# Measurement Uncertainty and Risk for Navy Calibrations

Dennis Jackson, Ph.D.
Mathematical Statistician
U.S. Navy
NSWC Corona Division
1-909-273-4492
jacksonDH@corona.navy.mil

**Abstract**

The US Navy has a policy of presenting the capability and accuracy of test equipment in terms of engineering tolerances and measurement reliability. Industry tends to discuss these same characteristics in terms of measurement uncertainty and confidence levels. The concepts are closely related and these relationships are described mathematically. Risk probabilities are key to assessing the meaning of capability and accuracy measures. In addition to discussing calibration risks, risks to the end item are also discussed.

**Introduction**

For the non-statistician, discussions of measurement uncertainty can sometimes be bewildering. For US Navy metrology engineers, there are also the additional concepts of engineering tolerances and measurement reliability to deal with. And somehow, all of these things relate to the usefulness of test equipment in testing real systems like radars and radios.

In this paper, we will discuss the exact meanings of these concepts, and show how they relate mathematically. In addition, we will look at two key bottom line concepts:

- Calibration risk – What is the probability that a given calibration process will result in an out of tolerance test instrument?

- System risk – What is the probability that a test instrument that belongs to a given calibration chain will incorrectly test the system it was designed for?

The research in this paper was a further development from the research presented in References [1] and [2].

**What is a Calibration?**

Calibration of a piece of measurement test equipment is performed by comparing the measurement values from a UUT (Unit Under Test) with the measurement values from a calibrator. This comparison is given mathematically by:

$$\textit{Comparison} = \textit{UUT Measurement} - \textit{Calibrator Measurement} \qquad (1)$$

This description applies to situations where the calibrator and the UUT take measurements on an artifact. The same description can be applied to a situation where the UUT takes measurements on a calibration standard which outputs an indicated value. The comparison is then given by:

$$\textit{Comparison} = \textit{UUT Measurement} - \textit{Calibrator Indicated Value} \tag{2}$$

Similarly, if the UUT is outputting an indicated value which is measured by a calibrator, the comparison is given by:

$$\textit{Comparison} = \textit{UUT Indicated Value} - \textit{Calibrator Measurement} \tag{3}$$

To determine whether a UUT is in tolerance, the calibration comparison is compared with engineering tolerances. A UUT is considered in tolerance if:

$$\textit{Lower Tolerance} \le \textit{Comparison} \le \textit{Upper Tolerance} \tag{4}$$

Test equipment that are Out of Tolerance (OOT) are adjusted using the comparison between the calibrator and the UUT. If the test equipment is in tolerance, it may be either adjusted or not adjusted depending on the policy of the calibration program. The U.S. Navy has the policy of not adjusting test equipment that are in tolerance.

**Calibration Intervals**

The test equipment is recalled to the calibration lab after a prescribed amount of time to be calibrated. This prescribed amount of time is called the calibration interval. Calibration intervals are determined in such a way as to assure that the test equipment is in tolerance.

One would prefer, of course, that the test equipment always be in tolerance. However, errors in measurement exhibit random behavior which makes it virtually impossible to guarantee this. It is only possible to ensure a given (lower than 100%) probability of in tolerance.

The probability that a piece of test equipment is in tolerance is called its measurement reliability. Mathematically this is written as:

$$\textit{Measurement Reliability} = \Pr(\textit{In Tolerance}) \tag{5}$$

$$= \Pr(\textit{Lower Tolerance} \le \textit{Comparison} \le \textit{Upper Tolerance}) \tag{6}$$

Measurement reliability tends to decrease with time since calibration. Therefore, we will write measurement reliability as a function of time or:

$$R(t) = \Pr(\textit{In Tolerance at time t after calibration}) \tag{7}$$

An example of a measurement reliability function is given in Figure 1.

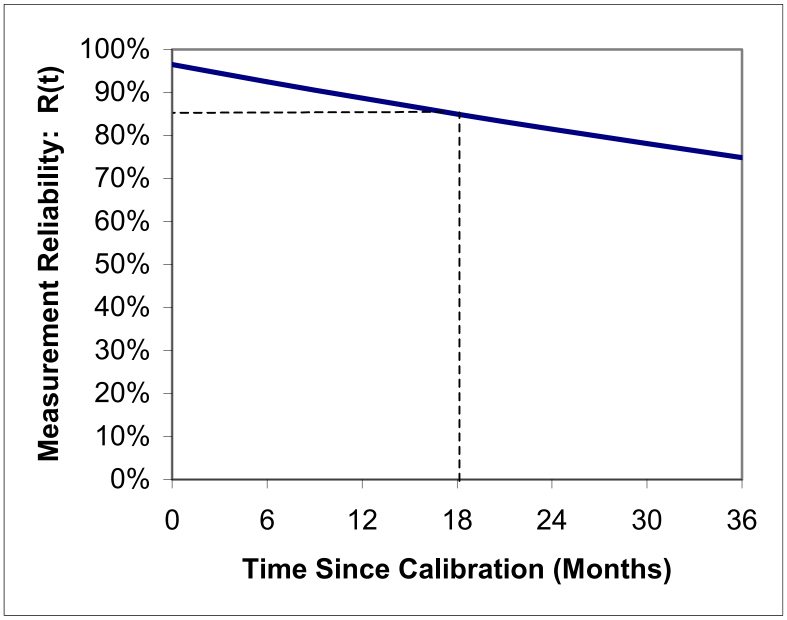


**Figure 1. Measurement Reliability**

Given a minimum acceptable or target measurement reliability, the calibration interval is determined by finding the amount of time until this target reliability is reached. In the example in Figure 1, an 85% target reliability is represented by a horizontal dashed line. This horizontal line reaches the Measurement Reliability curve at 18 months. Therefore, the appropriate calibration interval in this example would be 18 months. Such a calibration interval would ensure that there was at least an 85% probability that the test equipment was in tolerance during the entire calibration interval.

Mathematically, the calibration interval is found by solving the reliability function for the amount of time since calibration to reach the reliability target. If we define the calibration interval as $I$, and the reliability target as $R_T$, the calibration interval is found by solving the following equation:

$$R(I) = R_T \tag{8}$$

The solution to this equation is represented by:

$$I = R^{-1}(R_T) \tag{9}$$

**Measurement Error Model**

The measurement model, shown in Figure 2, relates the value of the quantity being measured, referred to as the ***measurand***, to the value obtained as a result of measurement. The difference between the two values is called the ***measurement error***. This will be expressed in mathematical form as:

$$Measurement = True\ Value + Error \tag{10}$$


**Figure 2. Measurement Model**

**Measurement Uncertainty**

Measurement uncertainty is a way of characterizing the range of values that the measurement error might take. A given measurement is drawn from a distribution of measurements such as the one shown in Figure 3.

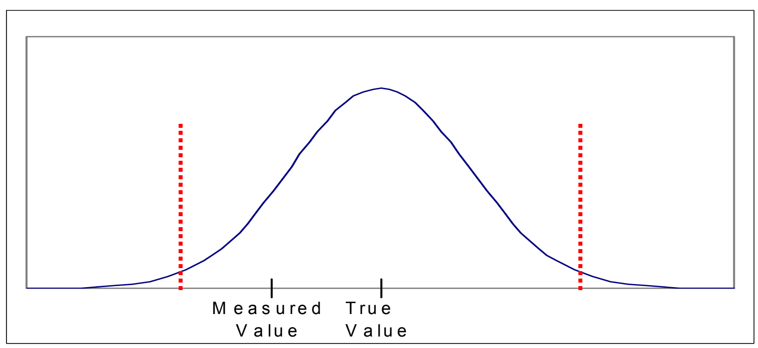


**Figure 3. Distribution of Measurements**

The distance from the true value to the dashed lines characterizes how big a measurement error could be and is called the uncertainty. Note that the uncertainty is not a specific measurement error, but rather helps to establish a range of values for the error. So we could say that

$$Measurement\ Range = True\ Value \pm Uncertainty \tag{11}$$

Or, equivalently:

$$True\ Value\ Range = Measurement \pm Uncertainty \tag{12}$$

The addition of an uncertainty to a measurement indicates that we don't know the true value exactly, but that we know a range of values for it.

Estimates of uncertainty are derived from the statistical concept of the standard deviation. The standard deviation is a characteristic or number that describes the amount of variability of a distribution of measurements.

For measurements that can be described using the normal, or bell-shaped, distribution the true value plus or minus a single standard deviation contains approximately 68% of the measurements that would result. Two standard deviations contain approximately 95% of the population of measurements, and three standard deviations contain approximately 99.7%. It is very common to express uncertainty as two standard deviations.

$$Uncertainty = 2 * Standard\ Deviation \tag{13}$$

Statistically, the uncertainty of a measurement is the same as the uncertainty of the measurement error. If we use the notation $u(x)$ to indicate the standard uncertainty (or standard deviation) of $x$, this would be expressed as:

$$Standard\ Uncertainty = u(Measurement) = u(Error) \tag{14}$$

**Measurement Uncertainty and Measurement Reliability**

Measurement uncertainty and measurement reliability would seem to be related concepts. They both have related probability statements. The measurement uncertainty is related to a confidence level expressed by:

$$Confidence\ Level = \Pr(-Uncertainty \leq UUT\ Measurement - True\ Value \leq +Uncertainty) \tag{15}$$

The measurement reliability is related to the engineering tolerances that are used during calibration. This probability is expressed by:

$$Measurement\ Reliability = \Pr(Lower\ Tolerance \leq UUT\ Measurement - Calibrator\ Measurement \leq Upper\ Tolerance) \tag{16}$$

Both probabilities are meant to place bounds on the location of the UUT Measurement with respect to the True Value. However, the measurement reliability uses the Calibrator Measurement in place of the True Value.

