Risk Analysis Methods for Complying with Z540.3[1]

Speaker/Author: Howard Castrup, Ph.D.
Integrated Sciences Group
14608 Casitas Canyon Rd.
Bakersfield, CA 93306
1-661-872-1683
1-661-872-3669 (fax)
hcastrup@isgmax.com

Abstract

Clause 5.3 of ANSI/NCSL Z540.3, "Requirements for the Calibration of Measuring and Test Equipment," contains a risk management requirement for cases where the results of calibration are employed to verify that quantities are within specified tolerances. This requirement limits the probability of incorrect acceptance decisions (false accept risk) to a maximum of 2%. Where the estimation of this probability is not feasible, the test uncertainty ratio (TUR) is required to be 4:1 or greater.

This paper provides the mathematical framework for computing false accept risk and gives guidelines for managing in-tolerance compliance decisions. The latter includes methods for developing test guardbands that correspond to specified risk levels.

Included in the paper is a discussion of the impact of measurement reliability and measurement uncertainty on false accept risk. A brief discussion is given on the application of the fallback 4:1 uncertainty ratio requirement.

1. Background

Measurement decision risk analysis attempts to quantify the probability of falsely accepting, as a result of calibration or testing, out-of-tolerance equipment parameters or falsely rejecting in-tolerance parameters. The former probability is called "false accept risk," while the latter is called "false reject risk."

The subject of measurement decision risk analysis emerged from consumer/producer risk analysis work done in the late '40s and early '50s by Alan Eagle [1] and Frank Grubbs and Helen Coon [2]. In the mid '50s, the methods of Eagle, Grubbs and Coon were adapted by Jerry Hayes [3] to address decision risk analysis for calibration and testing. Hayes' methods were extended and refined by John Ferling [4], Ken Kuskey [5], Steven Weber and Ann Hillstrom [6], and Howard Castrup [7] in the '70s and '80s, and were extended by Dave Deaver [8], Dennis Jackson [9], Castrup [10], Mimbs [11] and others in the years to follow.

The methods reported in the foregoing paragraph epitomize what is referred to in this paper as "program-level" risk control methods. With these methods, risks are evaluated for each equipment parameter of interest prior to testing or calibration by applying expected levels of unit under test (UUT) parameter in-tolerance probabilities and assumed calibration or test

---

[1] Presented at the 2007 NCSLI Workshop & Symposium, St. Paul, MN.

measurement process uncertainties. With program-level risk control, test limits called "guardband limits" are developed in advance and may be incorporated in calibration or test procedures. Measured values observed outside guardband limits trigger some corrective action, such as adjustment, repair or disposal.

In addition to program-level risk control, bench-level methods are available to control risks in response to equipment parameter values obtained during test or calibration. With bench-level methods, guardbands are superfluous, since corrective actions are triggered by on-the-spot risk or other measurement quality metric computation.

The most robust bench-level method is a Bayesian risk analysis methodology that was developed by Castrup [12] and Jackson [13] in the '80s and later published with the label SMPC (Statistical Measurement Process Control) [14].

Another bench-level approach, referred to as the "confidence level" method, evaluates the confidence that a measured parameter value is in-tolerance, based on the uncertainty in the measurement process. This is not a risk control method in the strictest sense of the term, since no assumptions are made regarding the in-tolerance probability of UUT parameters prior to test or calibration. More will be said on this later.

Over the past few decades, accuracy ratio requirements have been incorporated in various standards and guidelines [15], [16] which provide some implicit control of measurement decision risk. However, explicit risk control requirements have been lacking until recently. Such an explicit requirement is found in the new ANSI/NCSL Z540.3-2006 standard. Section 5.3 of this standard states that, where calibrations provide for verification that measured quantities are in-tolerance, the false accept risk shall not exceed 2%.

Where it is not practical to compute the false accept risk, the standard requires that the accuracy ratio for the measurement, referred to as a "test uncertainty ratio" or TUR, shall be greater than or equal to 4:1. The efficacy of this fallback is a matter of some contention, as is discussed later in this paper.

2. Introduction

This paper provides false accept and false reject risk criteria for managing test and calibration decisions and defines these criteria mathematically. Risk guardbanding is discussed for program-level risk control, and a Bayesian methodology for estimating risks, parameter biases and parameter in-tolerance probabilities is presented. A confidence level method is also discussed.

3. Risk Analysis Alternatives

Three methods of controlling measurement decision risk are described in this section. The first is a program-level control method, while the second and third are suggested as bench-level methods.

3.1 Program-level Control

As mentioned in Section 1, with program-level control methods, risks are estimated and evaluated for each equipment parameter of interest prior to testing or calibration. If the measurement decision risk for a given parameter is judged to be unacceptable, some program-level change is instituted. There are several alternatives, such as relaxing parameter tolerances, reducing measurement process uncertainty, shortening the UUT calibration interval, imposing guardband limits," etc.[2]

3.1.1 Risk Analysis Variables

The basic set of variables that are important for program-level measurement decision risk analysis is

x	-	a deviation from the subject parameter's nominal or reading value
$-L_1$ and L_2	-	the lower and upper tolerance limits for the parameter, expressed as deviations from the parameter's nominal or reading value
$-A_1$ and A_2	-	the lower and upper "acceptance" or "guardband" limits for the parameter, expressed as deviations from the parameter's nominal or reading value
$\mathfrak{L}$	-	the range of values between $-L_1$ and L_2 (the "performance region")
$\mathcal{A}$	-	the range of values between $-A_1$ and A_2 (the "acceptance region")
y	-	a measured value of x.

These variables will be employed in various probability relations. False accept and false reject risks will be constructed from these relations.

