N91-27107

P_15

EVALUATING THE EFFECT OF ACCURACY RATIOS ON THE PERCENT OF CALIBRATIONS WHICH ARE OUT OF TOLERANCE

Final Report

NASA/ASEE Summer Faculty Fellowship Program--1990

Johnson Space Center

V018760

Prepared by: Sharon E. Navard, Ph.D.

Academic Rank: Assistant Professor

University & Department: Virginia Commonwealth University
Department of Mathematical Sciences
Richmond, Virginia 23284-2014

NASA/JSC

Directorate: Safety, Reliability and Quality Assurance

Division: Quality Assurance and Engineering

Branch: Flight Systems Quality Engineering

JSC Colleague: Otho T. Dickerson

Date Submitted: August 10, 1990

Contract Number: NGT 44-005-803

ABSTRACT

The standard practice in calibration laboratories across the country, including the Measurement Standards and Calibration Laboratory (MSCL) at the Johnson Space Center, is to use accuracy ratios to determine if instruments are in-tolerance rather than computing the actual uncertainty associated with the instruments. In the past the accepted practice was to use an accuracy ratio of 10:1, but then state-of-the-art advanced to the point where the 10:1 ratio could no longer be maintained, and the ratio was arbitrarily lowered to 4:1. It is now becoming increasingly difficult to maintain the 4:1 accuracy ratio, and in some cases 1:1 is the best that can be achieved. However, the effect of using these small accuracy ratios on the number of mistakes made in classifying instruments as in- or out-of-tolerance is completely unknown.

In order to assess the effect of using accuracy ratios in calibration, a simulation program was written to compute the proportion of instruments determined to be out-of-tolerance which were actually in, denoted by α, and the proportion of instruments determined to be in-tolerance which were actually out, denoted by β. This was done for accuracy ratios of 1:1 to 10:1, for one to five progressive calibrations, under varying standard and instrument conditions. Selected results are presented and explained in this report; the full set of results, as well as the simulation program itself, can be obtained from the author.

INTRODUCTION

The Measurement Standards and Calibration Laboratory (MSCL) at the Johnson Space Center is responsible for calibrating all of the instruments used at JSC and by offsite contractors. Calibration involves determining if an instrument measures accurately within its tolerance as specified by the manufacturer and, if it does not, adjusting or repairing it so that it does. The time interval between calibrations for each instrument is determined so that, in theory, it will be re-calibrated before it drifts far out of tolerance.

Each calibration that is performed must possess a property known as *traceability*. This means that the standard which was used to do the calibration must be traceable back to either an intrinsic standard or to the national standards maintained by the National Institute of Standards and Technology (NIST). Instrinsic standards are those which are based on physical laws of nature or naturally occurring phenomena, such as the speed of light or the triple point of water. These standards are created or maintained within the Reference Standards Laboratory at JSC, and no interaction with the NIST is necessary for these calibrations. For those quantities for which no intrinsic standard is known, however, the national standards maintained by the NIST are the legal basis for a measurement system in the United States.

While the intrinsic and national standards are by definition the "true" values, any measurements made on them are, unfortunately, subject to error. In order for a calibration to be meaningful, the magnitude of the possible error must be known. There are two characteristics of the measurements which are of interest: the *accuracy*, which determines how close the measurements are, on average, to the true value, and the *precision*, which determines how close repeated measurements made with the same instrument are to each other. An instrument can be accurate but imprecise, and vice versa.

There are several levels of accuracy and precision of standards used in the MSCL. The most accurate are the *reference standards*, which are the most precise standards available, and are calibrated either in the MSCL directly from the intrinsic standards or by the NIST directly from the national standards. The reference standards are recalibrated periodically to maintain their level of accuracy, a process which can be very expensive and time consuming, especially if it involves physically returning the standards to the NIST. Thus, the integrity of the reference standards is closely guarded. Since merely handling the standards can reduce their accuracy, the reference standards are used only to calibrate the *transfer standards*, which in turn are used only to calibrate the *working standards*. It is the working standards which are actually used to calibrate customers' instruments. They are periodically recalibrated by the transfer standards, which are in turn recalibrated by the reference standards. Thus the accuracy of the reference standards is maintained as long as possible.

