3

David Deaver
Fluke Corporation (retired)
PO Box 9090
Everett, WA 98290
425 335 0615
Deaver@Kendra.com

Jack Somppi
Fluke Corporation
PO Box 9090
Everett, WA 98290
425 446 5469
Jack.Somppi@Fluke.com

Abstract – Recommendations on how to apply the requirements of minimizing the risk of the probability of a false accept decision to a maximum of 2%. The Z540.3 standard states: False Acceptance Decision Risk Specific application (5.3): "Where calibration provides for verification that measurement quantities are within specified tolerances, the probability that incorrect acceptance decisions (false accept) will result from calibration tests shall not exceed 2% and shall be documented." [1] This paper reviews application guidelines from the Z540.3 Handbook for this requirement and makes the recommends using the root difference of squares implementation of Method 6 for most calibration laboratories.

INTRODUCTION

ANSI/NCSL Z540.3 represents another significant paradigm shift in the evolution of metrology. There is considerable evidence of metrology in antiquity for trade, astronomy, time measurements and warfare. The importance of associating accuracy with measurements has long been recognized: "You must have accurate and honest weights and measures, so that you may live long in the land the Lord your God is giving you." [2] In the late 18th century, the concept of interchangeable parts was demonstrated for the production and maintenance of firearms, greatly increasing the interest in the accuracy and repeatability of measurements. The leaders in the new auto industry were those that adopted the assembly line which incorporated this concept of interchangeability. World War II brought about a massive mobilization for the production of the machines of war and a closer association of quality with the manufacturing process. This marriage led to a measurement assurance programs, use of statistical tools and a growing awareness of the concept of measurement uncertainty through the atomic age and the race to space. Until recently, detailed uncertainty analyses were performed only by the highest laboratories. Working laboratories made sure they had good procedures, standards, traceability, quality programs and relied heavily of the definition test specification ratio (TSR) as a figure of merit to determine the adequacy of the measurement process. In this context, TSR (also referred to as Test Uncertainty Ratio (TUR) or Test Accuracy Ratio (TAR)) was defined as the ratio of the specifications of the device under test to the specification of the standard(s) used for the calibration. MIL 45662A and Z540.1 established that TARs of at least 4:1 was a reasonable demonstration of adequacy of the measurement process. There was considerable academic effort expended in studying statistically the risks of false test decisions and in determining that, in most cases, the uncertainty of the measurement process was dominated by the standards. Most of the working labs benefited from the work of the these statisticians and mathematicians by adopting the guidelines but, for the most part not delving into the academics.

With the adoption of ISO/IEC 17025 in 1999, laboratories that want to be recognized to that standard were required to state their measurement uncertainty, pushing more rigorous treatment of uncertainty down to the accredited working laboratories. In addition, the standard stated that the uncertainty much be taken into account if claims of compliance with specifications were made.

Z540.3 RAISES THE BAR

Considerable debate surrounded the writing Section 5.10.4.2 in the ISO/IEC 17025 standard. As a result, the means of taking the measurement uncertainty into account, when making claims of compliance, was not prescribed. Section 5.3 of Z540.3 deals with these claims of compliance and the writing of that section was also hotly debated. Ultimately, the standard was able to give some guidance though it still allows a fairly broad range of interpretation. In releasing this standard, the adequacy of the calibration is stated in terms of risk rather than TSRs or even uncertainties. Section 5.3 states that the probability of false acceptance (PFA) must be constrained to less than 2% when claims of compliance are made. Z540.3 still has a couple escape clauses, however. If acceptable to the client, the lab does not have to make a claim of compliance. Secondly, if the TUR is at least 4:1, the claim of compliance with be deemed to meet the 2% PFA requirement. It is important to note, for Z540.3,the uncertainty used for the denominator for this TUR calculation is the total measurement uncertainty of the measurement process calculated in compliance with the Guide to Uncertainty of Measurement (GUM) [3] and stated at the 95% confidence level.

HOW TO COMPLY WITH THE 2% PFA REQUIREMENT

The working level labs that would like to step up to customer demand and improvements in the management system are finding that Section 5.3 is a formidable barrier. Few are prepared with the academic horsepower and resources to deal with this issue in the depth of those who have been studying measurement decision risk for years. Fortunately, those who have been working in this field, have published considerable guidance and produced a number of good software tools to help. But, there is still some effort that must be expended to understand the breadth of strategies and methods implementing decision rules. Because decision risk is new for many labs, the guidance Handbook for Z540.3 [4] devotes more ink to this one requirement than any other part of the standard. It shows six approaches that can be used to comply with the 2% PFA requirement. However, there is considerable difference in the justification of compliance in each method, resulting in huge differences in the cost of rejected units. The authors contend that all are valid methods but will identify the method they contend is most appropriate for the bulk of the working laboratories.