If the Calibrator Measurement is very close to the True Value, then Measurement Reliability and the uncertainty Confidence Level are very close to the same thing. This means that the engineering tolerance is a form of uncertainty with the Measurement Reliability being the approximate Confidence Level. The mathematical relationship between uncertainty, measurement reliability, and the engineering tolerances will be further discussed in subsequent sections.

**Estimating Uncertainty Using Calibration Results**

For two-sided symmetric tolerance limits we could express the observed measurement reliability as:

$$R_o(UUT) = \Pr(-Tolerance\ Limit \leq Comparison \leq +Tolerance\ Limit)$$

$$= 2F\left(\frac{ToleranceLimit}{u(Comparison)}\right) - 1 \tag{17}$$

where:

| | | |
|---|---|---|
| $R_o(UUT)$ | = | The observed measurement reliability at the end of the calibration interval for the UUT. |
| $F(z)$ | = | The Standard Normal distribution function |

In all calibration scenarios, the calibrator measurement (or indicated value) is compared with the UUT measurement (or indicated value). The squared standard uncertainty due to this comparison is:

$$u^2(Comparison) = u^2(UUT) + u^2(Calibrator) \tag{18}$$

This means that equation (17) can be solved for the Tolerance Limit (see the Appendix for further details) by:

$$Tolerance\ Limit = Z_{\left(\frac{1+R_o(UUT)}{2}\right)} \sqrt{u^2(UUT) + u^2(Calibrator)} \tag{19}$$

where:

| | | |
|---|---|---|
| $Z_p$ | = | The p$^{th}$ percentile of the Standard Normal distribution. |

Solving this equation for the uncertainty of the UUT gives:

$$u^2(UUT) = \left[\frac{Tolerance\ Limit}{Z_{\left(\frac{1+R_o(UUT)}{2}\right)}}\right]^2 - u^2(Calibrator) \tag{20}$$

The expression in (20) assumes that there is a known measurement uncertainty for the calibrator. This would be true if the calibrator came from an uncertainty testing calibration process (report of calibration). Unfortunately, if the calibrator comes from a tolerance testing process, its uncertainty will be a function of the calibrator's calibrator. This relationship will continue until there is a known uncertainty from an uncertainty testing calibration process. This is pictured in Figure 4.

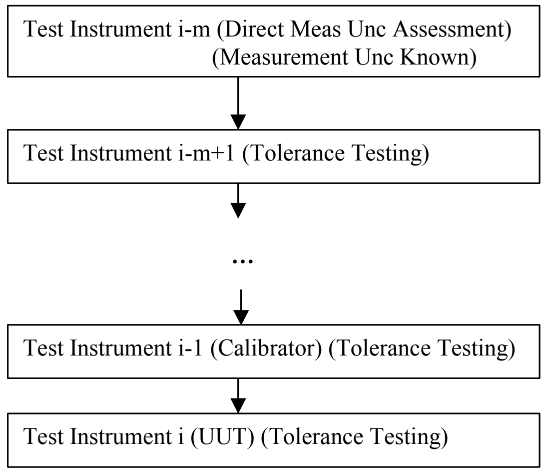


**Figure 4. Tolerance Testing Traceability Chain**

This recursive relationship gives rise to the following equation which takes into account the contribution of all of the calibration labs in the traceability chain.

$$u^2(UUT) = \left[\frac{UUT\ Tolerance\ Limit}{Z_{\left(\frac{1+R_o(UUT)}{2}\right)}}\right]^2 +$$

$$\sum_{k=1}^{m-1}(-1)^k\left[\frac{i-k^{th}\ Level\ Tolerance\ Limit}{Z_{\left(\frac{1+R_{o,aop}(i-k^{th}\ Level\ Test\ Instrument)}{2}\right)}}\right]^2 +$$

$$(-1)^m u^2(i-m^{th}\ Level\ Test\ Instrument) \qquad (21)$$

where:

| | | |
|---|---|---|
| $i$ | = | The level of the UUT in the Traceability Chain. |
| $k$ | = | The number of calibrators up from the UUT in the Traceability Chain. |
| $i$-$m$ | = | The i-m$^{th}$ calibrator up from the UUT in the Traceability Chain has a direct assessment of the measurement uncertainty. |
| $R_{o,aop}(k^{th}\ Level\ Test\ Instrument)$ | = | The observed Average Over Period (AOP) measurement reliability for the k$^{th}$ Test Instrument |

up from the UUT in the Traceability Chain.

**Uncertainty Estimation Example**

Consider a traceability chain consisting of a set of four test instruments each measuring 10 volts as shown in Table 1.

| Level | Unc (95% EOP) | EOP Std Unc | Engineering Tolerance | Obs Rel |
|---|---|---|---|---|
| 1 | 2.806 | 1.403 | | |
| 2 | 11.225 | 5.613 | 6.250 | 0.7200 |
| 3 | 45.541 | 22.771 | 25.000 | 0.7200 |
| 4 | 182.006 | 91.003 | 100.000 | 0.7200 |

**Table 1. Uncertainty Estimation Using Calibration Results**

The test instruments in the second, third, and fourth levels are part of a tolerance testing calibration process. The test tolerances (two sided) are 6.25 ppm, 25 ppm, and 100 ppm, respectively.

The test instrument at the first level is a part of an uncertainty testing process. The uncertainty, given as 2.806 ppm, is assumed to be 2 standard deviations. This means that the standard uncertainty is 1.403 ppm.

The standard uncertainties for the test instruments were calculated using the traceability uncertainty mathematics shown in equation (21). For a 72% End of Period reliability target, a tolerance of 100 ppm corresponds to a 95% uncertainty of 182.006 ppm.

**Approximate Uncertainty Estimates Using The TAR and the Measurement Reliability**

The Test Accuracy Ratio, or TAR, is defined as the ratio of the UUT engineering tolerance to the calibrator engineering tolerance, or:

$$TAR = \frac{UUT\ Engineering\ Tolerance}{Calibrator\ Engineering\ Tolerance}$$

For example, if a voltmeter was tested to be within ± 0.01 volts of the calibration standard, and the calibration standard being used to test the voltmeter had tolerances of ± 0.0025, the Test Accuracy Ratio would be:

$$TAR = \frac{0.01}{0.0025} = 4$$

Generally, this is expressed as “4:1”. 4:1 is the usual minimum US Navy technical requirement.

When the traceability chain has a consistent TAR, and the observed reliability is the same at every level, a close approximation for the standard uncertainty is given by:

$$u(UUT) \cong (UUT\,Tolerance\,Limit)\sqrt{\left[\frac{1}{Z_{\left(\frac{1+R_o(UUT)}{2}\right)}}\right]^2 - \left[\frac{1}{Z_{\left(\frac{1+R_{o,aop}(Calibrator)}{2}\right)}}\right]^2 \left(\frac{1}{TAR^2+1}\right)} \quad (22)$$

This equation can be solved to find the multiplier of the UUT uncertainty which gives the tolerance limit. The multiplier is given by:

$$Uncertainty\ Multiplier \cong \left(\left[\frac{1}{Z_{\left(\frac{1+R_o(UUT)}{2}\right)}}\right]^2 - \left[\frac{1}{Z_{\left(\frac{1+R_{o,aop}(Calibrator)}{2}\right)}}\right]^2 \left(\frac{1}{TAR^2+1}\right)\right)^{-1/2} \quad (23)$$

With a 4:1 TAR, the approximate uncertainty multiplier is 1.099 if the observed reliability target is 72% (General Purpose Test Equipment or GPTE). The approximate uncertainty multiplier for Special Purpose Test Equipment (SPTE) with the same TAR and an observed reliability target of 85% is 1.484.

**Four Necessary Elements For The Expression of Measurement Uncertainty**

A description of measurement uncertainty requires at least four elements:

1. The expanded ***uncertainty*** value usually expressed as ±*Uncertainty*
2. The ***multiplier*** used to obtain the expanded uncertainty from the standard uncertainty
3. The ***confidence*** associated with the uncertainty
4. A ***time*** limit that tells how long the uncertainty is valid

As an example, at the end of a 1 year calibration interval, the voltmeter at Level 4 in Table 1 has a standard uncertainty of 91.003 ppm. A simple expanded uncertainty would be given by twice the projected measurement standard deviation. Therefore, a correct and complete expression of the measurement uncertainty for this voltmeter would be:

1. ± 182.006 ppm
2. The expanded uncertainty is 2 times the standard uncertainty (91.003 * 2)
3. 95% Confidence (Actually 95.45%, since 2 is an approximation)
4. Valid for 1 year

A Navy analyst could also express the measurement uncertainty in terms of engineering tolerances and measurement reliability. This would be an approximation to the correct uncertainty values. This approximate expression for the measurement uncertainty for this voltmeter would be given by:

1. ± 100 ppm
2. The expanded uncertainty is 1.099 times the standard uncertainty (91.003 * 1.099)
3. 72% Confidence (Actually 72.82%, since measurement reliability is an approximation)
4. Valid for 1 year

The coverage probability, or confidence, must be included for the measurement uncertainty to have any usefulness. Without this figure, it is impossible to ascribe any meaning to the uncertainty value. The coverage probability predicts the proportion of measurements that would be expected to be within the expanded uncertainty of the true measurand value. For example, in the voltmeter above, a measurement would be within 182.006 ppm of the true voltage 95% of the time, or equivalently, within 100 ppm of the true voltage 72% of the time.

