Method A3 – Interval Test Method

Description of the Methodology

ABSTRACT

A methodology is described for testing whether a specific calibration interval is consistent with a given reliability target. the method is referred to as the *interval test method* or, alternatively, as *method A3*.[1] Interval testing computes binomial confidence limits from an observed reliability. If the reliability target falls outside the confidence limits, the interval is said to have failed the test. The methodology is embedded in the **ISG Method A3 Interval Tester**.

BACKGROUND

Since the late '50s, manufacturing, testing and calibration organizations have periodically calibrated measuring and test equipment (MTE) to ensure that items are in-tolerance during use. It was realized early on that it was virtually impossible to predict the time at which an item would transition from an in-tolerance state to an out-of-tolerance state. Alternatively, what has been attempted is to find an interval of time between calibrations that results in holding the percentage of items in use to a minimum acceptable level.

This percentage is called the **reliability target**. Intervals of time between calibrations are referred to as **resubmission times**. Resubmission times are to be contrasted with assigned recall cycles for ensuring in-tolerance. The latter are called **calibration intervals**.

STATISTICAL METHODS

Several methods have been devised over the years to control MTE in-tolerance percentages. Some of these methods employ sophisticated statistical techniques to mathematically model in-tolerance probability vs. time elapsed since calibration. These methods are labeled **statistical methods**. Statistical methods attempt to predict a calibration interval that corresponds to a specific end-of-period in-tolerance percentage. They require considerable calibration history for analysis and are often difficult to implement.

ALGORITHMIC METHODS

Other methods utilize simple to complex decision algorithms to adjust calibration intervals in response to in-tolerance or out-of-tolerance conditions observed during calibration. Typically, these approaches consist of instructions to lengthen or shorten calibration intervals in response to current or recent observations. Because of their nature, these methods are labeled **algorithmic methods**.

Algorithmic methods have achieved wide acceptance due to their simplicity and low cost of implementation. However, most algorithmic methods suffer from several drawbacks. The following list is fairly representative:

1. With most algorithmic methods, interval changes are in response to small numbers (usually one or two) of observed in-tolerance or out-of-tolerance conditions. It can be easily shown that any given in-tolerance or out-of-tolerance condition is a random occurrence. Adjusting an interval in response to small numbers of calibration results is, accordingly, equivalent to attempting to control a process by adjusting to random fluctuations. Such practices are inherently futile.
2. Algorithmic methods make no attempt to model underlying uncertainty growth mechanisms. Consequently, if an interval change is required, the appropriate magnitude of the change cannot be readily determined.
3. Algorithmic methods cannot be readily tailored to prescribed reliability targets that are commensurate with quality objectives. The level of reliability attainable with a given algorithmic method can be discovered only by trial and error or by simulation.[2,3,4]

---

[1] *Establishment and Adjustment of Calibration Intervals*, Recommended Practice RP-1, National Conference of Standards Laboratories, January 1996.

[2] Jackson, D., Ferling, J. and Castrup, H., "Concept Analysis for Serial Number Based Calibration Intervals," Proc. 1986 Meas. Sci. Conf., Irvine, January 23-24.

4. If an interval is attained that *is* consistent with a desired level of reliability, the results of the next calibration or next few calibrations will likely cause a change away from the correct interval. To see that this is so, consider cases where reliability targets are high, e.g., 90%. For a 90% target, if the interval is correct for an item, there is a 0.9 probability that it will be observed in-tolerance at any given calibration. Likewise, there is a 0.81 probability that it will be observed in-tolerance at two successive calibrations. With most algorithmic methods, such observations will cause an adjustment away from the item's current interval. Thus, algorithmic methods tend to cause a change away from a correct interval in response to events that are highly probable if the interval is correct.

5. With algorithmic methods, although a correct interval cannot be maintained, a time-averaged steady-state measurement reliability can be achieved. The typical time required ranges from fifteen to sixty years.[5]

