2. CONCEPTS OF ACCURACY, PRECISION, AND STATISTICAL CONTROL

Accuracy is an intuitively understandable and desirable requirement for measurements. Data which are known to be inaccurate or whose accuracy is unknown have little appeal to most users. Yet precision is sometimes confused with accuracy and the agreement of successive results can inspire a degree of confidence that the measurements may not merit.

Accuracy, the closeness of a measured value to the true value, includes the concepts of bias and precision and is judged with respect to the use to be made of the data. A measurement process must be unbiased to be capable of producing accurate values. It must be sufficiently precise, as well, or else the individual results will be inaccurate due to unacceptable variability. The following discussion is presented to clarify these concepts. The term uncertainty is used widely in describing the results of measurement and denotes an estimate of the bounds of inaccuracy. Strictly speaking, the actual error of a reported value is usually unknowable. However, limits of error ordinarily can be inferred, with some risk of being incorrect, from the precision and reasonable limits for the possible bias of the measurement process.

The concept of precision is concerned with the variability of the individual results of replicate measurements. A process which shows a small scatter is said to be precise and vice versa. Obviously such judgments are subjective and based on the intended use of the data. What might be considered as very precise for one purpose could be grossly imprecise for another. Under constant conditions, random errors are responsible for the observed scatter of measured values. These may be reduced to the point at which they are negligible with respect to the tolerable error of the measured value, or are limited by inherent characteristics of the instrumentation or the methodology used. The averages of several series of measurements will show a smaller variability than the individual values and the grand average of such is expected to approach a limiting value (limiting mean) as the number of measurements is increased.

The concept of bias is concerned with whether or not the limiting mean differs from the true (or accepted) value of the property measured. Here again, judgment is ordinarily involved since it is impossible to eliminate all error or even to know if this has been achieved. Such decisions are thus based on whether or not bias exists for all practical purposes.

In the case of individual measurements, each will exhibit some degree of inaccuracy, that is to say it will deviate from the true value. This will occur because of random error together with any bias of the measurement system. Indeed, it is highly improbable that any individual measurement made by an unbiased measurement system will be accurate, since the probability of zero random error is zero. Many individual values may appear to have the correct value but this is due to truncation resulting from insensitivity of the measurement process or from rounding of the data.

A measurement process should be sufficiently precise to minimize the number of replicate measurements required for the intended use. A very precise system may need only a few measurements, even one, to provide data that would not be significantly improved by further replication. Also, a measurement system must be sufficiently precise to identify whether or not biases of a

comparable magnitude are present in the system. While possible in principle, an unbiased measurement process of low precision may be incapable of providing accurate data, from a practical point of view, because of the large number of measurements required to reduce the uncertainty due to the random error to reasonable limits.

Precision may be evaluated by the redundant process of replicate measurement. Results on a single object or material may be used for this purpose, or the information obtained on a number of objects or samples (even duplicate measurements) may be pooled. Accordingly, there is no reason why a laboratory cannot evaluate its own precision without external assistance [25]. While reference standards may be helpful in this regard, they are not necessary for this purpose.

In order to properly estimate precision, a large number of measurements over an extended period of time are required. A small number of measurements tend to underestimate the standard deviation since small random errors are more probable than large ones and less likely to be observed during a limited set of measurements. Also, it is common experience that it is much easier to repeat a measurement on a given occasion than to reproduce it over a period of time. The repeatability, or short term standard deviation is needed to answer questions about the number of repetitive measurements that may be required while the long term standard deviation, or reproducibility is needed to answer such questions as the agreement of data obtained at different times, or the statistical control of a measurement process.

Though precise measurements can serve useful purposes when limited comparisons are required, accuracy is more often an essential requirement. Whenever the true value of the measured quantity is needed, or when data from different laboratories, different methodologies, or that from the same laboratory using the same method over a period of time needs to be interrelated, bias can be a serious problem. Bias of a measurement process can only be evaluated by comparison of a measured value with the "true" value of the parameter that is measured. This requires the use of a reference standard whose value is known within acceptable limits of uncertainty. Since the measured value typically will differ somewhat from the reference value, a statistically based decision must be made on the significance of any observed difference.

The precision of a measurement system may be influenced by a number of factors, each having its own precision. The precision of each factor, quantified in terms of the variance, contributes to the precision of the process. The variance is simply the square of the standard deviation, s. In measurement processes, the variances of the individual steps, s_i^2, add up to define the variance of the process, i.e., $s^2 = s_1^2 + s_2^2 + s_3^2 + \cdots + s_n^2$. Some of the steps (or factors) can be easily identified and the individual variances estimated. As steps are identifiable, improvements conceivably can be made when there are "assignable causes" for undesirable imprecision. Because of addition in quadrature, it is evident that one or a few sources of variance can be the major contributors to the total variance. Knowledge of the magnitude of the individual variances can indicate both directions for improvement and possible sources of trouble when "out-of-control" measurements occur.

