OIV-MA-AS1-11 Harmonized guidelines for internal quality control in analytical chemistry laboratories

Contents

1. Introduction
    1. Basic concepts
    2. Scope of this document
    3. Internal quality control and uncertainties
2. Definitions
    1. International definitions
    2. Definition of terms specific to this document
3. Quality assurance practices and internal quality control
    1. Quality assurance
    2. Choice of analytical method
    3. Quality control and aptitude tests
4. Internal quality control procedures
    1. Introduction
    2. General approach. Statistical control
    3. Internal quality control and fitness for purpose
    4. The nature of errors
5. IQC and withing run precision
    1. Precision and duplication
    2. Interpretation of duplicate data
6. Control materials in IQC
    1. Introduction
    2. The role of certified reference materials

3. Preparation of control material
4. Blank determinations
5. Traceability in spiking and recovery checks

7. Recommendations
8. Conclusions
9. References

Appendix 1. Shewart control charts

1. Introduction

1.1. Basic concept

This document sets out guidelines for the implementation of internal quality control (IQC) in analytical laboratories. IQC is one of a number of concerted measures that analytical chemists can take to ensure that the data produced in the laboratory are fit for their intended purpose. In practice, fitness for purpose is determined by a comparison of the accuracy achieved in a laboratory at a given time with a required level of accuracy. Internal quality control therefore comprises the routine practical procedures that enable the analytical chemist to accept a result or group of results as fit for purpose, or reject the results and repeat the analysis. As such, IQC is an important determinant of the quality of analytical data, and is recognised as such by accreditation agencies.

Internal quality control is undertaken by the inclusion of particular reference materials, here called "control materials", into the analytical sequence and by duplicate analysis. The control materials should, wherever possible, be representative of the test materials under consideration in respect of matrix composition, the state of physical preparation and the concentration range of the analyte. As the control materials are treated in exactly the same way as the test materials, they are regarded as surrogates that can be used to characterise the performance of the analytical system, both at a specific time and over longer intervals.

Internal quality control is a final check of the correct execution of all of the procedures (including calibration) that are prescribed in the analytical protocol and all of the other quality assurance measures that underlie good analytical practice. IQC is therefore necessarily retrospective. It is also required to be as far as possible independent of the analytical protocol, especially the calibration, that it is designed to

test.

Ideally both the control materials and those used to create the calibration should be traceable to appropriate certified reference materials or a recognised empirical reference method. When this is not possible, control materials should be traceable at least to a material of guaranteed purity or other well characterised material. However, the two paths of traceability must not become coincident at too late a stage in the analytical process. For instance, if control materials and calibration standards were prepared from a single stock solution of analyte, IQC would not detect any inaccuracy stemming from the incorrect preparation of the stock solution.

In a typical analytical situation several, or perhaps many, similar test materials will be analysed together, and control materials will be included in the group. Often determinations will be duplicated by the analysis of separate test portions of the same material. Such a group of materials is referred to in this document as an analytical "run". (The words "set", "series" and "batch" have also been used as synonyms for "run".) Runs are regarded as being analysed under effectively constant conditions. The batches of reagents, the instrument settings, the analyst, and the laboratory environment will, under ideal conditions, remain unchanged during analysis of a run. Systematic errors should therefore remain constant during a run, as should the values of the parameters that describe random errors. .As the monitoring of these errors is of concern, the run is the basic operational unit of IQC.

A run is therefore regarded as being carried out under repeatability conditions, *i.e.*, the random measurement errors are of a magnitude that would be encountered in a "short" period of time. In practice the analysis of a run may occupy sufficient time for small systematic changes to occur. For example, reagents may degrade, instruments may drift, minor adjustments to instrumental settings may be called for, or the laboratory temperature may rise. However, these systematic effects are, for the purposes of IQC, subsumed into the repeatability variations. Sorting the materials making up a run into a randomised order converts the effects of drift into random errors.