3.1.2 Probability Relations

In constructing false accept and false reject definitions, we make use of the following notation

$P(x \in \mathfrak{L})$	-	the *a priori* probability that x lies within $\mathfrak{L}$ prior to test or calibration. This is the probability that the parameter to be measured is in-tolerance at the time of measurement.
$P(y \in \mathcal{A})$	-	the probability that measured values of x will be found within $\mathcal{A}$. This is the probability that measured parameters will be accepted as in-tolerance.
$P(x \in \mathfrak{L}, y \in \mathcal{A})$	-	the probability that a parameter value lies within $\mathfrak{L}$ and its measured value lies within $\mathcal{A}$. Also written as $P(y \in \mathcal{A}, x \in \mathfrak{L})$. This is the joint probability that a parameter will be in-tolerance *and* will be observed to be in-tolerance.
$P(x \notin \mathfrak{L} \mid y \in \mathcal{A})$	-	the probability that the value of a parameter lies outside $\mathfrak{L}$, given that a measurement of its value will lie within $\mathcal{A}$. This is the conditional probability that a parameter will be out-of-tolerance given that it is

[2] The development of risk-based guardbands is discussed in sections 3.1.6 - 3.1.7.2 of this paper.

accepted as in-tolerance.

3.1.3 False Accept Risk

There are two alternative definitions for false accept risk:

Consumer Option (Option 1): The probability that accepted parameters are out-of-tolerance.

Producer Option (Option 2): The probability that out-of-tolerance parameters will be accepted.

The consumer option takes the point of view of the recipient of the calibrated or tested parameters. It is the probability that delivered items will be out-of-tolerance. The producer option, on the other hand, takes the point of view of the testing or calibrating agency. In this view, false accept risk is the probability that an out-of-tolerance parameter will be perceived as being in-tolerance during testing or calibration.

3.1.3.1 False Accept Risk Option 1

Consumer Option. In this view, the false accept risk is the probability an accepted parameter will be found to be out-of-tolerance. This is the probability that a parameter value x will lie *outside* $\mathfrak{L}$ (i.e., not within $\mathfrak{L}$), given that a measurement of its value was found to lie within $\mathcal{A}$:

$$
\begin{aligned}
FA_1 &= P(x \notin \mathfrak{L} \mid y \in \mathcal{A}) = \frac{P(x \notin \mathfrak{L}, y \in \mathcal{A})}{P(y \in \mathcal{A})} \\
&= 1 - \frac{P(x \in \mathfrak{L}, y \in \mathcal{A})}{P(y \in \mathcal{A})}
\end{aligned}
$$

3.1.3.2 False Accept Risk Option 2

Producer Option. In this view, the false accept risk is the probability that a parameter value x will lie outside $\mathfrak{L}$ and will be measured as lying within $\mathcal{A}$:

$$
\begin{aligned}
FA_2 &= P(x \notin \mathfrak{L}, y \in \mathcal{A}) \\
&= P(y \in \mathcal{A}) - P(x \in \mathfrak{L}, y \in \mathcal{A}).
\end{aligned}
$$

3.1.4 Comparing False Accept Risk Options

It is interesting to note that FA_1 is larger than FA_2. This can be seen from the relation

$$FA_1 = \frac{FA_2}{P(y \in \mathcal{A})} .$$

Oddly enough, when false accept risk (or consumer's risk) is discussed in journal articles, conference papers and in some risk analysis software, it is FA_2 that is being discussed. This may be somewhat puzzling in that FA_2 is not referenced to the consumer's perspective. The reason for this "knee-jerk" use of FA_2 is that the development of the concept of FA_1 did not occur until the late 1970s [7]. By then, professionals working in the field had become inextricably entrenched in thinking in terms of FA_2 exclusively.

3.1.5 False Reject Risk

False reject risk is the probability that a parameter value x lying within $\mathfrak{L}$ will be measured as being outside $\mathcal{A}$:

$$\begin{aligned} FR &= P(x \in \mathfrak{L}, y \notin \mathcal{A}) \\ &= P(x \in \mathfrak{L}) - P(x \in \mathfrak{L}, y \in \mathcal{A}). \end{aligned}$$

3.1.6 Controlling Risks with Guardbands

False accept and false reject risks can be controlled at the program-level by the imposition of guardband limits [Hutchinson], [Deaver]. The development of such limits involves additional probability relations.

3.1.6.1 Guardband Risk Relations

It is often useful to relate the range of acceptable values $\mathcal{A}$ to the range $\mathfrak{L}$ by variables called **guardband multipliers**. Let g_1 and g_2 be lower and upper guardband multipliers, respectively. If $-L_1$ and L_2 are the lower and upper parameter tolerance limits, and $-A_1$ and A_2 the corresponding acceptance limits, then

$$\begin{aligned} A_1 &= g_1 L_1 \\ A_2 &= g_2 L_2 . \end{aligned}$$

Suppose that g_1 and g_2 are both < 1. Then the acceptance region $\mathcal{A}$ is smaller than the performance region $\mathfrak{L}$.

If $\mathcal{A}$ is a subset of $\mathfrak{L}$ then

$$P(x \in \mathfrak{L}, y \in \mathcal{A}) < P(x \in \mathfrak{L}, y \in \mathfrak{L}),$$

and

$$P(y \in \mathcal{A}) < P(y \in \mathfrak{L}).$$

It is easy to see that, because $P(y \in \mathcal{A}) / P(y \in \mathfrak{L})$ is less than $P(x \in \mathfrak{L}, y \in \mathcal{A}) / P(x \in \mathfrak{L}, y \in \mathfrak{L})$, we have

$$\frac{P(x \in \mathfrak{L}, y \in \mathfrak{L})}{P(y \in \mathfrak{L})} > \frac{P(x \in \mathfrak{L}, y \in \mathcal{A})}{P(y \in \mathcal{A})}$$

so that

$$FA_{1,\mathcal{A}} < FA_{1,\mathfrak{L}} .$$

It can also be shown that, if $\mathcal{A} < \mathfrak{L}$ then

$$FA_{2,\mathcal{A}} < FA_{2,\mathfrak{L}} ,$$

and, it can easily be seen that, if $\mathcal{A} < \mathfrak{L}$ then

$$FR_{\mathcal{A}} > FR_{\mathfrak{L}} .$$

So, setting the guardband limits inside tolerance limits reduces false accept risk and increases false reject risk. Conversely, setting guardband limits outside tolerance limits increases false accept risk and reduces false reject risk.