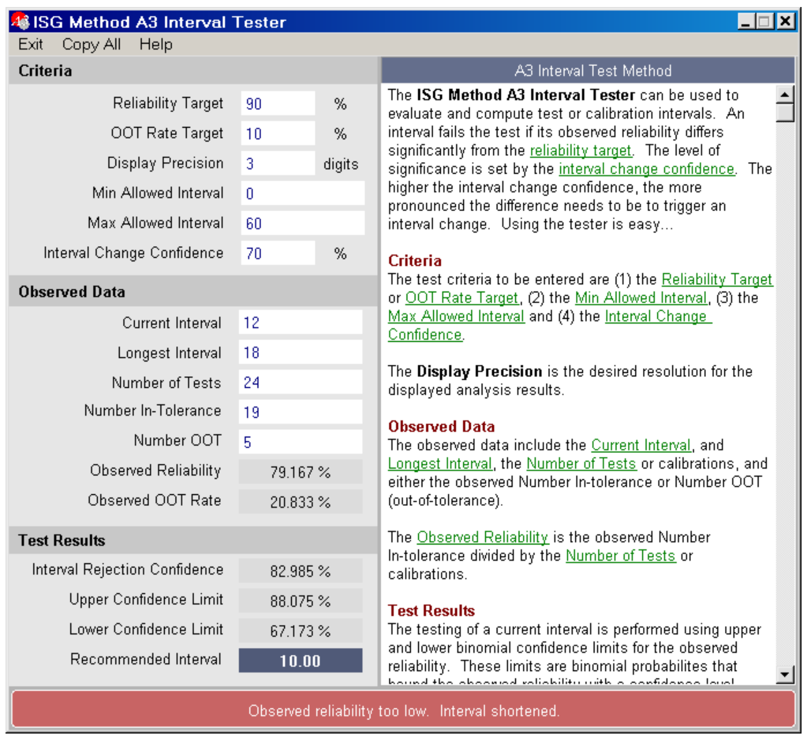
6. With algorithmic methods, interval changes are ordinarily computed manually by calibrating technicians, rather than established via automated methods. Accordingly, operating costs can be high.

METHOD A3

Method A3 was developed to overcome the deficiencies noted in 1, 3, 4 and 5 above. The description of the method and its implementation in the Method A3 Interval Tester is described in the following sections.

RELEVANT VARIABLES

- R_{obs} = Observed Reliability at the Current Interval
- R_{targ} = Reliability Target
- I_{rec} = Recommended Interval
- I_{trial} = trial value
- I_{cur} = Current Interval (current assigned interval)
- I_{long} = Longest Interval (longest observed resubmission time)
- I_{max} = Max Allowed Interval
- I_{min} = Min Allowed Interval
- Q = Interval Rejection Confidence
- C = Interval Change Confidence



The Method A3 Interval Tester

INTRODUCTION

For calibration interval analysis, the term *reliability* refers to the probability that an MTE item or parameter is in-tolerance. The *observed reliability* for an MTE item or parameter is defined as the fraction of items or parameters that are found to be in-tolerance when tested, calibrated or otherwise inspected.

When items are periodically tested or calibrated, the *observed reliability* is the fraction of items observed to be in-tolerance. If the resubmission time occurs at the end of the item's recall cycle, it is synonymous with the test or calibration interval. When items are received from an external source and inspected prior to being placed in service, the observed reliability is the fraction of items found to be in-tolerance during inspection.

In the present discussion, we represent the observed reliability with the variable R_{obs}. Testing an interval involves comparing R_{obs} against some the reliability target, denoted R_{targ}.

---

[3] Kuster, M., et al., "Calibration Interval Adjustment: The Effectiveness of Algorithmic Methods," Proc. 2009 NCSLI Workshop and Symposium, San Antonio, July 2009.

[4] Kuster, et al., "A Quantitative Comparison of Calibration Interval Adjustment Methods," Proc. 2010 Meas. Sci. Conf., Pasadena, March 25-26.

[5] Jackson, et al., op. cit.

With the A3 Interval Tester, the observed reliability is computed by dividing the **Number of Tests** (n) conducted at the Current Interval into the **Number In-Tolerance** (x) observed at the Current Interval. For example, if $n = 24$ and $x = 19$, the observed reliability is

$$\begin{aligned} R_{obs} &= \frac{x}{n} \\ &= \frac{19}{24} = 0.792. \end{aligned}$$

TESTING THE INTERVAL

An interval, labeled the **Current Interval**, is tested by comparing R_{obs} to R_{targ} to see if there is a significant difference between the two.

The magnitude of the difference and the **Test Confidence Level** determine whether or not the difference is significant. If so, the Current Interval is rejected and an adjusted interval is recommended.

COMPUTING TEST CONFIDENCE LIMITS

The algorithm used to test the Current Interval evaluates whether upper and lower confidence limits for R_{obs} contain the reliability target. If so, then the interval passes the test.