An expression of measurement uncertainty that is only valid immediately after calibration is misleading, because it is almost immediately invalid. For this reason, a time period must be associated with an uncertainty, and the uncertainty should apply to the end of the time period.

Nearly all uncertainties degrade with time and usage. It is for this very reason that periodic calibration is needed to maintain an acceptable level of uncertainty. The time limit provides an assurance that the uncertainty has not grown to the point that the test instrument cannot be used for a given application.

**Calibration Risk Analysis**

Usually, we think of calibration as testing only the error of the UUT. However, the comparison between the UUT and the calibrator contains error from both the UUT and the calibrator. This can be seen by deriving an expression for the comparison in terms of the measurement errors.

The measurement model for the Calibrator is:

$$\textit{Calibrator Measurement} = \textit{True Value} + \textit{Calibrator Error} \tag{24}$$

The measurement model for the UUT is:

$$\textit{UUT Measurement} = \textit{True Value} + \textit{UUT Error} \tag{25}$$

Using the measurement models from (24) and (25), it can be seen that the comparison is actually a comparison of the two errors:

$$\begin{aligned} \textit{Comparison} &= \textit{True Value} + \textit{UUT Error} - \\ &\qquad (\textit{True Value} + \textit{Calibrator Error}) \\ &= \textit{UUT Error} - \textit{Calibrator Error} \end{aligned} \tag{26}$$

Several important facts can be directly derived from this:

- Tolerance testing does not directly test the error of the UUT, rather it makes a test on the difference between the UUT and the calibrator errors.

- There is a risk that a UUT that is determined to be In Tolerance, will actually be Out of Tolerance. This is called the Probability of a False Accept (PFA).

- There is also a risk that a UUT that is determined to be Out of Tolerance, will actually be In Tolerance. This is called the Probability of a False Reject (PFR).

- Because there is a False Accept probability (PFA), and because the calibrator itself could be out of tolerance with respect to the UUT, there is a risk that the calibrated UUT could still be out of tolerance.

A ***false accept*** decision is shown in Figure 5.

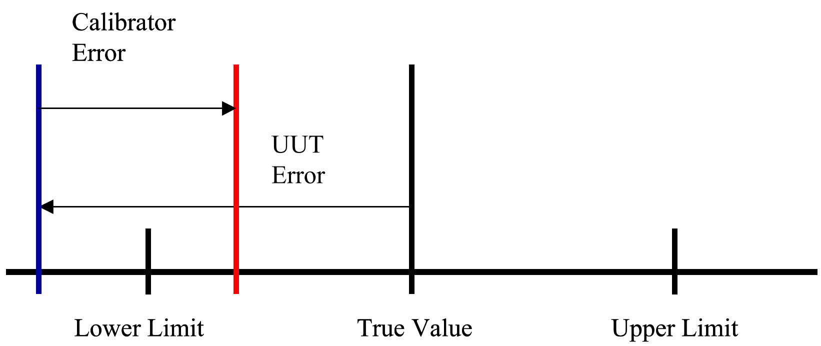


**Figure 5. False Acceptance**

In a false accept, the UUT really is Out of Tolerance, but the Calibrator Error brings it back In Tolerance. The probability of a false accept (PFA) is given by:

$$PFA = \Pr(\textit{UUT Out of Tolerance given Comparison In Tolerance}) \tag{27}$$

The evaluation of this and the other risk probabilities discussed in this section requires the use of the Bivariate Normal distribution function which is discussed in the Appendix.

A ***false reject*** decision is shown in Figure 6.

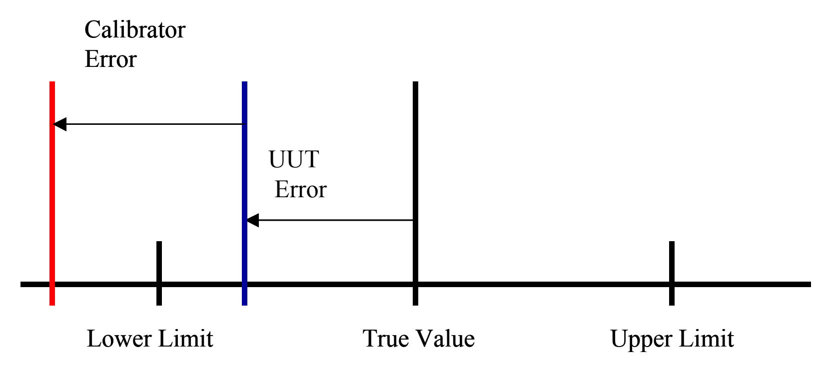


**Figure 6. False Rejection**

In a false reject, the UUT is really In Tolerance, but the Calibrator Error moves the comparison Out of Tolerance. The probability of a false reject (PFR) is given by:

$$PFR = \Pr(\textit{UUT In Tolerance given Comparison Out of Tolerance}) \tag{28}$$

For Navy calibrations, the UUT is only adjusted when it is Out of Tolerance. This means that false accepts will be released in an Out of Tolerance condition. In addition, if the calibrator is Out of Tolerance with respect to the UUT tolerance limits, then the UUT will be Out of Tolerance after adjustment, as well. The probability for an out of tolerance at the beginning of the use period (BOP) when the adjustment is only made when the UUT is out of tolerance (which corresponds to U.S. Navy policy) is given by:

$$\begin{aligned} BOPN &= Pr(\textit{UUT Out of Tolerance at BOP with Navy adjustment policy}) \\ &= Pr(\,(\textit{UUT Out of Tolerance and Comparison In Tolerance}) \textit{ or} \\ &\quad (\textit{Calibrator Out of Tolerance and Comparison Out of Tolerance})) \end{aligned} \tag{29}$$

**Calibration Risk Example**

In Table 2, we expand the uncertainty results given in Table 1 to show calibration risks. Table 2 shows the same traceability chain consisting of a set of four test instruments each measuring 10 volts. The uncertainties are expressed in parts per million (ppm).

| Level | Unc (95% EOP) | AOP Std Unc | EOP Std Unc | Engineering Tolerance | TAR | Obs Rel | PFA | PFR | BOP Adj OOT |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 2.806 | 1.403 | 1.403 | | | | | | |
| 2 | 11.225 | 4.125 | 5.613 | 6.250 | | 0.7200 | 0.050 | 0.179 | 0.036 |
| 3 | 45.541 | 16.934 | 22.771 | 25.000 | 4.000 | 0.7200 | 0.039 | 0.127 | 0.028 |
| 4 | 182.006 | | 91.003 | 100.000 | 4.000 | 0.7200 | 0.039 | 0.131 | 0.028 |

**Table 2. Tolerance Testing Calibration Chain Results**

For a 4:1 test uncertainty ratio, the risks of false decisions are not negligible.  Of the 72% of the In Tolerance test results, approximately 3.9% will actually be Out Of Tolerance.  This is the PFA number, or Probability of False Accept.  For 100 tests, this would mean that 72 would be found In Tolerance (on the average), and approximately 3 (0.039 x 72 = 2.8) would actually be Out of Tolerance.

The Probability of a False Reject (PFR) is even larger.  Of the 28% that would be found to be Out of Tolerance, approximately 13% would actually be In Tolerance.  Out of 100 tests, this would correspond to 28 Out of Tolerance decisions, of which approximately 4 (0.13 x 28 = 3.64) would actually be In Tolerance.

The bottom line for the calibration customer would be his risk of having an Out of Tolerance instrument after the calibration process.  With a 4:1 Test Accuracy Ratio, and an observed reliability of 72%, the customer has an almost 3% (2.8%) risk that his test instrument will be initially (BOP, or Beginning of Period) Out of Tolerance.  This means that out of a 100 items returned to the customer, approximately 3 would be out of tolerance.

**Calibration Risk and TAR**

Calibration risk can be made appropriately small by ensuring that the Test Accuracy Ratio between the calibration standard and the test equipment is at least 4:1.  The relationship between TAR levels and risk for General Purpose Test Equipment (GPTE) is shown in Figures 7 and 8.

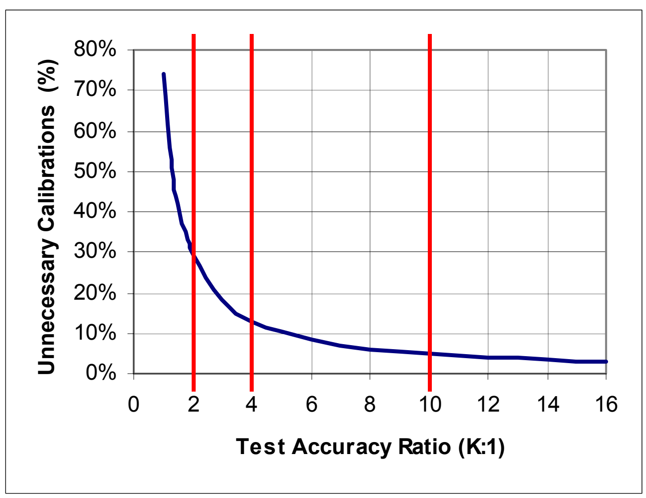


**Figure 7.  Unnecessary Calibration Risk (GPTE)**

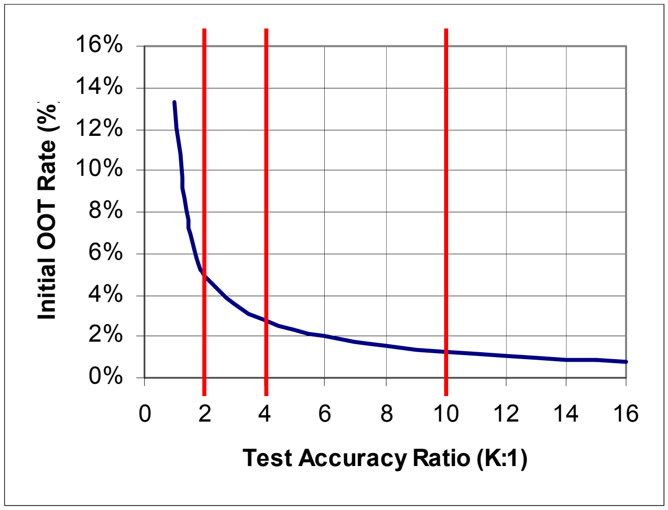


**Figure 8. Initial OOT Risk (GPTE)**

As can be seen in Figure 7, having a low TAR between the calibration standard and the test equipment will cause a lot of unnecessary calibration (adjustment). However, the larger risk is in sending the Navy out-of-tolerance test equipment. With a 2:1 TAR, it can be seen in Figure 8 that nearly 5% of the test equipment will be initially out-of-tolerance. The required 4:1 TAR improves this to less than 3%, while a 10:1 TAR drops the risk to nearly 1%.