3.1.7 Establishing Risk-Based Guardbands

Guardbands can often be set to achieve a desired level of false accept or false reject risk, R. Cases where it is not possible to establish guardbands are those where the desired level of risk is not attainable even with guardband limits set to zero. Moreover, solutions for guardbands are usually restricted to finding symmetric guardband multipliers, i.e., those for which $g_1 = g_2 = g$.[3]

3.1.7.1 False Accept Risk-Based Guardbands

False accept risk-based guardbands are established numerically by iteration.[4] The iteration adjusts the value of a guardband multiplier g, until $FA \cong R$.[5] The following algorithm illustrates the process:

Step 1: Set $g = 1$.
Step 2: Set $A_1 = gL_1$ and $A_2 = gL_2$
Step 3: Compute $P(y \in \mathcal{A})$ and $P(x \in \mathfrak{L}, y \in \mathcal{A})$
Step 4: Compute false accept risk FA (either FA_1 or FA_2, as appropriate).
Step 5: If $FA < R$, go to Case 1. If $FA > R$, go to Case 2 | Double. If $FA \cong R$, the process is complete.

Case 1: $FA < R$

Set $g = g / 2$
Repeat Steps 2 through 4
If $FA < R$, go to Case 1. If $FA > R$, go to Case 2 | Bisect. If $FA \cong R$, the process is complete.

Case 2: $FA > R$

Double: Set $g = 2g$
Repeat Steps 2 through 4
If $FA < R$, go to Case 1. If $FA > R$, go to Case 2 | Double. If $FA \cong R$, the process is complete.

Bisect: Set $g = g + g / 2$
Repeat Steps 2 through 4
If $FA < R$, go to Case 1. If $FA > R$, go to Case 2 | Bisect. If $FA \cong R$, the process is complete.

3.1.7.2 False Reject Risk-Based Guardbands

False reject risk-based guardbands are established in the same way as false accept risk-based guardbands. The following algorithm illustrates the process:

Step 1: Set $g = 1$
Step 2: Set $A_1 = gL_1$ and $A_2 = gL_2$
Step 3: Compute $P(x \in \mathfrak{L})$ and $P(x \in \mathfrak{L}, y \in \mathcal{A})$
Step 4: Compute false reject risk FR

[3] The subject of risk-based guardbands is discussed in detail in [17].

[4] The bisection method is recommended. An excellent selection of bisection algorithms is found in [18].

[5] The condition $FA \cong R$ indicates that the computed FA is within some pre-specified precision of the target risk R.

Step 5: If $FR < R$, go to Case 1 | Double. If $FR > R$, go to Case 2. If $FR \cong R$, the process is complete.

Case 1: $FR < R$

Double: Set $g = 2g$
Repeat Steps 2 through 4
If $FR < R$, go to Case 1 | Double. If $FR > R$, go to Case 2. If $FR \cong R$, the process is complete.

Bisect: Set $g = g + g/2$
Repeat Steps 2 through 4
If $FR < R$, go to Case 1 | Bisect. If $FR > R$, go to Case 2. If $FR \cong R$, the process is complete.

Case 2: $FR > R$

Set $g = g/2$
Repeat Steps 2 through 4
If $FR < R$, go to Case 1 | Bisect. If $FR > R$, go to Case 2. If $FR \cong R$, the process is complete.

3.1.2 Computing Risk

In computing risks, we use uncertainty estimates, along with *a priori* information, to compute the probabilities needed to estimate false accept and false reject risks.

3.1.2.1 Computing Probabilities

Let the value of a subject parameter be labeled x, the in-tolerance region for the parameter be labeled $\mathfrak{L}$ and the probability density function (pdf) for the parameter be denoted $f(x)$. Then the in-tolerance probability $P(x \in \mathfrak{L})$ is written

$$P(x \in \mathfrak{L}) = \int_{\mathfrak{L}} f(x)dx \,.$$

Let the measured value of x be labeled y, the acceptance region be labeled $\mathcal{A}$ and the joint pdf of x and y be denoted $f(x,y)$, then the probability that the subject parameter is both in-tolerance and observed to be in-tolerance is given by

$$\begin{aligned} P(x \in \mathfrak{L}, y \in \mathcal{A}) &= \int_{\mathfrak{L}} dx \int_{\mathcal{A}} dy\, f(x,y) \\ &= \int_{\mathfrak{L}} dx\, f(x) \int_{\mathcal{A}} dy\, f(y \mid x), \end{aligned}$$

where the function $f(y|x)$ is the conditional pdf of obtaining a value (measurement) y, given that the value being measured is x.

3.1.2.2 Parameter Distributions

The normal distribution is usually assumed for the measurement reference parameter. Several distributions have been found useful for the UUT parameter.

3.1.2.2.1 Measurement Reference Parameter

The measurement reference parameter pdf is the conditional pdf $f(y \mid x)$. This pdf is usually given by

$$f(y \mid x) = \frac{1}{\sqrt{2\pi}\sigma_y} e^{-(y-x)^2/2\sigma_y^2} ,$$

where σ_y is the standard uncertainty in the measurement, given by

$$\sigma_y = \sqrt{u_{mte}^2 + u_{other}^2} .$$

In this expression, u_{mte} is the uncertainty in the measurement reference parameter bias and u_{other} is the combined standard uncertainty for any remaining measurement process errors.