A BRIEF DESCRIPTION OF THE SIX METHODS

A detailed explanation of these six methods will not be attempted here but they are listed only briefly to illustrate they one method is recommended above the others. These descriptions are also not made very rigorously but are made to highlight the main differences in perspective. More detailed descriptions are in the Z540.3 Handbook and a more rigorous analysis of the risk is found in [5].

Method 1: Unconditional PFA, Test Point Population Data

This method estimates the PFA by making a calculation based on the probability density function (PDF) of both the measurement process and the individual point being measured. A calculation must be made of the convolution of the two PDFs. This is generally done by solving double integrals numerically using general purpose commercial software such as MathCad, MatLab, Maple, or Mathematica, referring to charts of PFA the author has published [6] using MathCad, creating an Excel spreadsheet built to calculate false decision risk [7][8], or purchasing commercially available software such as that commercially available from Integrated Sciences Group [9] designed specifically for the calculation of measurement decision risk. For a complex instrument, documenting the PDF for each measurement point of the unit under test (UUT) can be a considerable burden.

Method 2: Unconditional PFA, MT&E Population Data

The in-tolerance reliability of the unit under test estimates the PFA by making a calculation based on the PDF of the UUT. For a complex instrument, there can be a huge difference in the confidence level of an individual point and of that for the entire UUT. The authors suggest that, for complex instruments, instead of using the in-tolerance reliability directly as an estimate for the confidence of the individual points, a higher reliability be assigned to the individual points based on a model as discussed in [10]. Then the tools in Method 1 can be used to calculate the PFA.

Method 3: Conditional PFA, Acceptance Subpopulation

Like Method 1, this method is calculated using PDFs from both the measurement process and the individual points being measured. However, it assigns a different PDF for the UUT points than Method 1; a subset of only the accepted points instead of the PDF. This results in more aggressive guard bands (smaller acceptance limit) than those of Method 1.

Method 4: Conditional PFA, Bayesian

Like Method 1, this method is calculated using PDFs from both the measurement process and the individual points being measured. However, it also includes the measured value in the calculation of the PDF. Since this uses an prior assumption combined (updated) with the current measured value, the mathematics are similar to those used in Bayesian calculations. The resulting guard bands, however are the most aggressive of all six methods in the handbook.

Method 5: Guard Bands Based on Measurement Uncertainty

These are the simplest guard bands to calculate in that they do not involve the PDF of the UUT at all. They are calculated only from the measurement uncertainty. These result in very aggressive guard bands. The test limit is determined not from the aggregate PFA but are based on the worst case PFA that will be accepted for any individual measurement. The authors argue [11] that, for most labs, guard bands this aggressive, even though they are written into some documentary standards, are not justified and do not adequately share the PFA risk with the probability of a false reject (PFR).

Method 6: Guard Bands Based on TUR

The effort to calculate these guard bands, as in Method 5 are simple because they depend only on the measurement uncertainty when compared with the specification limits of the UUT, result in the TUR. The PDF of the UUT is not considered in the implementation of the method. As with most of the methods, however, the UUT PDF is considered at great length when developing the method and evaluating its effectiveness. The authors have published a number of these TUR based method in [11]. It is the TUR based implementations of Method 6 the authors suggest are the most appropriate for the bulk of the working labs.

Guard Band Required			
Large	**Method 5**		**Method 4**
		Method 2	**Method 3**
Small	**Method 6**		**Method 1**
	Small	**Effort**	**Large**

Table 1: Effort of Implementation and Size of Resulting Guard Bands

TWO TUR BASED GUARD BAND METHODS

In 1998 Michael Dobbert presented a paper [11] in which he also warned that aggressive guard bands often result in false reject rates that are too high. We concur with Mr. Dobbert's conclusion that the least aggressive guard bands that will satisfy the 2% PFA requirement are the most appropriate for most of the labs. He began his analysis by calculating PFA as a function of TUR and the confidence level of the UUT using Method 1. The author presented the results of the same calculations in charts and graphs in a 1993 NCSL paper [6]. However, Mr. Dobbert plotted his results differently. Instead of TUR, he used the confidence level of the UUT as the abscissa.