The problem with using this progression of standards is that accuracy is lost at each successive step. When a reference standard is used to calibrate a transfer standard, the measurements given by the reference standard are taken to be the actual values; but due to the inaccuracy and imprecision of the reference standard, these values will differ from the "true" values, thus making the reference standard inaccurate by this amount. Also, the imprecision of the transfer standard itself contributes to the inaccuracy, increasing its magnitude. When the transfer standard is then used to calibrate the working standard, its measurements are taken to be "true," and hence this larger inaccuracy is passed on, and is in turn increased by the imprecision of the transfer and working standards.

The result of this is that, when calibrating an instrument which has a certain precision as stated by the manufacturer, the calibrating standard must be even more precise. The precision of an instrument is stated as a *tolerance*, which is theoretically the largest possible magnitude of the difference between the measured and "true" values, assuming the instrument is accurate. The instrument is said to be calibrated, or *in-tolerance*, if a very high percentage (usually more than 99%) of measurements that it gives are within its stated tolerance of the "true" value. Thus to determine if an instrument is calibrated, it is necessary to know the precision and accuracy of the standard; and since the "true" value is unknown, mistakes will be made. The aim of this report is to determine how many mistakes are made under certain circumstances.

DETERMINING THE TOLERANCE OF AN INSTRUMENT

Determining the precision and accuracy of an instrument, and thus the uncertainty associated with it, can be a very difficult and time consuming problem. Extensive literature has been devoted to the subject--see, for example, Abernethy and Benedict (1985); Cameron (1976); Colclough (1987); Croarkin (1984); Eisenhart (1963); Ku (1966); Ku and Judish (1986); and Schumacher (1981). Unfortunately, due to the time involved in computing uncertainties, the common practice in calibration laboratories across the country is to employ the use of accuracy ratios rather than actually computing the uncertainty. A brief description of both methods follows.

Random and Systematic Errors

There are two types of errors which must be evaluated in order to compute the uncertainty of an instrument--the *random error*, which is a measure of precision, and the *systematic error*, which is a measure of the accuracy or offset. The total uncertainty, or tolerance, of the instrument is then the sum of the random and systematic errors.

Since the random error is a measure of precision, it can be estimated by taking repeated measurements with the instrument, under all of the different operational and environmental conditions with which the instrument will be used. The random error is then taken to be some multiple of the standard deviation of the measurements. When the manufacturer specifies a tolerance for an instrument, this tolerance is an estimate of the random error of the instrument; for most of the instruments used in the MSCL, the tolerance is set at three standard deviations. When there is more than one source of random error, such as in progressive calibrations or very complex calibrations, the random errors from the different sources can be combined in quadrature.

The systematic error is much more difficult to compute because, unlike the random error, it depends on the unknown "true" value and cannot be ascertained by repeated measurements. One method of estimating the systematic error is to use the tolerance of the standard used to do the calibration. When there is more than one source of systematic error, there is no general agreement on a method of combining them. Because they are not random, they cannot necessarily be expected to cancel each other out, so the most conservative method of combining them is a direct sum. However, some sources (see, for example, Schumacher (1988)) argue that since systematic errors from different sources are independent of each other, they can be viewed as random observations from another process and can thus be treated as standard deviations and combined in quadrature.

The ideal situation in doing progressive calibrations would be to compute the random and systematic errors at each step and keep track of them, combining them at each successive calibration. If this were done, the MSCL could report to each customer the actual uncertainty associated with his instrument, and also would be able to determine what proportion of instruments will be classified incorrectly as being in- or out-of-tolerance. Unfortunately, due to the time involved with computing random errors, this is not done; instead, accuracy ratios are employed to determine whether or not the instrument reads accurately within the manufacturer's specifications.

Accuracy Ratios

Calibration by using accuracy ratios is much faster and simpler than the method described above because no repeated measurements are taken and the overall uncertainty is not computed. If a test instrument with tolerance T is to be calibrated using an accuracy ratio of R:1, then the standard used to calibrate it must have a tolerance of T/R. A quantity is then measured using the standard, and the resulting reading is taken to be the "true value." The same quantity is then measured using the test instrument, and if the reading is within T of the assumed "true value," the test instrument is said to be in-tolerance; otherwise, it is said to be out-of-tolerance. Any time the reading on the test instrument is more than (0.7)T from the assumed true value, the test instrument is adjusted to read the same as the standard.