3.1.2.2.2 UUT Parameter

A few useful UUT parameter distributions are described below. Others are described elsewhere [19]. For each distribution, the parameter μ represents either the distribution's nominal value or mode value.

Normal Distribution
With this distribution, the pdf for x is given by

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma_x} e^{-x^2/2\sigma_x^2} ,$$

where σ_x is the *a priori* uncertainty in the UUT parameter bias. In this case, the population mean value is equal to the mode value μ.

Lognormal Distribution
The pdf for the lognormal distribution is given by

$$f(x) = \frac{1}{\sqrt{2\pi}\lambda \mid x-q \mid} \exp\left\{ -\left[\ln\left(\frac{x-q}{m-q} \right) \right]^2 \Big/ 2\lambda^2 \right\} .$$

In this expression, λ is referred to as the "shape" parameter. The parameter m is the distribution median value and the parameter q is a limiting value.

In general, for a lognormally distributed variable x, we have

Mode	μ
Median	$m = q + (\mu - q)e^{\lambda^2}$
Mean	$\langle x \rangle = q + (m-q)e^{\lambda^2/2}$
Variance	$(m-q)^2 e^{\lambda^2}(e^{\lambda^2} - 1)$
Standard Deviation	$\mid m-q \mid e^{\lambda^2/2}\sqrt{e^{\lambda^2} - 1}$

Uniform Distribution
The pdf for the uniform distribution is

$$f(x) = \begin{cases} \dfrac{1}{2a}, & -a \le x - \mu \le a \\ 0, & \text{otherwise .} \end{cases}$$

The parameters ± a represent minimum limiting values for the subject parameter.

Exponential Distribution
The exponential distribution is used in cases where the parameter tolerance is single-sided and the physical limit is equal to the mode or nominal value. The pdf is

$$f(x) = \lambda e^{-\lambda x} .$$

Student's t
The Student's t distribution is used to compute a standard uncertainty from an expanded uncertainty, or vice versa, in cases where a measurement error or parameter bias is normally distributed and an estimate of the associated uncertainty is accompanied by a finite degrees of freedom. The pdf for the t distribution is not often used explicitly.

Example: Normally Distributed Subject Parameter
Suppose that the parameter being measured follows an *a priori* (pre-measurement) normal distribution with mean μ and standard deviation σ_x.[6] Then the function $P(x\in\mathfrak{L})$ becomes

$$P(x \in \mathfrak{L}) = \frac{1}{\sqrt{2\pi}\sigma_x} \int_{\mathfrak{L}} e^{-x^2/2\sigma_x^2} dx .$$

Since measurements y of x follow a normal distribution with standard deviation σ_y and mean equal to x, the function $P(x\in\mathfrak{L}, y\in\mathcal{A})$ is written as

$$P(x \in \mathfrak{L}, y \in \mathcal{A}) = \frac{1}{2\pi\sigma_x\sigma_y} \int_{\mathfrak{L}} e^{-x^2/2\sigma_x^2} dx \int_{\mathcal{A}} e^{-(y-x)^2/2\sigma_y^2} dy .$$

The function $P(y\in\mathcal{A})$ is obtained by integrating $f(x,y)$ over all values of x

$$\begin{aligned} P(y \in \mathcal{A}) &= \frac{1}{2\pi\sigma_x\sigma_y} \int_{-\infty}^{\infty} e^{-x^2/2\sigma_x^2} dx \int_{\mathcal{A}} e^{-(y-x)^2/2\sigma_y^2} dy \\ &= \frac{1}{\sqrt{2\pi}\sigma} \int_{\mathcal{A}} e^{-y^2/2\sigma^2} dy , \end{aligned}$$

where

$$\sigma = \sqrt{\sigma_x^2 + \sigma_y^2} .$$

Consider the case of symmetric tolerance limits and guardband limits. Then the tolerance region can be expressed as ± L and the acceptance region as ± A, then the above expressions become

---

[6] For the normal distribution, the mean value is equal to the mode value.

$$P(x \in \mathfrak{L}) = \frac{1}{\sqrt{2\pi}\sigma_x} \int_{-L}^{L} e^{-x^2/2\sigma_x^2} dx$$
$$= 2\Phi\left(L/\sigma_x\right) - 1,$$

$$P(x \in \mathfrak{L}, y \in \mathcal{A}) = \frac{1}{2\pi\sigma_x\sigma_y} \int_{-L}^{L} e^{-x^2/2\sigma_x^2} dx \int_{-A}^{A} e^{-(y-x)^2/2\sigma_y^2} dy$$
$$= \frac{1}{\sqrt{2\pi}\sigma_x} \int_{-L}^{L} \left[\Phi\left(\frac{A-x}{\sigma_y}\right) - \Phi\left(-\frac{A+x}{\sigma_y}\right) \right] e^{-x^2/2\sigma_x^2} dx$$
$$= \frac{1}{\sqrt{2\pi}} \int_{-L/\sigma_x}^{L/\sigma_x} \left[\Phi\left(\frac{A-\sigma_x\zeta}{\sigma_y}\right) + \Phi\left(\frac{A+\sigma_x\zeta}{\sigma_y}\right) - 1 \right] e^{-\zeta^2/2} d\zeta,$$

and

$$P(y \in \mathcal{A}) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-A}^{A} e^{-y^2/2\sigma^2} dy$$
$$= 2\Phi\left(A/\sigma\right) - 1.$$

The value of $P(x \in \mathfrak{L}, y \in \mathcal{A})$ is obtained by numerical integration.

3.1.3 Risk Estimation Procedure

In estimating false accept and false reject risks, we follow the basic "recipe" shown below.

1. Establish the relevant quantities
    - The *a priori* subject parameter distribution.
    - The subject parameter in-tolerance probability prior to test or calibration.
    - The measurement reference parameter distribution.
    - The measurement reference parameter tolerances.
    - The in-tolerance probability for the measurement reference parameter at the time of test or calibration.
2. Estimate the measurement process uncertainty.
3. Compute the false accept and false reject risk.
4. Evaluate risks to determine if corrective action is needed.