The maintenance of low test equipment tolerances is dependent on two aspects of the METCAL program:

- Calibration intervals to ensure sufficient periodic calibration.
- Calibration standards that are traceable to International measurement units.

The traceability of the calibration standards is achieved through periodic calibration of the standards by higher level standards. Traceability assures that measurements made by any test equipment are comparable to any others. At each level, the TAR is used to ensure that the risk of making wrong decisions is minimized.

**The Role of Test Equipment**

US Navy test equipment is used to determine if systems will perform their function as designed. For example, the MK 612 Missile Test Set measures various characteristics of the STANDARD Missile system. If a section of the missile is found to be outside of specifications for a given characteristic, the faulty section is removed from the missile and returned to a depot for further testing and repair. Though this is expensive, it is much less expensive than leaving a ship vulnerable to attack because its defensive weapons won't work.

### Basics of In-Service Reliability

A single test of a US Navy system does not assure that it will perform reliably forever. If we define reliability as the probability that a system will perform its mission as designed, there is an unfortunate principle that seems to apply to almost all systems:

**Principle**: *Reliability degrades over time.*

This is demonstrated graphically in Figure 9. This shows a reliability curve for a system that starts with 95% reliability at the beginning of a deployment. After 10 years, this reliability drops to 60%. If this curve referred to missiles, this would mean that only 6 out of 10 missiles would perform their mission after 10 years.

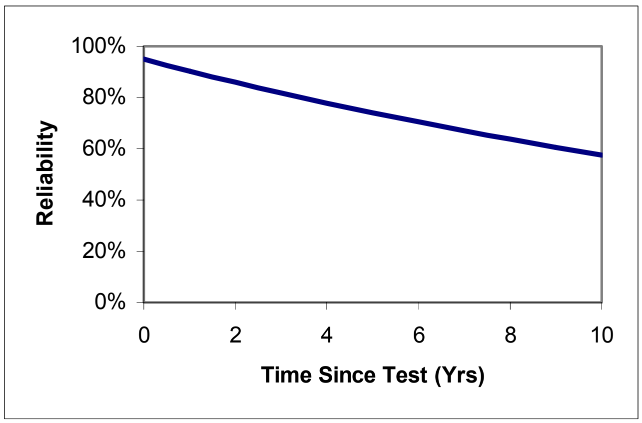


**Figure 9. Example Reliability Curve For a Military System**

Periodic testing is used to cull the unreliable systems from the inventory. This has the effect of returning the reliability to the 95% level in the example above. If you don't test periodically, you rapidly reach a condition where things just don't work anymore. "Things not working" can be inconvenient for folks at home, but it's downright disastrous for personnel in combat.

### What Happens With Bad Test Decisions?

It may come as a surprise that the reliability curve in Figure 9 did not start at 100%. After all, when a system has just been tested, it has 100% reliability, right? If we had perfect tests, that would be the case.

Unfortunately, testing is dependent on measurements, and measurements generally contain error. For example, a voltmeter measuring an exact 10 volt source, might yield a measurement of 10.02 volts. This would mean that the error in the measurement was 0.02 volts.

If this measurement were applied to a weapon system that was required to be between 9.99 and 10.01 volts, we would come to the erroneous conclusion that the system was out of specification.

Our faulty measurement would mean that the weapon system would be sent in for repair needlessly.

In a recent reliability study on STANDARD Missile, the results of testing were regularly found to be in question. Failed sections from a MK 612 test were retested at the depot. It was often found that the failure was not repeatable. Such failures are costly necessitating the breakdown and rebuild of missile systems as well as needless transportation of sections to and from the depot. False failures also mean fewer missiles are available for load out on ships to support US Navy missions.

On the other hand, if this measurement were applied to a weapon system that was required to be between 10.01 and 10.03 volts, we would incorrectly pass the system. This would mean that a faulty weapon system would be sent to the Fleet. Such an error can result in loss of mission, or even loss of lives.

Mistakenly sending bad items out to the Fleet is one of the causes of system reliability curves starting below 100% at the beginning of a deployment as shown in Figure 10. With a perfect test, the initial reliability would indeed be 100%. Unfortunately, weapon tests cannot test every feature. For example, the warhead and the rocket motor are not tested on a missile test set for obvious reasons. Since these items are not tested, this results in the partial test reliability curve which starts at 95%. When you take into account that the reliability curve comes from a partial test that is done with uncertainty, the reliability curve is lowered even farther as shown by the "Partial Test w/ Uncertainty" reliability curve.

This lower initial reliability has two big effects on the Fleet. Combat and other systems have a required level of reliability shown by the horizontal blue line in Figure 10. The place where the reliability curve meets the reliability requirement determines the maintenance cycle or certification interval for the system. A lower initial reliability means that the system must be tested more often which dramatically increases program office budgets.

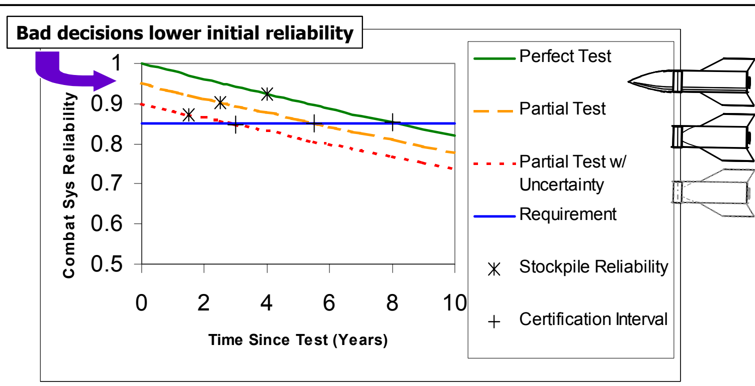


**Figure 10. Initial Reliability Lowered By Measurement Uncertainty**

In addition, the lower initial reliability translates to a lower level of reliability for the stockpile of systems. This means that the average reliability for deployed systems is lowered. In combat situations, this translates to failed missions and fatalities: Things just don't work when you need them to.

**How Do You Ensure A Good Test?**

How do you make sure that things do work when you need them to? One obvious answer is to make sure that testing does what it's supposed to by correctly identifying systems that need to be repaired or maintained, and by not mistakenly flagging systems that don't need to be. This is the entire objective of the US Navy METCAL Program. We do this by:

- Helping to ensure that the correct functions are tested.
- Providing for the correct equipment and measurement procedures.
- Providing for the needed Test Accuracy Ratio (TAR).
- Providing for periodic calibration of test equipment.
- Providing for measurement traceability to national standards.

The Test Accuracy Ratio for systems is defined as the ratio of the system specification tolerances to the test equipment tolerances, or:

$$\text{TAR} = \frac{\text{System Tolerance}}{\text{Test Equipment Tolerance}}$$

For example, if the voltage output for a characteristic of a weapon system was required to be within 0.01 volts of a nominal value, and the voltmeter being used to test the weapon system had tolerances of $\pm$ 0.0025, the Test Accuracy Ratio would be:

$$\text{TAR} = \frac{0.01}{0.0025} = 4$$

For system test accuracy ratios, 4:1 is the usual minimum US Navy technical requirement.

**System Risk Analysis**

The risk to a Navy system is found using the same mathematics as is used in finding the risk for a piece of test equipment. The major difference is that the bottom level of the traceability chain is a Navy system rather than a piece of test equipment. The mathematical details for risk calculation are described in the Appendix.