Since this method of calibration is much faster than the other, it is the method that is used in the calibration labs of all of the NASA centers as well as other calibration labs throughout the country. The JSC Metrology Requirements Manual (1990) specifically advocates the use of a 4:1 accuracy ratio. The reason for using it is expediency; the vast majority of the theory of calibration in the literature is about computing uncertainties. Very little work has been done on determining how well the use of accuracy ratios actually works. It is undesirable from a theoretical point of view because it loses all information about the actual uncertainty of the instrument. Furthermore, the proportion of mistakes made in determining if an instrument is in- or out-of-tolerance is unknown.

In the past, no one worried about what was lost by using accuracy ratios because the accepted practice was to use a 10:1 ratio, which intuitively seemed to guarantee accurate calibrations. However, as the technology improved, customers were able to obtain more precise instruments and it became impossible to maintain the 10:1 ratio. At this point, about twenty years ago, the accuracy ratio was arbitrarily lowered to 4:1 by tacit agreement among the calibration community. The effect of lowering the ratio was unknown, but the consensus seemed to be that it still did a good job and still no one worried about it; theoreticians continued to ignore what was being done in practice and continued to research computing uncertainties. However, the state-of-the-art has now improved to the point where the 4:1 ratio can no longer be maintained in many disciplines and pressure is being felt in the calibration community to lower the accuracy ratio to 3:1; the U. S. Navy has already done so. Even this won't solve the problem in some disciplines, however, where the best accuracy ratio that can be maintained is 1:1. Metrologists are finally starting to worry.

Because the calibration community has only recently become aware that it has a problem, very little work has been done on determining the effect of using accuracy ratios on the proportion of instruments which are incorrectly determined to be in- or out-of-tolerance. Nothing on the subject has been published in the major statistical journals, although some papers have recently been presented on the problem at metrology conferences. See, for example, Capell (1988); Capell (1989); and Schumacher (1988). Of these, only Schumacher addresses the specific problem of computing the proportion of instruments which are incorrectly classified. He actually computes the proportion of products which are incorrectly determined to meet or not meet specifications when the measuring instrument has a tolerance that is a fraction of the specifications, but this is equivalent to the case of using an accurate standard to calibrate a customer's instrument.

THE QUESTION

The specific question to be addressed is this: What proportion of instruments are incorrectly determined to be in- or out-of-tolerance, after one to five progressive calibrations, using accuracy ratios of 1:1 to 10:1? This should be answered for different proportions of the instruments being calibrated being actually out-of-tolerance, and for the standards being both in-tolerance and out-of-tolerance.

Notation

The following notation will be used in the remainder of this report.

- α = proportion of instruments determined to be out-of-tolerance which are actually in tolerance
- β = proportion of instruments determined to be in-tolerance which are actually out-of-tolerance
- I_0 = the reference standard
- I_i = the i^{th} instrument progressively calibrated after the reference standard
- E_i = the error distribution of I_i
- k = the number of standard deviations of E_i defining "in-tolerance"
- σ = standard deviation of reference standard error distribution, E_0

100p = percent that calibrating standard is out-of-tolerance
r:1 = accuracy ratio

Employing this notation, and assuming that the accuracy ratios are exactly correct, the reference standard I_0 will have tolerance $k\sigma$, while I_i will have tolerance $r^i k\sigma$, $i = 1, 2, \ldots, 5$. If I_i is in-tolerance when it is used as a standard, $i = 0, \ldots, 4$, then the standard deviation of the error distribution E_i is $r^i\sigma$; otherwise it is $(1 + p)r^i\sigma$.

Assumptions of the Model

In order to compute α and β, it is first necessary to define the specific probability model that is being observed. When working with measurement errors, a standard assumption to make is that the errors have a normal distribution with mean zero. It is also assumed that an r:1 accuracy ratio is exactly maintained throughout the progression of calibrations. Therefore,

$$E_i \sim N(0, [r^i\sigma]^2), \; i = 0, \ldots, 5.$$

If I_i is actually 100p% out-of-tolerance, then

$$E_i \sim N(0, [(1 + p)r^i\sigma]^2).$$

Note that if $p = 0$, then this distribution reduces to that given when the standard is in-tolerance. Henceforward, the general form of the distribution of E_i will be used.