3.2 Bench-level Control

Bench-level control methods are those that compute metrics in response to values obtained during test or calibration. Corrective action is triggered if the computed metrics are inconsistent with maximum allowable risk.

3.2.1 Bayesian Analysis

A method, based on Bayes' theorem, was developed in the late 1980s [12] that enabled (1) the analysis of false accept risk for UUT parameters, (2) the estimation of both UUT parameter and reference parameter biases, and (3) the uncertainties in these biases. The method has been referred to as Bayesian risk analysis or, simply, Bayesian analysis.

In applying Bayesian analysis, we estimate the risk of accepting a UUT parameter based on *a priori* knowledge and on the results of measuring the parameter's value during testing or calibration.

These results comprise what is called "post-test" or *a posteriori* knowledge which, when combined with *a priori* knowledge, allow us to compute the quantities of interest.

3.2.1.1 Risk Analysis for a Measured Value

The procedure for applying Bayesian analysis to estimate measurement decision risk associated with a measured parameter value is as follows:

1. Assemble all relevant *a priori* knowledge, such as the tolerance limits for the UUT parameter, the parameter's statistical distribution and the in-tolerance probability for the parameter at the time of measurement. This information is used to compute the UUT parameter bias uncertainty, i.e., the standard deviation of the UUT parameter bias distribution.
2. Perform a measurement or set of measurements. This may either consist of measuring the value of the UUT parameter with the measurement reference, measuring the value of the measurement reference with the UUT parameter or using both parameters to measure a common artifact.
3. Estimate the uncertainty of the measuring process
4. Estimate the UUT parameter and measurement reference biases using Bayesian methods.
5. Compute uncertainties in the bias estimates.
6. Act on the results. Either report the biases and bias uncertainties, along with in-tolerance probabilities for the parameters, or adjust each parameter to correct the estimated biases.

3.2.1.2 *A priori* (Pre-Test) Knowledge

The *a priori* knowledge for a Bayesian analysis may include several kinds of information. For example, if the UUT parameter is the pressure of an automobile tire, such knowledge may include a rigorous projection of the degradation of the tire's pressure as a function of time since the tire was last inflated or a SWAG estimate based on the appearance of the tire's lateral bulge. However *a priori* knowledge is obtained, is should lead to the following quantities:

- An estimate of the uncertainty in the bias of the UUT parameter value. This estimate may be obtained heuristically from containment limits and containment probabilities or by other means, if applicable [20].
- An estimate of the uncertainty in the measurement process, accounting for all error sources, including the bias in the measurement reference.

3.2.1.3 *A Posteriori* (Post-Test) Knowledge

The *a posteriori* knowledge in a Bayesian analysis consists of the results of measurement. As stated earlier, these results may be in the form of a measurement or a set of measurements. The measurements may be the result of readings provided by the measurement reference from measurements of the UUT parameter, readings provided by the UUT parameter from measurements of the measurement reference, or readings provided by both parameters, taken on a common artifact.

3.2.1.4 Bias Estimates

Bias estimates for the UUT parameter and measurement reference can be estimated using Bayesian analysis. The variables are

- $\bar{y}_s$ - The perceived value of the UUT parameter. This may be (1) the parameter's nominal value, (2) the mean value of a sample of measurements taken on an artifact by the UUT parameter, or (3) the mean value of a sample of direct measurements of the measurement reference by the UUT parameter.[7]
- n_s - Sample size of the UUT parameter measurement(s).
- s_s - UUT parameter sample standard deviation.
- u_{sb} - UUT parameter bias standard uncertainty.
- $\bar{y}_m$ - The perceived value of the measurement reference parameter. This may be (1) the reference parameter's nominal value, (2) the mean value of a sample of measurements taken on an artifact by the parameter, or (3) the mean value of a sample of direct measurements by the measurement reference of the UUT parameter value.
- n_m - Sample size of the reference parameter measurements.
- s_m - Measurement reference parameter sample standard deviation.
- u_{mb} - Measurement reference parameter bias standard uncertainty.
- $u_{process}$ - Standard uncertainty of the measurement process, excluding parameter bias uncertainty.

Let

$$r_s = u_s / u_m ,$$

and

$$r_m = u_m / u_s .$$

Then, UUT parameter and measurement reference parameter biases can be estimated according to

[7] The nominal value of a 10 VDC voltage source is an example of case (1). The measurement of ambient air pressure by a pressure gage (UUT parameter) and a dead weight tester (measurement reference parameter) is an example of case (2). The direct measurement of a reference mass (measurement reference parameter) by a precision balance (UUT parameter) is an example of case (3).

$$\text{UUT Parameter Bias} = \frac{r_s^2}{1+r_s^2}(\bar{y}_s - \bar{y}_m)$$

$$\text{Measurement Reference Parameter Bias} = \frac{r_m^2}{1+r_m^2}(\bar{y}_m - \bar{y}_s),$$

where

$$u_s = \begin{cases} u_{sb}, & \text{when estimating UUT parameter bias} \\ \sqrt{u_{sb}^2 + \dfrac{s_s^2}{n_s} + u_{process}^2}, & \text{when estimating measurement reference parameter bias,} \end{cases}$$

and

$$u_m = \begin{cases} \sqrt{u_{mb}^2 + \dfrac{s_m^2}{n_m} + u_{process}^2}, & \text{when estimating UUT parameter bias} \\ u_{mb}, & \text{when estimating measurement reference parameter bias.} \end{cases}$$

3.2.1.5 Bias Uncertainty Estimates

As indicated above, the variable $u_{process}$ is a combination of uncertainties due to process error sources, excluding the random components and the bias uncertainties. The quantities u_{sb} and u_{mb} are *a priori* estimates for UUT parameter and measurement reference parameter bias uncertainties. They are usually computed from tolerance limit and percent in-tolerance information on the populations of items from which the subject unit and measuring unit are drawn.