For a false accept, the tested system is really Out of Tolerance, but the Test Equipment Error makes it look as though it is In Tolerance. The probability of a false accept (PFA) is given by:

$$PFA = \Pr(\textit{System Out of Tolerance given Comparison In Tolerance}) \tag{30}$$

For a false reject, the system is really In Tolerance, but the Test Equipment Error makes it look as though it is Out of Tolerance. The probability of a false reject (PFR) is given by:

$$PFA = \Pr(\textit{System In Tolerance given Comparison Out of Tolerance}) \quad (31)$$

The BOPN term applies to calibrated test equipment that are only adjusted when they are Out of Tolerance. This risk probability would apply to Navy systems that are adjusted using the test equipment only when they are found to be Out of Tolerance. This probability is given by:

$$BOPN = Pr(\,(\textit{System Out of Tolerance and Comparison In Tolerance})\ \textit{or}$$

$$(\textit{Test Equipment Out of Tolerance and Comparison Out of Tolerance})) \quad (32)$$

Some Navy systems are always adjusted using the test equipment regardless of the tolerance condition. The probability that these adjusted systems will be Out of Tolerance at the beginning of their maintenance periods (BOP) is given by:

$$BOPA = \Pr(\textit{Test Equipment Out of Tolerance}) \quad (33)$$

For many Navy systems, no adjustment is attempted and Out of Tolerance systems are removed from the inventory. The testing of most missile systems is an example of such a process. The missile sections cannot be adjusted using the test equipment, so Out of Tolerance sections are sent to depots for repair. The probability that there are Out of Tolerance systems at the Beginning of the Maintenance Period (BOP) when there is no adjustment and Out of Tolerance systems are removed from the inventory is given by:

$$BOPR = Pr(\textit{System Out of Tolerance given Comparison In Tolerance}) \quad (34)$$

An example of risk calculations for a Navy system is given in Table 3.

| Level | Unc (95% EOP) | AOP Std Unc | EOP Std Unc | Engineering Tolerance | TAR | Obs Rel | PFA | PFR | BOPR |
|---|---|---|---|---|---|---|---|---|---|
| Std | 0.02810 | 0.01405 | 0.01405 | | | | | | |
| Cal | 0.11224 | 0.04125 | 0.05612 | 0.06250 | | 0.7200 | 0.050 | 0.180 | |
| TE | 0.33740 | 0.13575 | 0.16870 | 0.25000 | 4.000 | 0.8500 | 0.024 | 0.216 | |
| Sys | 1.53681 | | 0.76841 | 1.00000 | 4.000 | 0.8000 | 0.026 | 0.139 | 0.026 |

**Table 3. Navy System Risk Analysis When OOT's Are Removed**

Table 3 could apply to a missile system which had a target reliability of 80% at the end of its certification interval. Such a system would be tested using a Special Purpose Test Equipment (SPTE) which would have a target measurement reliability of 85% at the end of the calibration interval (EOP). The calibrator for the test equipment would be a General Purpose Test Equipment (GPTE) which would have a 72% EOP measurement reliability target. It is also assumed that 4:1 test accuracy ratios are obtained at every level of the traceability chain.

With these assumptions, the missile system would begin its maintenance cycle with a risk of 2.6% of being out of tolerance. This roughly corresponds to a 97.4% initial reliability.

## TAR and System Risk

What does a specific Test Accuracy Ratio buy you? The TAR can be related to the risk of making wrong decisions during system testing as shown in Figures 11 and 12. These figures apply to systems that are observed to be approximately 80% reliable at the time of their periodic test. They also assume the test equipment were Special Purpose (SPTE) which have an 85% EOP reliability target which themselves were calibrated by standards that have a 72% EOP reliability target.

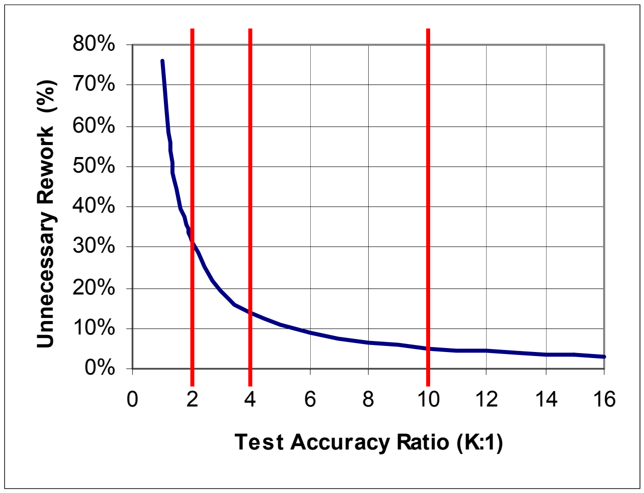


**Figure 11. Unnecessary Rework (False Rejects)**

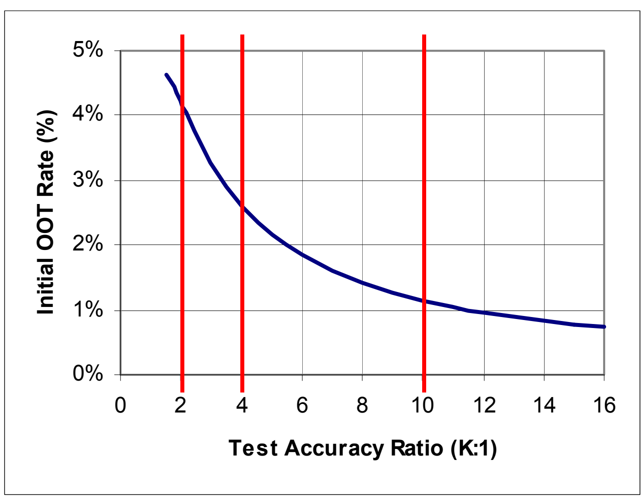


**Figure 12. Initial OOT Rate**

With a 4:1 TAR, approximately 14% of the test failures will be due to the test equipment rather than the tested system. This means that 14% of the repair and maintenance work done as a result of the testing will be unnecessary. By increasing the TAR to 10:1, we can decrease the unnecessary rework to as little as 5%. In cases where the TAR is 2:1 or lower, the percentage of unnecessary rework skyrockets to 31% and more.

The cost for maintenance has a tremendous impact on Program Office budgets. Low TARs not only increase unnecessary rework, they also decrease the observed reliability, which shortens the length of the maintenance cycle. The METCAL program can have a large effect on minimizing unnecessary expenditures.

Figure 12 shows the percentage of tested systems that will be returned to the Fleet in an Out-Of-Tolerance (OOT), or faulty condition. For a 4:1 TAR, 2.6% of the tested systems will still be out of specification. A 2:1 TAR will increase the initial OOT rate to over 4%, while a 10:1 TAR decreases this rate to almost 1%. This initial OOT rate translates directly to reduced reliability in deployed systems.

## Summary

In summary, the concepts of measurement uncertainty and uncertainty confidence level are closely related to the US Navy concepts of engineering tolerances and measurement reliability. Though they are not interchangeable, this paper provides mathematical relationships between the two.

Measurement reliability is a necessary concept for the development of calibration intervals. And strictly speaking, calibration intervals give validity to expressions of measurement uncertainty. Because measurement uncertainty degrades with time, an expression of uncertainty should be projected to the end of a calibration interval. This ensures that the uncertainty is valid during the entire interval period.

The measurement uncertainties of the calibrator and a piece of test equipment provide the means to determine the risk that the test equipment is Out of Tolerance. These uncertainties applied to the engineering tolerances for a US Navy system provide the means to determine risk that the system is Out of Tolerance. The risk to the end system is probably the most important measure of the success of a calibration system.

## References

1. Jackson, D., and Castrup, H., 2001, “Uncertainty Propagation Theory and Derivations”, NSWC Corona Division

2. Jackson, D., and Castrup, H., 2001, “Uncertainty Propagation in Calibration Support Hierarchies”, 2002 NCSLI Proceedings

## Appendix: Risk Theory and Derivations

### Introduction

This appendix develops the theory for measurement traceability mathematics. The notation used here is meant to be rigorous and a familiarity with mathematical statistics is assumed.

The basis of the traceability mathematics is a measurement error model. This is derived and assumptions are made concerning the relevant error sources. Though many of the results are independent of distribution assumptions, the normal distribution is assumed in order to provide a framework for some of the results.

### Measurement Error Model

The measurement model, shown in Figure A.1, relates the value of the quantity being measured, referred to as the ***measurand***, to the value obtained as a result of measurement. The difference between the two values is called the ***measurement error***. This will be expressed in mathematical form as:

$$y(t) = Y + e(t) \tag{A.1}$$

where:

| | | |
|---|---|---|
| $t$ | = | The time since calibration for the measurement equipment. |
| $Y$ | = | The (true) value of the measurand. |
| $y(t)$ | = | The result of the measurement on Y. |
| $e(t)$ | = | The measurement error. |

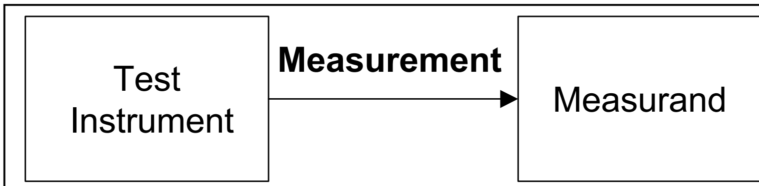


**Figure A.1. Measurement Model**

### Measurements in a Traceability Chain

Calibration standards or measuring and test equipment (MTE) of each echelon in a measurement support hierarchy are submitted for test or calibration to the next highest echelon in the hierarchy. Figure A.2 shows a traceability chain consisting of *i* test instruments.