It is necessary to know not only the error distribution, but also the distribution of the measurements themselves. With E_i defined as above, the distribution of measurements X_i taken from I_i will also have a normal distribution with some mean μ_i and the same standard deviation; that is,

$$X_i \sim N(\mu_i, [(1 + p)r^i\sigma]^2), \; i = 0, \ldots, 5.$$

Thus far the assumptions have been very straightforward, but at this point one is faced with a dilemma: what is μ_i? Ideally, μ_i will be equal to the "true" value at each step, which would mean that the instruments are all exactly accurate. Unfortunately, this will almost certainly not be the case; a systematic error will probably be present. But what is the magnitude of the systematic error? The metrologists at the MSCL were unable to give any practical insight into this problem; systematic errors are the ones which are nearly impossible to estimate, even if the time and resources to do so are available. It is thus necessary to use some intuitive reasoning.

The reference standard I_0 must be treated differently from the others because it is calibrated using actually *known* values, that is, the intrinsic or national standards. For simplicity's sake, denote the "true" value by μ. When I_0 is being calibrated, it should be made to register exactly μ. However, due to the imprecision of the instrument itself, it will be set to read a value slightly different from μ. Since $E_0 \sim N(0, [(1 + p)\sigma]^2)$, μ_0 will on the average be equal to μ, but will vary from it an amount determined by the distribution of E_0. Thus,

$$\mu_0 \sim N(\mu, [(1 + p)\sigma]^2).$$

The exact value of μ_0 will depend on the particular observations made by the instrument at the time of the calibration. In practice, calibration of the reference standard is done at the NIST, if national standards are used, or in the Reference Standards Lab of the MSCL, if intrinsic standards are used. In either case, repeated measurements are made, and σ is then estimated. At this step, accuracy ratios are not used.

Determination of μ_i, $i = 1, \ldots, 5$, requires further reasoning. The situation is not as simple as that of determining μ_0 because, instead of actually setting μ_i based on observations from I_i, it is now necessary to know what μ_i is before any observations are ever taken. This is because of the method associated with using accuracy ratios--one observation will be taken from the standard I_{i-1} and the test instrument I_i, and these observations will then be compared to determine if I_i is in-tolerance.

At this point, it is necessary to make some assumptions, which may and may not be correct. First, assume that I_i is, on the average, in-tolerance. However, a certain percentage of the instruments being calibrated are found to be out-of-tolerance; at the MSCL, this percentage is 10% or more. Thus the possible values of μ_i must be allowed to shift about μ in either direction, out to a distance that is determined by the percent of instruments coming in which are actually out of tolerance. For simplicity, it will again be assumed that μ_i has a normal distribution with mean μ, but now the standard deviation must be larger than that of E_i. In fact, it is assumed that

$$\mu_i \sim N(\mu, [D(1 + p)r^i\sigma]^2),$$

where D is the drift factor that determines what percent of the instruments coming in to be calibrated are actually out of tolerance.

Similar to the case of μ_0, the exact value of μ_i will depend on the particular instrument which is being calibrated.

Defining the measurement distributions as above, the values of α and β will vary depending on the values of p and D. However, due to the nature of the normal distribution, the particular values of μ and σ are unimportant; the probabilities will be the same no matter what values are used. Therefore it is assumed without loss of generality that $\mu = 0$ and $\sigma = .01$.

THE EFFECT OF USING ACCURACY RATIOS

Now that the specific model has been defined, it is now possible to compute α and β. Unfortunately, it is possible to obtain an analytical solution only in a limited situation--when both the standard and test instrument are accurate. Even then it is only possible to obtain the solution for one calibration, because when this calibration is performed a systematic error is introduced. Therefore, in order to evaluate α and β, it is necessary to perform a simulation.

Simulation Program

The simulation program was written in SAS/IML, the Interactive Matrix Language of the statistical package SAS. The random number generator used was the SAS function RANNOR. The parameter values used were:

k = 2.4, 3;
p = 0, .1, .25, .5, 1, 2, 3; and
D = 15%, 32%, 65%, 92%.