For instance, if the percent in-tolerance for the UUT parameter population is $100 \times p$ %, the tolerance limits are $-L_1$ and L_2, and the population's probability density function is $f(x,u_{sb})$, then u_{sb} is determined by inverting the equation

$$p = \int_{-L_1}^{L_2} f(x, u_{sb})dx.$$

For example, if x is normally distributed with mean μ, variance u_{sb}^2, and $L_1 = L_2 = L$, then

$$u_{sb} = \frac{L}{\Phi^{-1}\left(\dfrac{1+p}{2}\right)},$$

where $\Phi^{-1}(\cdot)$ is the inverse normal distribution function. Uncertainties in the parameter bias estimates are computed from

$$\text{UUT Parameter Bias Uncertainty} = \frac{r_s^2}{1+r_s^2}u_m,$$

and

$$\text{Measurement Reference Parameter Bias Uncertainty} = \frac{r_m^2}{1+r_m^2}u_s.$$

3.2.1.6 UUT Parameter In-Tolerance Probability

Following measurement, the in-tolerance probability of a two-sided symmetrically toleranced UUT parameter may be estimated from the relation

$$\begin{aligned} P_{in} &= P(-L \le x \le L) \\ &= \Phi(a_+) + \Phi(a_-) - 1, \end{aligned}$$

where

$$a_\pm = \frac{\sqrt{1+r_s^2}}{u_s}\left[L_s \pm \frac{r_s^2}{1+r_s^2}(\bar{y}_s - \bar{y}_m) \right],$$

In this expression, $\pm L_S$ are the UUT parameter tolerance limits, and the function $\Phi(a_\pm)$ is defined according to

$$\begin{aligned} \Phi(a_\pm) &= P(x < a_\pm) \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a_\pm} e^{-x^2/2} dx \,. \end{aligned}$$

Extension to single-sided and asymmetrically tolerance parameters is straightforward.[8]

3.2.1.7 Applying Bayesian Estimates

Corrective action may be called for if P_{in} is less than a predetermined specified limit. If the 2% requirement of Z540.3 is applied, corrective action is called for if $P_{in} < 0.98$.

3.2.2 Confidence Level Method

With this method of analysis, the confidence that a UUT parameter value lies within its tolerance limits is computed, given a measured value or "deviation from nominal" x. This method is applied when an estimate of the UUT parameter in-tolerance probability is not feasible. As such, it is not a true "risk control" method, but rather an application of the results of measurement uncertainty analysis.

For the present discussion, assume that the measurement process error is normally distributed with mean x and variance u_{proc}^2.

3.2.2.1 Two-Sided Tolerances

Suppose that the lower and upper tolerance limits for the UUT parameter are $-L_1$ and L_2, respectively. Then, given a measured deviation from nominal x, the in-tolerance confidence level is obtained from

[8] See, for example [21].

$$
\begin{aligned}
P_{in} &= \frac{1}{\sqrt{2\pi}u_{proc}} \int_{-L_1}^{L_2} e^{-(y-x)^2/2u_{proc}^2} dy \\
&= \Phi\left(\frac{L_2 - x}{u_{proc}}\right) - \Phi\left(-\frac{L_1 + x}{u_{proc}}\right) \\
&= \Phi\left(\frac{L_2 - x}{u_{proc}}\right) + \Phi\left(\frac{L_1 + x}{u_{proc}}\right) - 1,
\end{aligned}
$$

where the function Φ is the cumulative normal distribution function.

3.2.2.2 Single-Sided Upper Limits

For cases where a "not to exceed" upper tolerance limit L applies, P_{in} becomes

$$
\begin{aligned}
P_{in} &= \frac{1}{\sqrt{2\pi}u_{proc}} \int_{-\infty}^{L} e^{-(y-x)^2/2u_{proc}^2} dy \\
&= \Phi\left(\frac{L - x}{u_{proc}}\right).
\end{aligned}
$$

3.2.2.3 Single-Sided Lower Limits

For cases where a minimum allowable lower tolerance limit L applies, P_{in} becomes

$$
\begin{aligned}
P_{in} &= \frac{1}{\sqrt{2\pi}u_{proc}} \int_{L}^{\infty} e^{-(y-x)^2/2u_{proc}^2} dy \\
&= 1 - \Phi\left(\frac{L - x}{u_{proc}}\right).
\end{aligned}
$$

3.2.2.4 Applying Confidence Level Estimates

As with the Bayesian method, corrective action may be called for if P_{in} is less than a predetermined specified limit. In keeping with the Z540.3 requirement, corrective action is called for if $P_{in} < 0.98$.

4. The 4:1 Alternative

As mentioned earlier, Z540.3 allows meeting a TUR requirement of 4:1 in cases where estimating false accept risk is not practical. This requirement will now be discussed in some detail.

4.1 The Z540.3 Definition

Z540.3 defines TUR as the ratio of the span of the UUT tolerance to twice the "95%" expanded uncertainty of the measurement process used for calibration. A caveat is provided in the form of a note stating that this requirement applies only to two-sided tolerances. Actually, it can be rigorously applied only in cases where the two-sided tolerances are symmetric.

Mathematically, the 4:1 Z540.3 TUR definition is stated for tolerance limits $-L_1$ and L_2 as

$$\mathrm{TUR} = \frac{L_1 + L_2}{2U},$$

where U is equal to the standard uncertainty u of the measurement process multiplied by a coverage factor k

$$U = ku.$$

In Z540.3, $k = 2$.