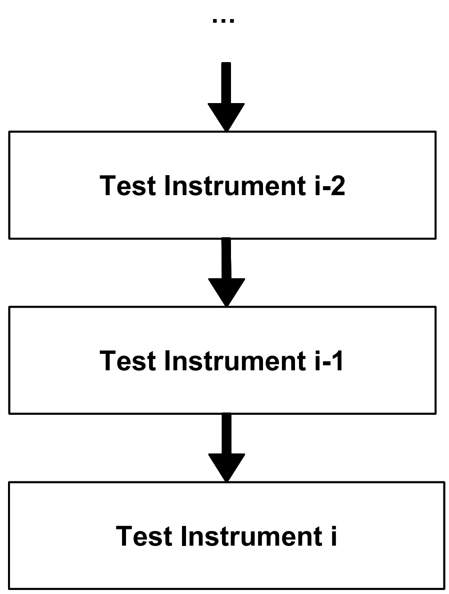


**Figure A.2. Traceability Chain**

In a calibration scenario, there is a calibrator and a unit under test, or UUT. The UUT will be given the subscript *i*, where *i* could take values from 1 to N. The calibrator, being the next test instrument up in the chain will generally be given the subscript *i-1*. The measurement model for the UUT will be given as:

$$y_i(t_i) = Y + e_i(t_i) \tag{A.2}$$

where:

| | | |
|---|---|---|
| $i$ | = | The $i^{th}$ level in the traceability chain (i = 0,1,2,…,N). |
| $t_i$ | = | The time since calibration for the $i^{th}$ test instrument in the traceability chain. |
| $Y$ | = | The (true) value of the measurand. |
| $y_i(t_i)$ | = | The result of the measurement on Y by the $i^{th}$ test instrument in the traceability chain. |
| $e_i(t_i)$ | = | The measurement error for the $i^{th}$ test instrument in the traceability chain. |


**Figure A.3. Subscripted Measurement Model**

The result of a measurement on $Y$ by the calibrator, shown in Figure A.4, is $y_{i-1}(t_{i-1})$ which is modeled by:

$$y_{i-1}(t_{i-1}) = Y + e_{i-1}(t_{i-1}) \tag{A.3}$$

where:

| | | |
|---|---|---|
| $t_{i-1}$ | = | The time since calibration for the $i\text{-}1^{st}$ test instrument in the traceability chain when calibrating the $i^{th}$ test instrument in the traceability chain. |
| $t_{i-2}$ | = | The time since calibration for the $i\text{-}2^{nd}$ test instrument in the traceability chain when calibrating the $i\text{-}1^{st}$ test instrument in the traceability chain. |
| $t_{i-k}$ | = | The time since calibration for the $i\text{-}k^{th}$ test instrument in the traceability chain when calibrating the $i\text{-}k\text{+}1^{st}$ test instrument in the traceability chain. |

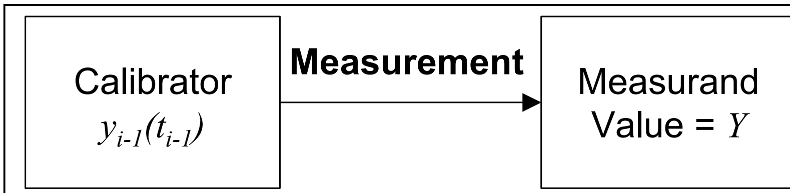


**Figure A.4. Calibrator Measurement Model**

**Statistical Assumptions**

Though many of the results developed in this paper will be general in nature, certain statistical assumptions will be useful for the development of probability and uncertainty results. The errors will be assumed to follow the normal distribution. The notation used to indicate that $x$ follows the normal distribution with mean $\mu$ and variance $\sigma^2$ is as follows:

$$x \sim N(\mu, \sigma^2)$$

This means that the probability distribution function for $x$ is given by the bell shaped curve or:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

The error component of the measurement error model given in (A.2) is assumed to be distributed as:

$$e_i(t_i) \sim N(\mu_i(t_i), \sigma_i^2(t_i)) \qquad \text{(A.4)}$$

The standard uncertainty of the measurement is the same as the error standard deviation, $\sigma_i(t_i)$. For the test instrument at the $i^{th}$ level in the traceability chain, the uncertainty will also be represented by:

$$u_i(t_i) = \sigma_i(t_i) \qquad \text{(A.5)}$$

The inclusion of $\mu_i(t_i)$ in the measurement error distribution means that there can be a drift or bias from the true value in the measurement.

It should be noted that the assumptions discussed in this section refer to the distribution of the errors in measurement rather than the distribution of the measurements themselves. The reason for this is that the true value of the measurand, $Y$, is not a fixed value, generally. During a given calibration session, the value of $Y$ will probably fluctuate. The conditional distribution of $y_i(t_i)$ for a given value of $Y$ (denoted by $y_i(t_i) \mid Y$) is given by:

$$y_i(t_i) \mid Y \sim N(Y + \mu_i(t_i), \sigma_i^2(t_i)) \qquad \text{(A.6)}$$

In most calibration setups, the actual value of $Y$ is irrelevant, since it is lost in the difference between the calibrator and the UUT, or:

$$y_i(t_i) - y_{i-1}(t_{i-1}) = e_i(t_i) - e_{i-1}(t_{i-1}) \qquad \text{(A.7)}$$

However, it should be realized that equation (A.7) is only valid if the UUT and calibrator measurements are approximately simultaneous. Simultaneity is a feature of most calibration measurement scenarios, with the possible exception of an external unknown artifact.

**Tolerance Testing**

In a tolerance test, the difference between the calibrator and the UUT is compared with a set of tolerance limits.

In general, the tolerance test should be performed on a single measurement, rather than a sample. The tolerances are meant to represent the uncertainty of the UUT, which includes the random error component. Using a sample of measurements would artificially reduce the uncertainty of the UUT during the test, without reducing the actual uncertainty of the UUT.

The comparison for a tolerance test is given by:

$$d_i = y_i(t_i) - y_{i-1}(t_{i-1}) \qquad \text{(A.8)}$$

The tolerance test that will be developed in the following sections will be the ***two-sided symmetric*** test. In this scenario, a UUT is considered to be in tolerance if:

$$-L_i \leq d_i \leq L_i \tag{A.9}$$

Simple extensions to this development can be made to fit the single-sided tests.

**Uncertainty Based on Tolerance Testing Results**

Though the uncertainty of a UUT under a tolerance test is not established parametrically, the uncertainty can be found using the reliability objective and the calibration interval. Measurement reliability is defined as the probability that a test instrument is observed in tolerance. This will be given by $R_{o,i}(t_i)$ which is expressed mathematically as:

$$R_{o,i}(t_i) = Pr(-L_i \leq d_i \leq L_i) \tag{A.10}$$

If we expand the comparison variable $d_i$ using the measurement model in (A.2), we have:

$$d_i = y_i(t_i) - y_{i-1}(t_{i-1}) = e_i(t_i) - e_{i-1}(t_{i-1}) \tag{A.11}$$

From the statistical assumptions in (A.4),

$$e_i(t) \sim N(\mu_i(t_i), \sigma_i^2(t_i)) \tag{A.12}$$

However, since there is very little parametric data recorded for tolerance testing, it will be very difficult to explicitly establish the expectation trend for $\mu_i(t_i)$. Without much loss of generality, we will assume for the tolerance testing scenario:

$$e_i(t_i) \sim N(0, \sigma_i^2(t_i)) \tag{A.13}$$

This has the effect of assuming that the bias growth is entirely due to an increase in variability. Using this assumption, we have:

$$e_i(t_i) - e_{i-1}(t_{i-1}) \sim N(0, \sigma_i^2(t_i) + \sigma_{i-1}^2(t_{i-1})) \tag{A.14}$$

Therefore,

$$R_{o,i}(t_i) = Pr(-L_i \leq e_i(t_i) - e_{i-1}(t_{i-1}) \leq L_i)$$

$$= Pr\left(\frac{-L_i}{\sqrt{\sigma_i^2(t_i) + \sigma_{i-1}^2(t_{i-1})}} \leq \frac{e_i(t_i) - e_{i-1}(t_{i-1})}{\sqrt{\sigma_i^2(t_i) + \sigma_{i-1}^2(t_{i-1})}} \leq \frac{L_i}{\sqrt{\sigma_i^2(t_i) + \sigma_{i-1}^2(t_{i-1})}}\right)$$

$$= F\left(\frac{L_i}{\sqrt{\sigma_i^2(t_i) + \sigma_{i-1}^2(t_{i-1})}}\right) - F\left(\frac{-L_i}{\sqrt{\sigma_i^2(t_i) + \sigma_{i-1}^2(t_{i-1})}}\right)$$

$$= 2F\left(\frac{L_i}{\sqrt{\sigma_i^2(t_i) + \sigma_{i-1}^2(t_{i-1})}}\right) - 1 \tag{A.15}$$

where $F(z)$ is the cumulative Standard Normal distribution function given by

$$F(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

Solving (A.15) for the variance of the UUT gives:

$$\sigma_i^2(t_i) = \left[\frac{L_i}{F^{-1}\left(\frac{1 + R_{o,i}(t_i)}{2}\right)}\right]^2 - \sigma_{i-1}^2(t_{i-1}) \tag{A.16}$$

**Traceability Results for Tolerance Testing**

If there is a direct estimate of $\sigma_{i-1}^2(t_{i-1})$, then (A.16) can be used to obtain the uncertainty for the UUT. This would be the case if the $i\text{-}1^{st}$ level in the traceability chain uses an uncertainty testing calibration process.

Without expressing $\sigma_{i-1}^2(t_{i-1})$ as a specific function of $t_{i-1}$, (A.16) cannot be solved. However, it is possible to consider a slight modification of the measurement model for tolerance testing that will allow a solution that can apply to the traceability chain.

Rather than looking at the measurement errors at the end of the calibration interval for the tolerance testing calibrators in the traceability chain, we consider a measurement randomly chosen from the calibration interval. Thus, rather than using the end-of-period (EOP) reliability target, $R_{o,i-k}(t_{i-k})$, we use the average-over-period (AOP) reliability, $R_{o,i-k,aop}$ given by

$$R_{o,i\text{-}k,aop} = \int_0^{I_{i-k}} \frac{R_{o,i-k}(t)}{I_{i-k}} dt \tag{A.17}$$

where:

$R_{o,i-k}(t)$ = The measurement reliability as a function of the time since calibration as give by either (A.15) or the reliability function used to determine the calibration interval.

$I_{i-k}$ = The calibration interval for the $i\text{-}k^{th}$ level tolerance testing test instrument.