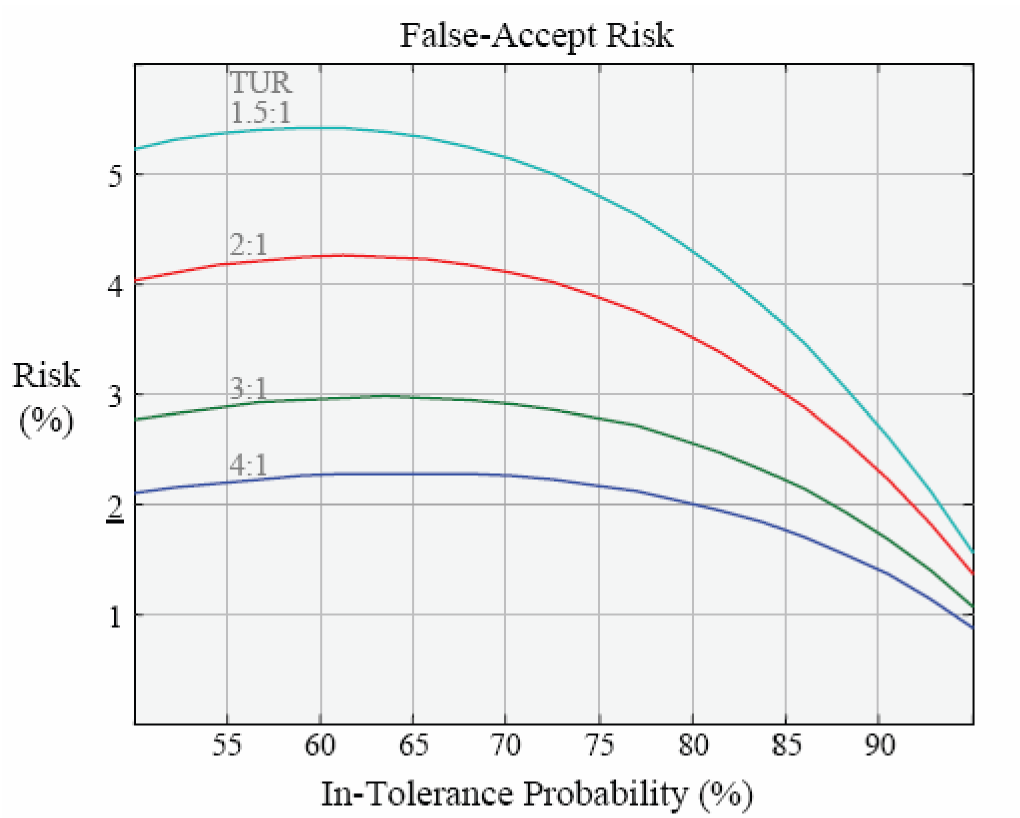

Figure 1: Dobbert's PFA Without Guard Bands [11]

The resulting plots show clearly the curves exhibit a peak PFA. He suggested that just enough guard band be applied using Method 1 to reduce the PFA to the required 2%. He made the calculation for many TURs and empirically fitted a curve to be able to express the required guard band as a function only of TUR:

Eq. 1 $$A = T - U \times M = T - U \times 1.04 - e^{(0.38 \log(TUR) - 0.54)}$$ (Eq. 5 in [11])

Where A is the acceptance limit, T is the specification limit, U is the measurement uncertainty (stated at the 95% confidence interval), M is the multiplier to calculate the guard band determined empirically and the TUR is calculated using the complete measurement uncertainty calculated at the 95% confidence limit.

To aid in the comparison with the Dobbert method with the other methodologies presented in [10] which uses a guard band factor K where TL = SL x K, Eq. 2 shows the relation ship between K and M.