4.2 A Critique of the 4:1 Requirement

In a telephone conversation with Jerry Hayes, the author of the original 4:1 requirement, he said "I can't believe they're still using that old accuracy ratio requirement." When Jerry developed this requirement, it was with the knowledge that test and calibration facilities would be hard pressed to estimate measurement decision risks using the computing machinery available at the time and needed a simple criterion that provided some control of false accept risk.

In the present day, sufficient computing power is readily available and risk estimation methods are so well documented that it is difficult to understand why a crude 1955 risk control patch is still being implemented in lieu of more refined and, ultimately, more productive criteria.

With this in mind, characteristics of the Z540.3 TUR deserve mention.

- The requirement does provide a ratio expressed in terms of measurement uncertainty relative to UUT tolerance limits. It is, at best, a crude risk control tool, i.e., one that does not controls risks to any specified level. Moreover, in some cases, it may be superfluous. For instance, what if all UUT parameters of a given manufacturer/model were in-tolerance prior to test or calibration? In this case the false accept risk is zero regardless of the TUR.
- As Mimbs points out [24], failing to account for the *a priori* in-tolerance probability of the UUT parameter is a major flaw in the TUR requirement as stated in Z540.3.[9]
- The requirement is not applicable to all measurement scenarios. It does not apply when tolerances are asymmetric or single-sided.
- The requirement treats the expanded uncertainty $2u$ as a 95% confidence limit. To say that an expanded uncertainty determined in this way is analogous to a 95% confidence level is like saying that an infant banging the floor with a rattle is analogous to Tiger Woods hitting a 5-iron.

4.2.1 Equivalent Accuracy Ratios (TURs)

If risks are under control, there is no real need to establish or meet TUR requirements. However, the use of TUR as a measurement quality metric has become so ingrained that it deserves at least some mention.

In days past, when instrument tolerances were specified as simple ± limits, it made sense to quantify the quality of a measurement in terms of the ratio of the UUT parameter tolerance limits to the measurement reference parameter tolerance limits. This ratio was referred to as the

[9] Guardbands consisting of multiples of measurement process uncertainty also suffer from this flaw. If the *a priori* bias uncertainty estimate is not taken into account, such guardbands are not risk-based.

"accuracy ratio." Its very simplicity guaranteed that it would be extensively used. In fact, nominal accuracy ratio requirements soon emerged in Government contractual standards, e.g., MIL-STD-45662A [15].

In practice, not all tolerances are of the simple ± variety. In fact, some tolerances are single-sided, making it impossible to sensibly divide one set into the other.

In addition, for some technologies, a nominal accuracy ratio requirement, such as 4:1, cannot be met, and simple accuracy ratio prescriptions are of no value.

These considerations led to the development of something called an "equivalent accuracy ratio." This is the TUR that corresponds to the level of risk associated with a test or calibration under what are called "baseline" conditions.

4.2.1.1 Risk Baselines

A risk baseline is determined by specifying a nominal accuracy ratio (e.g., 4:1), UUT parameter and measurement reference parameter nominal in-tolerance probabilities (e.g., 95%), a risk key (false accept or false reject), and, if the key is false accept risk, the risk option (consumer or producer). Once a baseline is established, the computation of baseline risks is simple. An example is shown in the accompanying graphic for a 4:1 baseline with UUT parameter and measurement reference parameter in-tolerance probabilities of 95%.

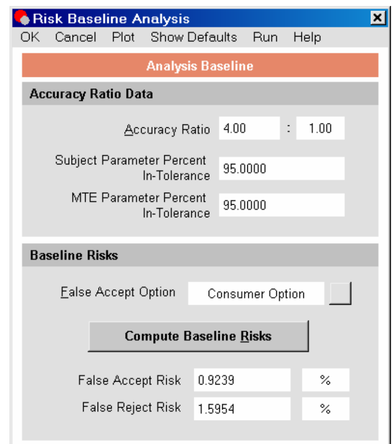

Figure 1. Baseline Definition. A typical risk baseline. In the case shown, the Consumer Risk option has been selected

Once the false accept and false reject risk are computed for a measurement scenario of interest, an equivalent accuracy ratio is determined. This is done by calculating the accuracy ratio that would be required to produce a baseline risk equal to the computed risk, given the stated parameter in-tolerance probabilities.

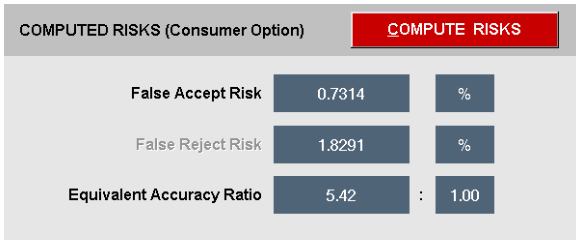

Figure 2. Equivalent Accuracy Ratio Example. The computed risk for a case where the nominal accuracy ratio is 2:1, but the UUT parameter in-tolerance probability is 97% and the measurement reference parameter in-tolerance probability is 99.73%. Since these probabilities are higher than the baseline in-tolerance probabilities, both the subject parameter and measurement reference parameter bias uncertainties are smaller than what they would be under baseline conditions. Hence, the equivalent accuracy ratio is higher than the baseline accuracy ratio.

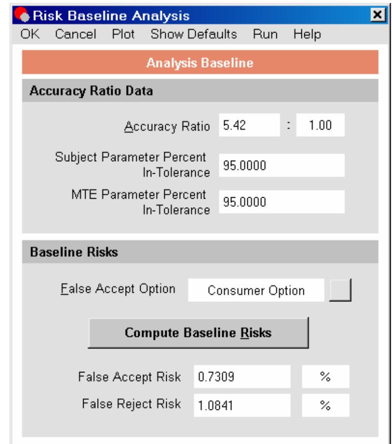

Figure 3. Equivalent Accuracy Ratio Example (cont.). The equivalent accuracy ratio is set equal to the accuracy ratio that yields a baseline risk approximately equal to the computed risk (to the displayed level of precision) under baseline conditions.