For each combination of these parameters, 10,000 iterations were performed using accuracy ratios of 1:1 to 10:1, and progressing 5 steps down from the reference standard.

The algorithm for the program is as follows. At each iteration, for i = 1 to 5, observations x_{i-1} and x_i are obtained from I_{i-1} and I_i by first generating the means, and then the observations themselves. Once the mean is generated for an instrument when it is the test intrument, the same mean is used at the next step when that same instrument becomes the standard. Since the "true" value is zero, I_i is defined to be *actually* in-tolerance if $|x_i| < kr^i\sigma$, where r is the accuracy ratio. Since the perceived true value is x_{i-1}, I_i is *perceived* to be in-tolerance if $|x_{i-1} - x_i| < kr^i\sigma$. There are therefore four possible outcomes: the instrument is actually in-tolerance, and is also perceived to be in-tolerance; the instrument is actually out-of-tolerance and is also perceived to be out-of-tolerance; the instrument is actually in-tolerance but is perceived to be out-of-tolerance (an α error); or the instrument is actually out-of-tolerance but is perceived to be in-tolerance (a β error). The outcome for each step is observed, and the total number occurring of each outcome out of the 10,000 iterations is counted. In keeping with the procedure actually practiced in the MSCL, if $|x_{i-1} - x_i| > (.7)kr^1\sigma$, I_i is adjusted to read the same as I_{i-1}. This procedure is done for accuracy ratios of 1:1 to 10:1.

Results

The output of the program for selected parameter values is shown in Tables 1 - 4; output using the other parameter values follows a similar pattern. In general, it can be seen that, as one would expect, no matter what the parameter values are, α and β decrease as the accuracy ratio is increased. For a particular accuracy ratio, α is generally larger than β, with the magnitude of the difference decreasing as the accuracy ratio is increased. This is desirable because α is a less severe error than β; it is also the same pattern that was observed in Schumacher (1988). When the accuracy ratio is 1:1, α and β both increase as progressive calibrations are made from one step to the next, but this increase vanishes as the accuracy ratio is increased. For accuracy ratios of 2:1 and 3:1, there is an increase in α and β from the first step to the second, but they remain constant afterward; for accuracy ratios of 4:1 and higher, α and β remain fairly constant for all five steps down.

Comparing Tables 1 and 2, it can be seen that as "in-tolerance" is changed from 3σ to 2.4σ, the percent of errors made increases. This is to be expected, because errors are more likely to be made when the observation is near the edge of the tolerance range, and when the tolerance range is reduced, a larger proportion of observations will be near the cutoff points. Comparing Tables 1 and 3, it can be seen that, as one would expect, if the standard is out-of-tolerance, more mistakes will be made. The farther out-of-tolerance the standard is, the more mistakes will be made. Finally, comparing Tables 1 and 4, one can see that as more instruments are out-of-tolerance coming in, fewer mistakes are made. While this may seem surprising at first glance, it makes sense because the way the model was defined, fewer observations are near the edge of the tolerance range. The farther the instruments drift out-of-tolerance, the easier it is to detect, and fewer mistakes will be made.

TABLE 1.- α AND β WHEN TOLERANCE = 3σ, STANDARD IS IN-TOLERANCE, AND 15% OF INSTRUMENTS COMING IN ARE OUT-OF-TOLERANCE

	Accuracy Ratio									
	1:1		2:1		3:1		4:1		5:1	
Step	α	β	α	β	α	β	α	β	α	β
1	.131	.043	.050	.026	.030	.022	.023	.017	.018	.015
2	.210	.052	.063	.033	.037	.023	.026	.016	.020	.014
3	.257	.054	.063	.031	.036	.025	.025	.017	.020	.015
4	.286	.056	.068	.033	.036	.022	.029	.022	.021	.013
5	.322	.058	.065	.034	.036	.022	.026	.016	.020	.017

	Accuracy Ratio									
	6:1		7:1		8:1		9:1		10:1	
Step	α	β	α	β	α	β	α	β	α	β
1	.014	.013	.011	.010	.01	.008	.010	.008	.012	.007
2	.015	.016	.013	.011	.011	.009	.010	.009	.009	.009
3	.016	.012	.013	.011	.012	.009	.008	.010	.009	.007
4	.018	.012	.013	.010	.012	.010	.010	.009	.008	.010
5	.017	.013	.013	.010	.012	.010	.010	.009	.010	.007