The following approximation can be used, if needed:

$$R_{o,i\text{-}k,aop} = \sqrt{R_{o,i-k,eop}} \tag{A.18}$$

where:

$R_{o,i\text{-}k,eop}$ = The end of period, EOP, measurement reliability target.

Applying (A.16) to the assumptions concerning prior testers:

$$\sigma_{i-k}^2 = \left[ \frac{L_i}{F^{-1}\left( \dfrac{1+R_{o,i-k,aop}}{2} \right)} \right]^2 - \sigma_{i-k-1}^2 \tag{A.19}$$

where:

$\sigma_{i-k}^2$ = The variance of a measurement from the *i-k*$^{th}$ Test Instrument taken at a random point during the calibration interval.

Equation (A.19) is a recursive relation and can be extended back to the first test instrument in the Traceability chain that is calibrated using the obtained measurement uncertainty methodology. Assuming that the first such test instrument is at the *i-m*$^{th}$ level, we obtain:

$$\sigma_i^2(t_i) = \left[ \frac{L_i}{F^{-1}\left( \dfrac{1+R_{o,i}(t_i)}{2} \right)} \right]^2 + \sum_{k=1}^{i-m-1} (-1)^k \left[ \frac{L_{i-k}}{F^{-1}\left( \dfrac{1+R_{o,i-k,aop}}{2} \right)} \right]^2 + (-1)^m \sigma_{i-m}^2 (t_{i-m}) \tag{A.20}$$

**Approximate Relationship Between TAR and Uncertainty**

Using equation (A.20), one can determine the relationship between various test accuracy ratios (TAR's) in the traceability chain, and the resulting UUT uncertainty. A useful approximation to this relationship can be found assuming that all calibrations in the calibration chain have the same TAR value, A. This assumption is expressed as:

$$L_i / L_{i-1} = A \tag{A.21}$$

Using this assumption, and the geometric series, we have:

$$\sigma_i^2(t_i) = \left[\frac{L_i}{F^{-1}\left(\frac{1+R_{o,i}(t_i)}{2}\right)}\right]^2 + \sum_{k=1}^{i-m-1}(-1)^k\left[\frac{L_{i-k}}{F^{-1}\left(\frac{1+R_{o,i-k,aop}}{2}\right)}\right]^2 + (-1)^m\sigma_{i-m}^2(t_{i-m})$$

$$\cong \left[\frac{L_i}{F^{-1}\left(\frac{1+R_{o,i}(t_i)}{2}\right)}\right]^2 + \sum_{k=1}^{\infty}(-1)^k\left[\frac{L_{i-k}}{F^{-1}\left(\frac{1+R_{o,i-k,aop}}{2}\right)}\right]^2$$

$$\cong \left[\frac{L_i}{F^{-1}\left(\frac{1+R_{o,i}(t_i)}{2}\right)}\right]^2 + \sum_{k=1}^{\infty}(-1)^k\left[\frac{L_i / A^k}{F^{-1}\left(\frac{1+R_{o,i-1,aop}}{2}\right)}\right]^2$$

$$\cong \left[\frac{L_i}{F^{-1}\left(\frac{1+R_{o,i}(t_i)}{2}\right)}\right]^2 - \left[\frac{L_i}{F^{-1}\left(\frac{1+R_{o,i-1,aop}}{2}\right)}\right]^2\left(\frac{1}{A^2+1}\right) \tag{A.22}$$

A two stage approximation is also valuable to allow a change in the TAR for the last calibration in the traceability chain. This approximation assumes the following TAR values:

$$L_i / L_{i-1} = A_1 \tag{A.23}$$

$$L_{i-k} / L_{i-k-1} = A_2 \quad for\ k \geq 1 \tag{A.24}$$

Using these assumptions, and the geometric series, we have:

$$\sigma_i^2(t_i) = \left[\frac{L_i}{F^{-1}\left(\frac{1+R_{o,i}(t_i)}{2}\right)}\right]^2 + \sum_{k=1}^{i-m-1}(-1)^k\left[\frac{L_{i-k}}{F^{-1}\left(\frac{1+R_{o,i-k,aop}}{2}\right)}\right]^2 + (-1)^m\sigma_{i-m}^2(t_{i-m})$$

$$\cong \left[\frac{L_i}{F^{-1}\left(\frac{1+R_{o,i}(t_i)}{2}\right)}\right]^2 - \left[\frac{L_i / A_1}{F^{-1}\left(\frac{1+R_{o,i}(t_i)}{2}\right)}\right]^2 + \sum_{k=2}^{\infty}(-1)^k\left[\frac{L_{i-k}}{F^{-1}\left(\frac{1+R_{o,i-k,aop}}{2}\right)}\right]^2$$

$$\cong \left[\frac{L_i}{F^{-1}\left(\frac{1+R_{o,i}(t_i)}{2}\right)}\right]^2 - \left[\frac{L_i / A_1}{F^{-1}\left(\frac{1+R_{o,i-1,aop}}{2}\right)}\right]^2 + \sum_{k=2}^{\infty}(-1)^k\left[\frac{(L_i / A_1)/ A_2^{k-1}}{F^{-1}\left(\frac{1+R_{o,i-2,aop}}{2}\right)}\right]^2$$

$$\cong \left[\frac{L_i}{F^{-1}\left(\frac{1+R_{o,i}(t_i)}{2}\right)}\right]^2 - \left[\frac{L_i / A_1}{F^{-1}\left(\frac{1+R_{o,i-1,aop}}{2}\right)}\right]^2 + \left[\frac{(L_i / A_1)}{F^{-1}\left(\frac{1+R_{o,i-2,aop}}{2}\right)}\right]^2\left(\frac{1}{A_2^2+1}\right) \quad (A.25)$$

**UUT Reliability for Tolerance Testing**

The observed measurement reliability, $R_{o,i}(t_i)$, of a UUT that is tolerance tested is not the probability that the UUT is in tolerance. Rather, it is the probability that the comparison, $\Delta_i$, between the UUT and the calibrator is in tolerance. This is an important difference because the uncertainty, or standard deviation, of the comparison is larger than the uncertainty of the UUT.

The observed reliability was shown in (A.15) to be:

$$R_{o,i}(t_i) = Pr(-L_i \leq e_i(t_i) - e_{i-1}(t_{i-1}) \leq L_i)$$

$$= 2F\left(\frac{L_i}{\sqrt{\sigma_i^2(t_i) + \sigma_{i-1}^2(t_{i-1})}}\right) - 1 \quad (A.26)$$

The probability that the UUT is in tolerance, or UUT reliability, is given by:

$$R_i(t_i) = Pr(Y - L_i \leq y_i(t_i) \leq Y + L_i)$$

$$= Pr( - L_i \leq e_i(t_i) \leq L_i)$$

$$= Pr\left(\frac{-L_i}{\sigma_i(t_i)} \leq \frac{e_i(t_i)}{\sigma_i(t_i)} \leq \frac{L_i}{\sigma_i(t_i)}\right)$$

$$= F\left(\frac{L_i}{\sigma_i(t_i)}\right) - F\left(\frac{-L_i}{\sigma_i(t_i)}\right)$$

$$= 2F\left(\frac{L_i}{\sigma_i(t_i)}\right) - 1 \tag{A.27}$$

**Risk Associated with Tolerance Testing**

Since tolerance testing is performed using calibrators that have uncertainty, there is a chance that a tolerance test will result in a wrong decision. The probability of a wrong decision is known as risk, and this risk is a key characteristic of tolerance testing policy.

There are three major types of wrong decisions that can be made during a tolerance test, or as the result of a tolerance test:

- ***False Accept*** – This means that the tolerance test finds the UUT to be in tolerance, when it is really out of tolerance.
- ***False Reject*** – This means that the tolerance test finds the UUT to be out of tolerance, when it is really in tolerance.
- ***Beginning of Period Out of Tolerance*** – This means that the UUT is still out of tolerance after the testing process at the beginning of the use period (BOP).

**Risk and The Bivariate Normal Distribution**

The expression of risk probabilities will be dependent on the use of the Bivariate Normal distribution function. This is given by:

$$G(h,k,\rho) = \int_{-\infty}^{k}\int_{-\infty}^{h} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{z_1^{\ 2} - 2\rho z_1 z_2 + z_2^{\ 2}}{2(1-\rho^2)}\right) dz_1 dz_2 \tag{A.28}$$

$$= \int_{-k}^{\infty}\int_{-h}^{\infty} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{z_1^{\ 2} - 2\rho z_1 z_2 + z_2^{\ 2}}{2(1-\rho^2)}\right) dz_1 dz_2 \tag{A.29}$$

It should be noted here, that in order to obtain the opposite quadrant probabilities, the necessary Bivariate Normal probability relationships are given as follows:

$$\int_{-\infty}^{k}\int_{h}^{\infty} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{z_1^{\ 2} - 2\rho z_1 z_2 + z_2^{\ 2}}{2(1-\rho^2)}\right) dz_1 dz_2 = F(-h) - G(-h, -k, \rho) \tag{A.30}$$

$$\int_{k}^{\infty}\int_{-\infty}^{h} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{z_1^{\ 2} - 2\rho z_1 z_2 + z_2^{\ 2}}{2(1-\rho^2)}\right) dz_1 dz_2 = F(h) - G(h, k, \rho) \tag{A.31}$$

The numerical computation of this integral was accomplished using 20 point Gaussian Quadrature and the following equality developed by Drezner and Wesolowsky (1990):

$$G(h,k,\rho)=\int_0^{\rho}\frac{1}{2\pi\sqrt{1-r^2}}\exp\left(-\frac{h^2-2rhk+k^2}{2(1-r^2)}\right)dr+F(h)F(k) \qquad \text{(A.32)}$$