The confidence limits for R_{obs} are determined using the binomial distribution function. Given a Number of Tests n and a Number In-Tolerance x, the **upper confidence limit** p_U is obtained from the relation

$$\sum_{k=0}^{x} \binom{n}{k} p_U^k \left(1 - p_U\right)^{n-k} = \begin{cases} \dfrac{1-C}{2}, & 0 < x < n \\ 1, & x = n \\ 1 - C, & x = 0, \end{cases}$$

where $\boldsymbol{C}$ is the **Interval Change Confidence** level. Similarly, the **lower confidence limit** p_L is obtained from

$$\sum_{k=x}^{n} \binom{n}{k} p_L^k \left(1 - p_L\right)^{n-k} = \begin{cases} \dfrac{1-C}{2}, & 0 < x < n \\ 1 - C, & x = n \\ 0, & x = 0. \end{cases}$$

If $p_L \le R_{targ} \le p_U$, then the Current Interval passes the test. Otherwise, it is rejected.

TEST RESULTS

Following interval testing and adjustment, the results are shown in the Test Results section of the Method A3 Interval Tester. The displayed fields are Rejection Confidence, Upper Confidence Limit, Lower Confidence Limit and Recommended Interval.

DESIRED DISPLAY PRECISION

Results can be view to a desired level of precision by entering the number of digits to be displayed following the decimal in the **Desired Displayed Precision** box. Up to eight digits are allowed.

INTERVAL REJECTION CONFIDENCE

This is the confidence Q for rejecting the Current Interval, given the entered data. It is obtained by numerical iteration of Q in which values of p_1 and p_2 are obtained at each iteration that satisfy

$$\sum_{k=x}^{n}\binom{n}{k}p_1^k\left(1-p_1\right)^{n-k}=\frac{1-Q}{2}$$

and

$$\sum_{k=0}^{x}\binom{n}{k}p_2^k\left(1-p_2\right)^{n-k}=\frac{1-Q}{2}\ .$$

The process begins with a value of Q near 1.0 (e.g., 0.9999999). In solving for Q, use is made of the bisection method .[6] At each step of the process, estimates for p_1 and p_2, are obtained using the incomplete beta functions[7]

$$I_{p_1}(x,n-x+1)=\sum_{k=x}^{n}\binom{n}{k}p_1^k\left(1-p_1\right)^{n-k}=(1-Q)/2$$

and

$$I_{p_2}(x+1,n-x)=Q/2\ .$$

For each step, initial estimates are obtained using an approximate solution p for the function $I_p(a,b)$

$$p\cong\frac{a}{a+be^{2w}}\ ,$$

where

$$w=\frac{y_p(h+\lambda)^{1/2}}{h}-\left(\frac{1}{2b-1}-\frac{1}{2a-1}\right)\left(\lambda+\frac{5}{6}-\frac{2}{3h}\right),$$

$$h=2\left(\frac{1}{2b-1}+\frac{1}{2a-1}\right)^{-1},$$

$$\lambda=\frac{y_p^2-3}{6}\ ,$$

$$y_p=\Phi^{-1}(1-P)\ ,$$

and where Φ is the inverse normal distribution function. At each iteration, the parameters a, b and P are defined as shown in the table below

Initial Value	Parameter	Value
p_1	a	x
	b	$n-x+1$
	P	$(1-Q)/2$
p_2	a	$x+1$
	b	$n-x$
	P	$Q/2$

The iteration process stops when

$$p_1<R_{targ}\quad \text{and}\quad R_{targ}-p_1<\varepsilon\ ,$$

or

$$p_2>R_{targ}\quad \text{and}\quad p_2-R_{targ}<\varepsilon\ ,$$

where ε is a preset level of precision.[8] If either limit is found during the iteration, the solved for value of Q is displayed in the **Interval Rejection Confidence box**.

[6] See, for example, Press, et al., **Numerical Recipes,** 2nd Ed., Cambridge University Press, 1992.

[7] See, for example Abramowitz, M. and Stegun, I., *Handbook of Mathematical Functions*, NBS Applied Mathematics Series, **55**, June 1964.

[8] In the ISG Method A3 Interval Tester, $\varepsilon = 0.00001$.

UPPER CONFIDENCE LIMIT

This is the value of p_U that corresponds to the entered Interval Change Confidence C.

LOWER CONFIDENCE LIMIT

This is the value of p_L that corresponds to the entered Interval Change Confidence C.

RECOMMENDED INTERVAL

This is the interval I_{rec} computed as described in the Interval Establishment Algorithm.

ADJUSTING INTERVALS

If the Current Interval is rejected, a new recommended interval is computed using a simple algorithm. The process takes place in two stages. In the first stage, the interval is adjusted to a trial value. In the second stage, the trial value is refined to ensure it's feasibility and compliance with administrative constraints.