TABLE 2.- α AND β WHEN TOLERANCE = 2.4σ, STANDARD IS IN-TOLERANCE, AND 15% OF INSTRUMENTS COMING IN ARE OUT-OF-TOLERANCE

	Accuracy Ratio									
	1:1		2:1		3:1		4:1		5:1	
Step	α	β	α	β	α	β	α	β	α	β
1	.167	.055	.066	.035	.041	.027	.030	.022	.022	.017
2	.246	.058	.085	.039	.050	.031	.033	.021	.023	.018
3	.304	.067	.084	.038	.045	.030	.034	.021	.027	.019
4	.342	.059	.084	.041	.047	.027	.029	.020	.025	.019
5	.365	.055	.084	.038	.048	.029	.033	.022	.024	.016

	Accuracy Ratio									
	6:1		7:1		8:1		9:1		10:1	
Step	α	β	α	β	α	β	α	β	α	β
1	.020	.016	.016	.012	.017	.01	.012	.012	.011	.009
2	.019	.016	.016	.011	.016	.013	.011	.011	.012	.009
3	.020	.015	.019	.015	.015	.011	.014	.012	.011	.011
4	.019	.014	.017	.012	.014	.016	.015	.012	.011	.01
5	.017	.016	.013	.014	.016	.011	.014	.012	.011	.010

TABLE 3.- α AND β WHEN TOLERANCE = 3σ, STANDARD IS 50% OUT-OF-TOLERANCE, AND 15% OF INSTRUMENTS COMING IN ARE OUT-OF-TOLERANCE

	Accuracy Ratio									
	1:1		2:1		3:1		4:1		5:1	
Step	α	β	α	β	α	β	α	β	α	β
1	.171	.048	.074	.036	.040	.023	.028	.019	.024	.016
2	.262	.052	.081	.035	.047	.025	.033	.022	.028	.016
3	.319	.055	.083	.037	.045	.030	.034	.019	.026	.019
4	.364	.052	.087	.036	.047	.028	.036	.018	.025	.020
5	.394	.054	.085	.035	.044	.028	.034	.024	.026	.018

	Accuracy Ratio									
	6:1		7:1		8:1		9:1		10:1	
Step	α	β	α	β	α	β	α	β	α	β
1	.018	.013	.014	.014	.013	.011	.011	.009	.013	.008
2	.022	.017	.016	.013	.013	.012	.014	.011	.014	.008
3	.020	.014	.017	.013	.016	.012	.012	.012	.011	.008
4	.022	.015	.019	.011	.016	.012	.013	.011	.011	.009
5	.020	.016	.016	.014	.016	.012	.011	.01	.011	.011

TABLE 4.- α AND β WHEN TOLERANCE = 3σ, STANDARD IS IN-TOLERANCE, AND 32% OF INSTRUMENTS COMING IN ARE OUT-OF-TOLERANCE

	Accuracy Ratio									
	1:1		2:1		3:1		4:1		5:1	
Step	α	β	α	β	α	β	α	β	α	β
1	.110	.067	.047	.039	.037	.028	.023	.022	.019	.016
2	.180	.084	.067	.052	.039	.031	.029	.024	.022	.020
3	.212	.099	.066	.050	.038	.031	.030	.022	.021	.017
4	.247	.099	.068	.050	.037	.033	.029	.025	.021	.017
5	.273	.102	.060	.046	.038	.036	.029	.024	.023	.021

	Accuracy Ratio									
	6:1		7:1		8:1		9:1		10:1	
Step	α	β	α	β	α	β	α	β	α	β
1	.015	.014	.014	.013	.013	.013	.011	.010	.010	.010
2	.019	.018	.017	.013	.012	.011	.012	.013	.011	.011
3	.016	.018	.016	.013	.015	.013	.010	.013	.011	.010
4	.019	.016	.017	.016	.012	.011	.010	.010	.011	.011
5	.018	.017	.016	.016	.015	.014	.011	.011	.010	.011

CONCLUSIONS

When using the results of this simulation to determine the effect of accuracy ratios on the percent of instruments which are incorrectly determined to be in- or out-of-tolerance, it is necessary to remember that the results obtained apply only to the specific model that was used. While most of the assumptions seem fairly reasonable, the method of determining the means of the instrument measurement distributions may not closely approximate what is actually happening in practice. Furthermore, the definition of "in-tolerance" when using accuracy ratios is not the same as the standard definition. When using accuracy ratios, the instrument is said to be in-tolerance if the one observation taken is within a certain tolerance of the "true" value. However, the one observation taken may be an extreme value from the measurement distribution, and in fact most observations taken from that instrument would be outside of the tolerance. This possibility was not considered in this program.

