18.4.3.6 Control and testing of assumptions

"*Quality metrology* concerns itself with the control of measurements and their results which enter into examinations of the quality of materials, devices,…measuring instruments…" [OIML, *Vocabulary of Legal Metrology*, International Organization of Legal Metrology (1978)]. The testing of assumption validity for the *Chemical Measurement Process*, and thereby its results, necessarily constitutes a fundamental part of *Quality Metrology*. The control and assessment of imprecision and bias of the CMP -- ie, Quality Assurance -- is accomplished via assumption or *hypothesis testing*, where the null hypothesis is generally taken to be the absence of bias or of an added component of random error.

Assumption Testing

The principal concepts involved in the statistical theory of hypothesis testing are presented in section 18.4.3.7 with reference to analyte detection. Testing for bias or added imprecision rests upon the same principles. That is, one must postulate null (H_o) and alternative (H_A) hypotheses, and then define a test statistic and critical value, based upon the acceptable level for the error of the first kind α -- also known as the *significance level* of the test. The *power of the test*, which is described by its *operating characteristic* [OC curve], is defined as the probability of correctly "accepting" the alternative hypothesis, given α. The power is thus 1-β, where β is the probability of the error of the second kind.

Three points deserve emphasis:

(1) "Acceptance" of an hypothesis, based on such statistical testing must not be taken literally. More correctly, one simply fails to reject the hypothesis in question. For example, non-detection of an analyte does not prove its absence. Put another way, "acceptance" [non-rejection] may reflect inadequate power [1-β, given α] for the test and alternative hypothesis in question.

(2) Assumption (hypothesis) testing, itself, rests upon assumptions. The vast majority of statistical tests performed on the CMP and its results, for example, rely upon the assumptions of randomness and normality. Robust estimators and non-parametric or distribution-free tests may be employed when certain common assumptions may not be valid.

(3) Assumption tests emerge in many facets of chemical measurement, ranging from analyte detection [section 18.4.3.7], to tests of randomness and independence, to tests of means (and bias) using *z*- or *t*-statistics, and variance (and model) tests using χ^2 or *F* statistics. Such test statistics play a central role in maintaining CMP quality both within and among laboratories; the resultant quality assurance

is generally considered from the perspective of internal or external control (See also Sections 18.6 and 18.7, respectively).

Notes:

(1) Significance tests may be one-sided or two-sided. Testing for the presence of analyte in excess of the blank (detection test) is one- sided, since the true value of the net analyte concentration cannot be negative. Testing for the presence of bias, on the other hand, is generally two-sided.

(2) In many cases, such as testing for the presence of a particular analyte or the presence of systematic error, the null state cannot, in principle be attained; nor can the null hypothesis be proved. Recognizing the impossibility of attaining or proving absolute purity or absolute accuracy, it has been suggested that H_o be displaced from zero to an incremental value consistent with the relevant metrological objectives. In such circumstances, attention would be shifted for example from the Detection Limit to the *discrimination limit*, where the null state would be that characterized by a small, acceptable analyte concentration.

Internal Control

Within a given laboratory employing a given method of analysis, control of the CMP can be assessed in part by repeated measurements of test samples, such as *reference materials* (RM), having characteristics (composition) similar to the test samples of interest. The control in this case is limited to control of the mean (absence of trends, etc.) and control of the variance -- ie, the two quantities that reflect the stability of the CMP. *Control charts* are used to maintain a record of such internal control, where critical or control levels are derived from the mean and standard deviation (or ranges) of sets of observations. (At least four observations per set are advisable, to take advantage of the central limit theorem.) When *Certified reference materials (CRM)* or other materials of known composition are available, one may estimate bias as well, within the uncertainty bounds of the CRM. The procedures for accomplishing internal (and external) control, especially from the perspective of the CRM, have been documented by the International Organization for Standardization. (See also section 18.8.)

Repeatability

as measured by the *repeatability standard deviation*, is an accepted measure of internal variance. Its definition requires that "mutually independent test results [be] obtained with the same method on a test material in the same laboratory with the same equipment by the same operator within a short interval of time" (ISO 3534/1993). Thus, repeatability reflects the best achievable internal precision, and realistic uncertainty estimates must take into account possible variations in the constrained factors, as well as possible sources of uncompensated bias. Note that a false level of

precision (repeatability) ensues if the observations are not truly mutually independent. Successive readings from an instrument, for example, do not give a valid measure of repeatability for the CMP; rather, they are solely an indication of the instrumental repeatability. (See sections 18.2 and 18.4.4.)

External Control

Control may be assessed from without via "blind" replicates (for CMP stability) or "blind" CRMs (for CMP accuracy), submitted without foreknowledge of the measuring laboratory. A common failure of such external control is that the test samples are not totally blind. That is, the appearance or scheduling of the external samples may be sufficient to alert the internal analyst (possibly only subconsciously) to apply extra care, or even lack thereof. *Collaborative tests* comprise the other form of external control, where a number of (presumably) equivalent laboratories assay test portions from the same homogeneous material. ISO Guide 33 treats CMP assessment via an interlaboratory program; and ISO Guide 35 discusses this approach for the certification of CRMs. (See Section 18.7.)

Reproducibility,

as measured by the *reproducibility standard deviation*, is the external complement to repeatability. Conditions here are defined such that "test results are obtained with the same method on a test material in different laboratories with different equipment by different operators" (ISO 3534/1993). Thus, if the method in question is unbiased, reproducibility meets the objective of varying all factors so that the total error becomes random and thereby experimentally (statistically) estimable. In the International Vocabulary of Basic and General Terms in Metrology (see Section 18.2), the definition appears a little more flexible, in that a list of six types of changing factors is presented (including the method of measurement), accompanied by the notes that a specification of conditions actually subject to change should be indicated, and that the dispersion of results would serve as the quantitative measure of reproducibility.

Control, internal or external, need not be limited to measurement stability and accuracy. Control or assessment of assumed physical (or functional) models as well as random error models (cumulative distribution functions, autocovariance functions) may also be addressed. Both of these elements of modern multivariable and multicomponent measurements are leading to the emergence of a data analogue of Standard (Certified) Reference Materials (SRM), i.e., *standard test data (STD)*. STD, which represent fully characterized simulations of real analytical signals, have the great merit of providing quality assessment for the evaluation step of the CMP -- the step that is becoming at the same time more common and more complex and more remote from the direct control of the operator, through the advent of sophisticated computational and instrumentation modules. See Fig. 1 for a graphical representation of the CRM and STD control points for the Chemical Measurement Process.