Interval Adjustment Algorithm

If $R_{obs} > R_{targ}$ Then ' The interval may be too short
  If $Q = 1$ Then
    $I_{trial} = 2\ I_{cur}$
  Else
    $y = (R_{obs} - R_{targ}) / (1 - Q)$
    If $y > 25$ Then
      $I_{trial} = \text{Min}(I_{max}, 1.2\ I_{long})$
    Else
      $I_{trial} = \text{Int}\left(2^y I_{cur} + 0.5\right)$
    End If
  End If
Else ' The interval may be too long
  $I_{trial} = \text{Int}\left[10^{(R_{obs} - R_{targ})Q} I_{cur} + 0.5\right]$
End If

Interval Establishment Algorithm

' Establish I_{rec} adhering to the interval criteria
If $I_{trial} > 1.2\ I_{long}$ and $I_{long} > 0$ Then
  $I_{trial} = 1.2\ I_{long}$
End If
If $I_{trial} > 2\ I_{cur}$ Then
  $I_{trial} = 2\ I_{cur}$
End If
If $I_{trial} < I_{cur} / 2$ and $I_{cur} / 2 < 1.2\ I_{long}$ Then
  $I_{trial} = I_{cur} / 2$
End If
If $I_{trial} > I_{max}$ and $I_{max} > 0$ Then
  $I_{trial} = I_{max}$
End If
If $I_{trial} < I_{min}$ Then
  $I_{trial} = I_{min}$
End If
$I_{rec} = I_{trial}$

SETTING THE INTERVAL CHANGE CONFIDENCE

As indicated earlier, Method A3 was developed in part to prevent interval changes in response to observations that are consistent with a correct interval. The stringency with which such changes are prevented is established by setting the Interval Change Confidence.[9]

In tests of statistical significance, confidence levels are customarily set at 95%. If the Interval Change Confidence is set to this value in the Method A3 Interval Tester, the limits p_L and p_U are 95% confidence limits for R_{obs}, and the interval will be rejected with 95% confidence if R_{targ} lies outside the range (p_L , p_U).

LENGTHENING INTERVALS

Reliability Target Considerations

It is important to bear in mind, that the interval is rejected only if R_{obs} is significantly different from R_{targ}, where the level of significance is established from the Interval Change Confidence. Hence, since the maximum possible value of R_{obs} is 1.0 (i.e., 100%), if R_{targ} is high, the interval will not be lengthened unless a substantial amount of data are collected to support rejecting the Current Interval. Consider the following:

[9] In all cases considered in this section, a nonzero Current Interval is entered, the Min Allowed Interval is set to 0; Max Allowed Interval and Longest Interval fields are set to 10000.

Suppose that R_{targ} = 95% and the Interval Change Confidence is 95%. A recommendation to lengthen the interval will not occur until 59 in-tolerances are observed out of 59 tests. If R_{targ} = 90%, then a recommendation to lengthen the interval will occur if 29 in-tolerances are observed out of 29 tests. If R_{targ} = 85%, then a recommendation to lengthen the interval will occur if 19 in-tolerances are observed out of 19 tests.

Interval Change Confidence Considerations

Varying the Interval Change Confidence yields equally interesting results. For instance, suppose that R_{targ} = 95% and the Interval Change Confidence is 80%. A recommendation to lengthen the interval will occur when 32 in-tolerances are observed out of 32 tests. If the Interval Change Confidence is 70%, the recommendation will occur when 24 in-tolerances are observed out of 24 tests.

SHORTENING INTERVALS

Reliability Target Considerations

We have seen that the higher the reliability target, the more difficult it is to lengthen an interval. The opposite is true for shortening intervals. For example, suppose that R_{targ} = 95% and the Interval Change Confidence is 95%. A recommendation to shorten the interval will occur if 2 out-of-tolerances are observed out of 2 tests. If the Interval Change Confidence is 95%, but R_{targ} = 64%, then a recommendation to shorten the interval will not occur until 3 out-of-tolerances are observed out of 3 tests.

Interval Change Confidence Considerations

As with interval increases, recommendations to decrease the interval occur more readily for lower Interval Change Confidence values. For instance, suppose that R_{targ} = 90% and the Interval Change Confidence is 95%. A recommendation to shorten the interval will not occur until 2 out-of-tolerances are observed out of 2 tests. However, if the Interval Change Confidence is 89%, then a recommendation to shorten the interval will occur if 1 out-of-tolerance is observed out of 1 test.