With this in mind, it is now possible to answer the questions that were asked. First of all, the effect of lowering the accuracy ratio from 4:1 to 3:1 is, in almost all cases, an increase of approximately 0.01 in both α and β. While this may sound like a small price to pay, if one considers the percent increase in the number of errors made, it ranges from approximately 25% to 50%, depending on the parameter values--a substantial increase. Secondly, using a 1:1 accuracy ratio is substantially worse, and the magnitude of the errors increases as the progressive calibrations are made. The size of the errors made using the 1:1 ratio is almost certainly unacceptable in any situation.

Nevertheless, situations do exist where accuracy ratios of 1:1 to 3:1 are the best that can be maintained. What can be done in these situations? One suggestion is to abandon accuracy ratios altogether and keep track of the systematic and random errors at each successive calibration. The metrologists will immediately counter this suggestion with the fact that they have neither the time nor the personnel to do this for every instrument they calibrate.

Perhaps a compromise can be reached. Since the transfer and working standards are not calibrated very often, it may be practical to take the time to make repeated observations on them to obtain an accurate estimate of their uncertainties. This might lead to a smaller tolerance for the working standards than that obtained by using the accuracy ratios, which would in turn make larger accuracy ratios possible for calibrating customers' instruments. If the tolerance gets bigger instead of smaller, this would demonstrate that the problem with calibration is even more serious than previously believed. In either case, the MSCL would have a better idea of how accurate its calibrations are than is now known.

REFERENCES

Abernethy, R. B.; and Benedict, R. P.: Measurement Uncertainty: A Standard Methodology. *ISA Transactions*, vol. 24, no.1, 1985, pp. 75-79.

Cameron, Joseph M.: Measurement Assurance. *Journal of Quality Technology*, vol. 8, no. 1, Jan. 1986, pp. 53-55.

Capell, Frank: From 4:1 to SPC. Paper presented at the National Conference of Standards Laboratories 1988 Workshop and Symposium (Washington, D. C.), Aug. 14 - 18, 1988.

Capell, Frank: More on 'From 4:1 to SPC.' *National Conference of Standards Laboratories 1989 Workshop and Symposium Technical Presentations*, July 9-13, 1989, pp. 8-1 - 8-15.

Colclough, A. R.: Two Theories of Experimental Error. *Journal of Research of the National Bureau of Standards*, vol. 72, no. 3, May-June 1987, pp. 167-185.

Croarkin, Carroll: Measurement Assurance Programs Part II: Development and Implementation. National Bureau of Standards Special Publication 676-II, Apr. 1984.

Eisenhart, Churchill: Realistic Evaluation of the Precision and Accuracy of Instrument Calibration Systems. *Journal of Research of the National Bureau of Standards - C. Engineering and Instrumentation*, vol. 67C, no. 2, April-June 1963, pp. 161-187.

JSC Metrology Requirements Manual. JSC M 8070C, April 1990.

Ku, Harry H.: Notes on the Use of Propagation of Error Formulas. *Journal of Research of the National Bureau of Standards - C. Engineering and Instrumentation*, vol. 70C, no. 4, Oct.-Dec. 1966, pp. 263-273.

Ku, Harry H.; and Judish, Robert M.: Fundamentals of Error Analysis. *Proceedings of the IEEE*, vol. 74, no. 1, Jan. 1986, pp. 25-27.

Schumacher, Rolf B. F.: Systematic Measurement Errors. *Journal of Quality Technology*, vol. 13, 1981, pp. 10-24.

Schumacher, Rolf B. F.: MIL-STD 45662: Conflicting Positions on Out-of-Tolerance Measurements. *Proceedings of the 1988 Measurement Science Conference*, Jan. 1988, pp.99-105.