**False Accept Risk**

The probability of a false accept for the $i^{th}$ test instrument in the traceability chain, or $PFA_i$, is given by:

$$PFA_i = \Pr(\mathit{UUT\ Out\ Of\ Tolerance\ given\ the\ Comparison\ In\ Tolerance})$$

$$= \Pr(|e_i(t_i)| \geq L_i \text{ given } -L_i \leq d_i \leq L_i)$$

$$= \Pr(|e_i(t_i)| \geq L_i \text{ given } -L_i \leq e_i(t_i) - e_{i-1}(t_i) \leq L_i)$$

$$= \frac{\Pr(|e_i(t_i)| \geq L_i \text{ and } -L_i \leq e_i(t_i) - e_{i-1}(t_{i-1}) \leq L_i)}{\Pr(-L_i \leq e_i(t_i) - e_{i-1}(t_{i-1}) \leq L_i)}$$

$$= \frac{\Pr(|e_i(t_i)| \geq L_i \text{ and } -L_i \leq e_i(t_i) - e_{i-1}(t_{i-1}) \leq L_i)}{R_{o,i}(t_i)}$$

$$= \frac{2\left[G\left(\frac{-L_i}{\sigma_i},\frac{L_i}{\sigma_d},\rho_i\right)-G\left(\frac{-L_i}{\sigma_i},\frac{-L_i}{\sigma_d},\rho_i\right)\right]}{R_{o,i}(t_i)} \qquad \text{(A.33)}$$

where:

$$\sigma_i^2=\sigma_i^2(t_i) \qquad \text{(A.34)}$$

$$\sigma_d^2 = \sigma_i^2(t_i)+\sigma_{i-1}^2(t_{i-1}) = \left[\frac{L_i}{F^{-1}\left(\frac{1+R_{oi}(t_i)}{2}\right)}\right]^2 \qquad \text{(A.35)}$$

$$\rho_i = \mathrm{Cor}(e_i(t_i),\ e_i(t_i) - e_{i-1}(t_{i-1}))$$

$$= \frac{\sigma_i(t_i)}{\sqrt{\sigma_i^2(t_i)+\sigma_{i-1}^2(t_{i-1})}} \qquad \text{(A.36)}$$

**False Reject Risk**

The probability of a false reject for the $i^{th}$ test instrument in the traceability chain, or $PFR_i$, is given by:

$PFR_i = Pr(UUT\ In\ Tolerance\ given\ the\ Comparison\ Out\ Of\ Tolerance)$

$= \Pr(-L_i \le e_i(t_i) \le L_i\ given\ |d_i| \ge L_i)$

$= \Pr(-L_i \le e_i(t_i) \le L_i\ given\ |e_i(t_i) - e_{i-1}(t_i)| \ge L_i)$

$$= \frac{\Pr(-L_i \le e_i(t_i) \le L_i\ and\ |e_i(t_i) - e_{i-1}(t_{i-1})| \ge L_i)}{\Pr(|e_i(t_i) - e_{i-1}(t_{i-1})| \ge L_i)}$$

$$= \frac{\Pr(-L_i \le e_i(t_i) \le L_i\ and\ |e_i(t_i) - e_{i-1}(t_{i-1})| \ge L_i)}{1 - R_{o,i}(t_i)}$$

$$= \frac{2\left[G\left(\frac{L_i}{\sigma_i}, \frac{-L_i}{\sigma_d}, \rho_i\right) - G\left(\frac{-L_i}{\sigma_i}, \frac{-L_i}{\sigma_d}, \rho_i\right)\right]}{1 - R_{o,i}(t_i)} \tag{A.37}$$

**Beginning of Period Risk**

The probability that the UUT is still out of tolerance after the testing process depends on the adjustment policy. The probability for an out of tolerance at the beginning of the use period (BOP) when the adjustment is only made when the UUT is out of tolerance (which corresponds to U.S. Navy policy) is given by:

$BOPN_i = Pr(UUT\ Out\ of\ Tolerance\ at\ BOP\ with\ Navy\ adjustment\ policy)$

*=Pr( (UUT Out of Tolerance and Comparison In Tolerance) or*

*(Calibrator Out of Tolerance and Comparison Out of Tolerance))*

$= \Pr(|e_i(t_i)| \ge L_i\ and\ |d_i| \le L_i) +$

$\Pr(|e_{i-1}(t_{i-1})| \ge L_i\ and\ |d_i| \ge L_i)$

$$= 2\left[G\left(\frac{L_i}{\sigma_i}, \frac{-L_i}{\sigma_d}, \rho_i\right) - G\left(\frac{-L_i}{\sigma_i}, \frac{-L_i}{\sigma_d}, \rho_i\right)\right] +$$

$$2\left[F\left(\frac{-L_i}{\sigma_{i-1}}\right) + G\left(\frac{-L_i}{\sigma_{i-1}}, \frac{-L_i}{\sigma_d}, \rho_{i-1}\right) - G\left(\frac{-L_i}{\sigma_{i-1}}, \frac{L_i}{\sigma_d}, \rho_{i-1}\right)\right]$$

$$= PFA_i\,(R_{o,i}(t_i)) + 2\left[F\left(\frac{-L_i}{\sigma_{i-1}}\right) + G\left(\frac{-L_i}{\sigma_{i-1}}, \frac{-L_i}{\sigma_d}, \rho_{i-1}\right) - G\left(\frac{-L_i}{\sigma_{i-1}}, \frac{L_i}{\sigma_d}, \rho_{i-1}\right)\right] \tag{A.38}$$

where:

$$\sigma_{i-1}^2 = \sigma_{i-1}^2(t_{i-1})$$

$$= \left[ \frac{L_i}{F^{-1}\left(\frac{1+R_{oi}(t_i)}{2}\right)} \right]^2 - \sigma_i^2(t_i) \tag{A.39}$$

$$\rho_{i-1} = \mathrm{Cor}(e_{i-1}(t_{i-1}),\ e_i(t_i) - e_{i-1}(t_{i-1}))$$

$$= \frac{\sigma_{i-1}(t_{i-1})}{\sqrt{\sigma_i^2(t_i) + \sigma_{i-1}^2(t_{i-1})}} \tag{A.40}$$

When the UUT is always adjusted after calibration, the probability that the UUT is out of tolerance at the beginning of the period of use (BOP) corresponds to the probability that the calibrator is out of tolerance with respect to the UUT's engineering tolerances. This should be a reasonably unlikely event when the uncertainty of the calibrator is much smaller than the UUT. The probability of an initial out of tolerance (at BOP) when the UUT is always adjusted is given by:

$$BOPA_i = Pr(UUT\ Out\ of\ Tolerance\ at\ BOP\ after\ adjustment)$$

$$= Pr(Calibrator\ Out\ of\ Tolerance)$$

$$= \Pr(|e_{i-1}(t_{i-1})| \geq L_i)$$

$$= 2F\left(\frac{-L_i}{\sigma_{i-1}}\right) \tag{A.41}$$

**Use of the Risk Probabilities**

The difference between these four risk probabilities is as follows:

- $PFA_i$ is the proportion of in tolerance test results that are false. Here, the UUT is actually out of tolerance and bad equipment is being released.

- $PFR_i$ is the proportion of out of tolerance test results that are false. Here, the UUT is actually in tolerance and good test equipment is being rejected and processed.

- $BOPN_i$ is the proportion of UUT's that are released to customers in an out of tolerance condition when the adjustment policy is to adjust the UUT only when an out of tolerance is detected during calibration.

- $BOPA_i$ is the proportion of UUT's that are released to customers in an out of tolerance condition when the adjustment policy is to adjust the UUT after every calibration.

The BOP risk probabilities are probably the key measures of the quality of a metrology program. The entire reason for performing calibrations is to ensure that test equipment are in tolerance. The BOP risk probabilities indicate the extent to which the metrology program has succeeded in this endeavor.

**Risk Applied to Tested Systems**

Test equipment are used to test an end item or system. The process used to test a system using a piece of test equipment is mathematically equivalent to that used to test a UUT using a calibrator. As a result, the risk equations developed previously can be applied directly to the systems tested by test equipment where the $i^{th}$ subscript refers the tested system and the $i\text{-}1^{st}$ subscript refers to the test equipment. Using these definitions, we have the following:

- $PFA_i$ is the proportion of in tolerance test results that are false.

- $PFR_i$ is the proportion of out of tolerance test results that are false. Here, the system is actually in tolerance and a good system will be rejected and processed.

- $BOPN_i$ is the proportion of systems that are released to customers in an out of tolerance condition when the adjustment policy is to adjust the system only when an out of tolerance is detected during the test.

- $BOPA_i$ is the proportion of systems that are released to customers in an out of tolerance condition when the test equipment is used to adjust system.

Quite often, systems are tested, but not adjusted, using test equipment. The Out of Tolerance systems are removed from the inventory. The risk that such a test will leave a system in an out of tolerance condition is given by:

$$
\begin{aligned}
BOPR_i &= Pr(\textit{System Out of Tolerance at BOP when OOT's removed}) \\
&= Pr(\textit{System Out of Tolerance given Comparison In Tolerance}) \\
&= \Pr(|e_i(t_i)| \geq L_i \textit{ given } |d_i| \leq L_i) \\
&= \frac{2\left[G\left(\frac{-L_i}{\sigma_i}, \frac{L_i}{\sigma_d}, \rho_i\right) - G\left(\frac{-L_i}{\sigma_i}, \frac{-L_i}{\sigma_d}, \rho_i\right)\right]}{R_{o,i}(t_i)} \\
&= PFA_i \qquad \text{(A.42)}
\end{aligned}
$$