It is conceivable that variance can be reduced to very low levels, with diligent effort. Laboratories commonly improve their precision as they gain experience with their methodology. Ordinarily, a laboratory will improve its quality control practices to the point where the precision attained is adequate for a particular application or when peer performance has been attained. Because measurement must be pragmatic, cost-benefit decisions will often dictate how far to go. For example, it is a matter of record that laboratories using the same methodology will differ in their precisions. This may be due to differences in levels of skill but also to different levels of tolerance for permissible error.

Unlike random errors, systematic errors or biases from several sources are not necessarily randomly distributed; hence one must consider that biases can add up algebraically. That is to say, the total bias $B = B_1 + B_2 + \cdots + B_n$. Thus, a large number of small biases can equal or even exceed a large bias from a single source. While the effect of random error decreases as the number of measurements, n, is increased ($s_{\bar{x}} = s_x/\sqrt{n}$), the effect of bias is independent of the number of measurements.

A stable measurement system is expected to generate reproducible data. Statistical control may be defined as the attainment of a state of predictability. Under such a condition, the mean of a large number of measurements will approach a limiting value (limiting mean) and the individual measurements should have a stable distribution, described by their standard deviation. Under such a condition, the limits within which any new measured value would be expected to lie can be predicted with a specified probability, the confidence limits for a measurement or mean of set of measurements can be calculated, and the number of measurements required to obtain a mean value with a given confidence may be estimated.

It is axiomatic that attainment of statistical control is the first objective of a measurement process. This is just another way of stating that it must achieve stability. Yet, it has the further connotation that the data produced are statistically describable. Eisenhart has stated -- "Until a measurement operation has been 'debugged' to the extent that it has attained a state of statistical control it cannot be regarded in any logical sense as measuring anything at all" (12).

When a measurement system is altered or disturbed, a new or modified measurement system may result with a limiting mean and/or a standard deviation different from the previous values. During normal use of a measurement system, changes can occur as well, unbeknown to the laboratory personnel. A well designed quality assurance program will monitor the system for such changes and indicate when corrective actions are required.

Modern quality assurance is based on the premise that measurement can be established as a process that can be in a state of statistical control, achievable by applying the principles of quality control [7]. The output of such a process can be described statistically and limits can be assigned for the confidence of single measurements. In addition, bounds for bias can be estimated intelligently; hence, limits of uncertainty can be established for the data. Such data can be used to make decisions with statistically supported confidence since its quality is known.

The uncertainty statement is a necessary and critical part of reporting the results of calibration. As stated by Croarkin [11]:

> The uncertainty statement assigns credible limits to the accuracy of the reported value stating to what extent that value may differ from its reference base. In practice it quantifies the magnitude of any possible discrepancy between the value actually obtained in the laboratory and the value which would be obtained at NBS for the same property of an object. An uncertainty provides both a measure of the worth of the values reported by the measurement laboratory and an estimate of the systematic error accruing to any organization that makes use of these values.

> The uncertainty statement is composed of (i) all sources of systematic error that contribute to the offset from the reference base and (ii) a limit to random error that quantifies the variability that is inherent in the measurement process as it transfers from a "known" or calibrated artifact or measurement system to an "unknown."

The estimate of the standard deviation of a measurement process is the appropriate statistic for quantifying random error. Limits to random error are computed so as to cover a large percentage of possible measurement outcomes; i.e., limits to random error can be computed at the 99.73 percent confidence level. In some applications, the factor three is sufficient to achieve this confidence level; for this handbook where results may depend on relatively small number of measurements, exact limits based on Student's t-distribution are recommended.

The systematic errors included in the uncertainty statements in the SOPs are based upon the uncertainties associated with the reference standards. It is assumed that other sources of systematic error are negligible. If this assumption is not true, the systematic error from other sources must be included in the uncertainty statement. Consequently, the SOPs recommend that the uncertainty associated with a reported value be computed as the sum of possible bias due to the uncertainty in the values reported for the reference standards plus the appropriate t-statistic for the desired confidence level times the standard deviation of the measurement process.

The uncertainty associated with a measurement reported by a laboratory may be used as part of the basis of quality assessment. The confidence level associated with an uncertainty statement permits a user of the data to under-stand "how good" the reported value is and whether or not it is "good enough" for the user's needs. The uncertainty statements from different laboratories can be compared for the same types of measurements. Since the uncertainty statement represents credible limits on both random and systematic errors, the overall measurement capability of different laboratories can be compared. The validity of reported uncertainty statements can then be evaluated by other means, such as round robin testing. If the limits for random and systematic errors comprising the uncertainty statement are reported separately, the relative magnitudes of these errors for different laboratories may be compared to determine how well each component is controlled. This may provide a useful insight into the quality of a laboratory's measurement process.