Eq. 2 $$K = \frac{A}{T} = 1 - \frac{U}{T} \bullet M = 1 - \frac{M}{TUR}$$

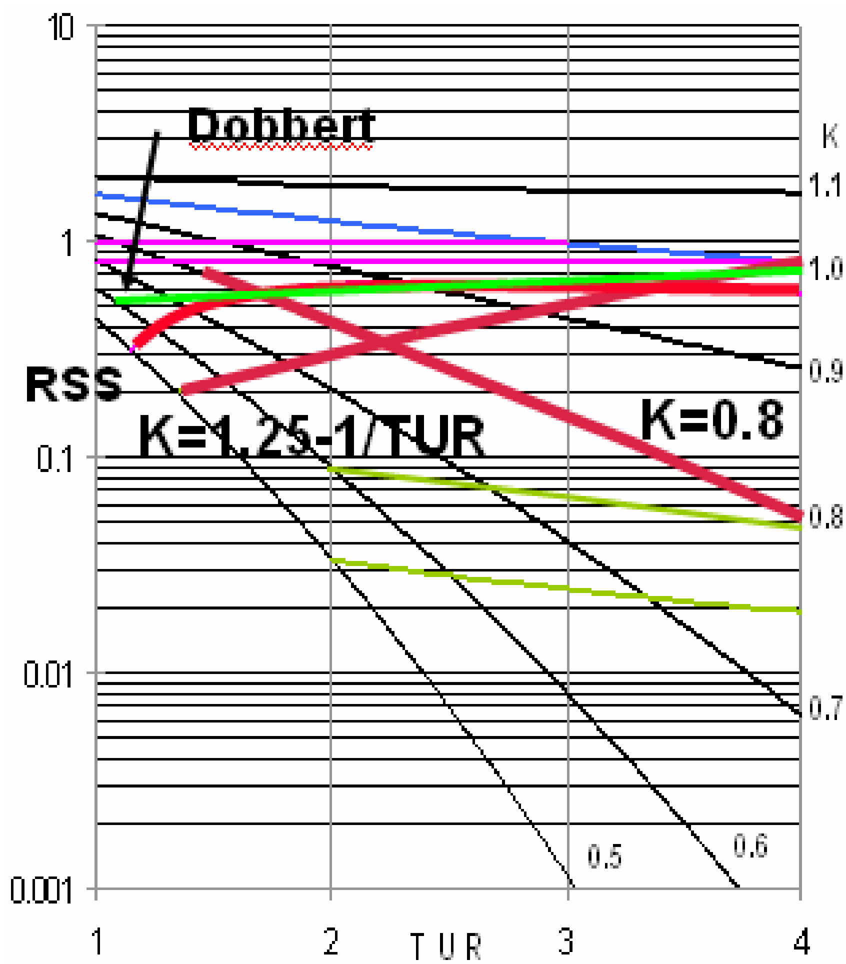

Figure 2: PFA of Dobbert Guard Bands Relative to Other Strategies in [10]

In Figure 2, showing PFA at the 95% confidence level from [10], the Dobbert guard bands are plotted along with those investigated in that paper. It is interesting that the Dobbert guard band appears to be approximated very closely by the RSS, or root-difference-of-squares (RDS) method.

TUR	$K_{Dobbert}$	M	K_{RDS}
1.1	**0.60**	0.44	0.42
1.5	**0.76**	0.36	0.75
2.0	**0.86**	0.28	0.87
2.5	**0.91**	0.21	0.92
3.0	**0.95**	0.16	0.94
3.5	**0.97**	0.10	0.96
3.99	**0.99**	0.05	0.97
4.00	**1.00**		0.97

Table 2: Dobbert Guard bands Compared with RDS Guard bands

IMPLEMENTATION IN MET/CAL

For the many calibration laboratories use the Fluke MET/CAL software, the Dobbert and RDS guard bands are quite easy to implement. For MET/CAL 7.10 and higher, setting the guardband parameter to "rds" in a VSET or TSET command implements the root difference of squares guard bands. Dobbert guard bands can be implemented by specifying "tur" for the guardband parameter and then building a simple table, the guardband_table with relates returns a guard band factor, K, from the TUR value. The Dobbert guard band values vs. TUR are a very smooth, well-behaved function and MET/CAL can be set to interpolate between values in the table so the table does not need to have many entries. The TUR and K values within the bold lines of Table 2 constitute an acceptable guardband_table.

CONCLUSION

The authors' study of PFA in [6] and [10] conducted in the early 1990s re-affirmed the Fluke practice of using the root difference of squares method for guard banding when taking the uncertainty of measurement into account. The authors concur with the work done by Michael Dobbert to implement Method 6 of meeting the Z540.3 2% PFA requirement. It is a justification that, in our opinion, provides compliance with the standard with the least impact to false rejects of the six methods. However, the authors would contend that the root differences of squares method is the preferred method of implementing Z540.3 Handbook Method 6.

REFERENCES

[1] *Requirements for the Calibration of Measuring and Test Equipment*, 2006, ANSI/NCSL Z540.3-2006

[2] Deuteronomy 15:15, New International Version (NIV Bible)

[3] *U.S. Guide to the Expression of Uncertainty in Measurement*, ANSI/NCSL International, Z540.2-1997 (R2002)

[4] *Handbook for the ANSI/NCSL Z540.3-2006*, 2009, ANSI/NCSL International

[5] Castrup, H., *Risk Analysis Methods for Complying with Z540.3,* 2007 NCSL International Conference and Symposium

[6] Deaver, D., *How to Maintain Your Confidence, in a World of Declining TURs,* 1993 National Conference of Standards Laboratories Conference and Symposium

[7] Nicholas, R., *Measurement Decision Risk Simplified*, 1999 Measurement Science Conference and Symposium

[8] Nicholas, R., *How to Build your own Unilateral Consumer Risk Calculation Tool, Including the Treatment of Biases,* 2008 Measurement Science Conference

[9] Integrated Sciences Group, Bakersfield, CA, www.igsmax.com

[10] Deaver, D., *Having Confidence in Specifications,* 1993 National Conference of Standards Laboratories Conference and Symposium

[11] Dobbert, M., *A Guard-Band Strategy for Managing False-Accept Risk,* 2008 NCSL International Conference and Symposium