INTERVAL CHANGE CONFIDENCE GUIDELINES

From the foregoing, it is apparent that reducing the Interval Change Confidence promotes more frequent rejection of the Current Interval and attendant interval changes. As we have seen, however, reducing this confidence makes it easier to shorten intervals as well as to lengthen them. In addition, for high values of R_{targ}, intervals are resistant to increases, whereas they are amenable to decreases.

Since intervals are considerably more resistant to increases than decreases for high reliability targets, it might be considered justifiable to adjust the Interval Change Confidence to produce increases in a “reasonable” way, keeping in mind the fact that relaxing this confidence leads to more frequent interval decreases as well as increases.

All other things being equal, a general rule might be stated as “the higher the Reliability Target, the lower the Interval Change Confidence.” Of course, this rule is statistical heresy but might be instituted for matters of practicality. To wit, consider the prescription shown in Table 1 that is based on the claim that interval increases are justified if 20 in-tolerances are observed out of 20 tests. As the table shows, we reach a kind of “cut-off” at R_{targ} = 85%, where a conventional confidence level can be employed.

If a minimum reasonable criterion for increasing intervals is considered to be 30 in-tolerances observed out of 30 tests, then Table 2 applies. As Table 2 shows, for this minimum criterion, a conventional confidence level is reached at a reliability target of 90%.

Other criteria are possible. Suppose for instance, that the minimum reasonable interval increase criterion is 49 in-tolerances observed out of 50 tests (98% in-tolerance). In this case, Table 3 applies.

Table 1

Interval Change Confidence vs. Reliability Target
for a 20 In-tolerance out of 20 Tests
Minimum Interval Increase Condition

Reliability Target	Max Interval Change Confidence (%)	Minimum Interval Increase Condition (# in-tol / # tests)	Minimum Interval Decrease Condition (# OOT / # tests)
0.95	64	20 out of 20	0 out of 1
0.90	87	20 out of 20	0 out of 1
0.85	96	20 out of 20	0 out of 2
0.80	98	20 out of 20	0 out of 3

Table 2

Interval Change Confidence vs. Reliability Target
for a 30 In-tolerance out of 30 Tests
Minimum Interval Increase Condition

Reliability Target	Max Interval Change Confidence (%)	Minimum Interval Increase Condition (# in-tol / # tests)	Minimum Interval Decrease Condition (# in-tol / # tests)
0.95	**78**	**30 out of 30**	**0 out of 1**
0.90	**95**	**30 out of 30**	**0 out of 2**
0.85	**99.2**	**30 out of 30**	**0 out of 3**
0.80	**98**	**30 out of 30**	**0 out of 3**

Table 3

Interval Change Confidence vs. Reliability Target
for a 49 In-tolerance out of 50 Tests
Minimum Interval Increase Condition

Reliability Target	Max Interval Change Confidence (%)	Minimum Interval Increase Condition (# in-tol / # tests)	Minimum Interval Decrease Condition (# in-tol / # tests)
0.95	44	49 out of 50	0 out of 1
0.90	93	49 out of 50	0 out of 2
0.85	99.4	49 out of 50	0 out of 3

OTHER FACTORS

From the algorithms shown under Adjusting Intervals, we see that interval adjustments are sensitive to certain ancillary factors, such as Min Allowed Interval, Max Allowed Interval and Longest Interval.

MIN / MAX ALLOWED INTERVAL

These criteria sometimes affect whether an interval is rejected or not. For example, if the Current Interval is 90 and the Max Allowed Interval is 90, an interval increase will not be recommended.

These parameters also serve as constraints on recommended interval changes, as is shown in the interval adjustment algorithms. To illustrate, suppose that the Max Allowed Interval is 10000, the Current Interval is 90, the Longest Interval is 200, R_{targ} = 90% and the Interval Change Confidence is 95%. If 30 in-tolerances have been observed out of 30 tests, the recommended interval will be 180.

Given the same circumstances with a Max Allowed Interval of 120, the recommended interval will be 120.

LONGEST INTERVAL

The Longest Interval is the longest observed resubmission time. The Method A3 Interval Tester will not recommend an interval that is more than 1.2 times this value.

Suppose that the Max Allowed Interval is 10000, the Current Interval is 90, the Longest Interval is 200, R_{targ} = 90% and the Interval Change Confidence is 95%. If 30 in-tolerances have been observed out of 30 tests, the recommended interval will be 180.

Given the same circumstances with a Longest Interval of 120, the recommended interval will be 144.

CAPTURING THE ANALYSIS RESULTS

The analysis results can be captured to the Windows clipboard by clicking **Copy All** on the menu. The results can then be pasted into a spreadsheet or other application.