The above figure portrays a case where the nominal accuracy ratio is only 2:1. However, because of the high UUT parameter in-tolerance probability, the equivalent accuracy ratio turns out to be over 5:1.

It should be said that the development of an equivalent accuracy ratio requires the estimation of measurement decision risk. Since this is the case, the equivalent accuracy ratio is of no real use, in that estimated risks can be compared directly with risk criteria to determine if corrective action is needed. Equivalent accuracy ratio is provided solely to help ease the evolution from meeting TUR requirements to meeting risk criteria.

5.0 Conclusion

Z540.3 takes a bold step in the direction of controlling measurement quality by requiring that measurement decision risks be held to maximum acceptable levels. This requirement will no doubt find many calibration organizations in an unprepared state, since a knowledge of risk analysis methodology is required to meet the challenge of the new standard.

Pursuant to this, three risk control methods have been described in this paper. One method, referred to as a "program-level" control method, involves computing false accept risk for expected or assumed test or calibration measurement uncertainties and anticipated UUT in-tolerance probabilities. Two varieties of false accept risk were defined.[10]

Two bench-level methods were also described. A true risk control Bayesian method and a "pseudo" risk control confidence level method. Each method involves calculations that are

[10] While it is not explicitly stated in Z540.3 which definition is to be applied, preliminary work being done by a Z540.3 handbook writing committee suggests that the producer option is preferred. This is sometimes referred to as "unconditional false accept risk" or UFAR.

simple enough to apply at the test or calibration bench by technicians using small footprint software applications.

In addition, a brief discussion of TUR was given in which the drawbacks of relying on TUR as a measurement quality metric were presented.

6.0 References

[1] Eagle, A., "A Method for Handling Errors in Testing and Measuring," *Industrial Quality Control*, March, 1954.

[2] Grubbs, F. and Coon, H., "On Setting Test Limits Relative to Specification Limits," *Industrial Quality Control*, March, 1954.

[3] Hayes, J., "Factors Affecting Measuring Reliability," U.S. Naval Ordnance Laboratory, TM No. 63-106, 24 October 1955.

[4] Ferling, J., *Calibration Interval Analysis Model*, SAI Comsystems Technical Report, Prepared for the U.S. Navy Metrology Engineering Center, Contract N00123-87-C-0589, April 1979.

[5] Kuskey, K., *Cost-Benefit Model for Analysis of Calibration-System Designs in the Case of Random-Walk Equipment Behavior*, SAI Comsystems Technical Report, Prepared for the U.S. Navy Metrology Engineering Center, Contract N00123-76-C-0589, January 1979.

[6] Weber, S. and Hillstrom, A., *Economic Model of Calibration Improvements for Automatic Test Equipment*, NBS Special Publication 673, April 1984.

[7] Castrup, H., "Evaluation of Customer and Manufacturer Risk vs. Acceptance Test In-Tolerance Level," TRW Technical Report 99900-7871-RU-00, April 1978. Later published in *Proc. NCSL Workshop & Symposium*, Gaithersburg, November 1980.

[8] Deaver, D., "How to Maintain Your Confidence," *Proc. 1993 NCSL Workshop and Symposium*, Albuquerque, July 1993.

[9] Jackson, D., "Measurement Risk Analysis Methods," Proc. Meas. Sci. Conf., Anaheim, 2005.

[10] Castrup, H., "Calibration Requirements Analysis System," *Proc. NCSL Workshop and Symposium*, Denver, July 1989.

[11] Mimbs, S., "ANSI/NCSL Z540.3: Measurement Decision Risk and the 2% Rule," NASA Presentation, March 2007.

[12] Castrup, H., *Intercomparison of Standards: General Case*, SAI Comsystems Technical Report, U.S. Navy Contract N00123-83-D-0015, Delivery Order 4M03, March 16, 1984.

[13] Jackson, D.H., Instrument Intercomparison: *A General Methodology, Analytical Metrology Note AMN 86-1*, U.S. Navy Metrology Engineering Center, NWS Seal Beach, January 1, 1986; and "Instrument Intercomparison and Calibration," *Proc. Meas. Sci. Conf.*, Irvine, January 1987.

[14] Castrup, H., " Analytical Metrology SPC Methods for ATE Implementation," *Proc. NSCL Workshop & Symposium*, Albuquerque, August 1991.

[15] MIL-STD 45662A, *Calibration Systems Requirements*, U.S. Dept. of Defense, 1 August 1988.

[16] ANSI/NCSL Z540-1-1994, *American National Standard for Calibration Laboratories and Measuring and Test Equipment—General Requirements*, NCSL, August 1994.

[17] Castrup, H., "Risk-Based Control Limits," Proc. Measurement Science Conference, Anaheim, January 2001.

[18] Press, et al., *Numerical Recipes, 2nd Ed*., Cambridge University Press, 1992, New York

[19] Castrup, H., "Distributions for Uncertainty Analysis," Proc. International Dimensional Workshop, Knoxville, May 2001. (Revised April 2007. See www.isgmax.com/articles_papers.asp).

[20] Castrup, H., "Estimating Parameter Bias Uncertainty," *Proc. NCSLI Workshop & Symposium*, Washington, D.C., July-August 2001.

[21] *AccuracyRatio 1.6*, © 2004, Integrated Sciences Group, www.isgmax.com.

[22] Hutchinson, B., "Setting Guardband Limits to Satisfy MIL-STD 45662A Requirements, Proc. NCSL Workshop & Symposium, Albuquerque, August 1991.

[23] Deaver, D., "Guardbanding with Confidence," Proc., NCSL Workshop & Symposium, Chicago, July - August 1994.

[24] Mimbs, S., " Measurement Decision Risk – The Importance of Definitions," Proc. NCSLI Workshop & Symposium, St. Paul, August 2007.