1.2. Scope of this document

This document is a harmonisation of IQC procedures that have evolved in various fields of analysis, notably clinical biochemistry, geochemistry and environmental studies, occupational hygiene and food analysis(3-9). There is much common ground in the procedures from these various fields. Analytical chemistry comprises an even wider range of activities and the basic principles of IQC should be able to encompass all of these. The present document provides guidelines that will be applicable in most instances. This policy necessarily excludes a number of IQC practices that are

restricted to individual sectors of the analytical community. In addition in some sectors it is common to combine IQC as defined here with other aspects of quality assurance practice. There is no harm in such combination, but it must remain clear what are the essential aspects of IQC.

In order to achieve a harmonisation and provide basic guidance on IQC, some types of analytical activity have been excluded from this document. Issues specifically excluded are as follows.

(i) *Quality control of sampling.* While it is recognised that the quality of the analytical result can be no better than that of the sample, quality control of sampling is a separate subject and in many areas is not fully developed. Moreover, in many instances analytical laboratories have no control over sampling practice and quality.

(ii) *In-line analysis and continuous monitoring*. In this style of analysis there is no possibly of repeating the measurement, so the concept of IQC as used in this document is inapplicable.

(iii) *Multivariate IQC*. Multivariate methods in IQC are still the subject of research and cannot be regarded as sufficiently established for inclusion here. The current document regards multianalyte data as requiring a series of univariante IQC tests. Caution is necessary in the interpretation of this type of data to avoid inappropriately frequent rejection of data.

(iv) *Statutory and contractual requirements*.

(v) *Quality assurance measures* such as checks on instrumental stability before and during analysis, wavelength calibration, balance calibration, tests on resolution of chromatography columns, and problem diagnostics are not included. For present purposes they are regarded as part of the analytical protocol, and IQC tests their effectiveness together with the other aspects of the methodology.

1.3. Internal quality control and uncertainty

A prerequisite of analytical chemistry is the recognition of "fitness for purpose", the standard of accuracy that is required for an effective use of the analytical data. This standard is arrived at by consideration of the intended uses of the data although it is seldom possible to foresee all of the potential future applications of analytical results. For this reason in order to prevent inappropriate interpretation, it is important that a statement of the uncertainty should accompany analytical results, or be readily available to those who wish to use the data.

Strictly speaking, an analytical result cannot be interpreted unless it is accompanied

by knowledge of its associated uncertainty at a stated level of confidence. A simple example demonstrates this principle. Suppose that there is a statutory requirement that a foodstuff must not contain more than 10 µg g-1 of a particular constituent. A manufacturer analyses a batch and obtains a result of 9 µg g-1 for that constituent. If the uncertainty of the result expressed as a half range (assuming no sampling error) is 0.1 µg g-1 (i.e. the true result falls, with a high probability, within the range 8.9¤9.1) then it may be assumed that the legal limit is not exceeded. If, in contrast, the uncertainty is 2 µg g-1 then there is no such assurance. The interpretation and use that may be made of the measurement thus depends on the uncertainty associated with it.

Analytical results should therefore have an associated uncertainty if any definite meaning is to be attached to them or an informed interpretation made. If this requirement cannot be fulfilled, the use to which the data can be put is limited. Moreover, the achievement of the required measurement uncertainty must be tested as a routine procedure, because the quality of data can vary, both in time within a single laboratory and between different laboratories. IQC comprises the process of checking that the required uncertainty is achieved in a run.

2. Definitions

2.1. International definitions

Quality assurance. All those planned and systematic actions necessary to provide adequate confidence that a product or service will satisfy given requirements for quality(10).

Trueness: closeness of the agreement between the average value obtained from a large series of test results and an accepted reference value(11).

Precision: closeness of agreement between independent test results obtained under prescribed conditions(12).

Bias: difference between the expectation of the test results and an accepted reference value(11).

Accuracy: closeness of the agreement between the result of a measurement and a true value of the measurand(13).

Note 1. Accuracy is a qualitative concept.

Note 2. The term *precision* should not be used for *accuracy*.

Error: result of a measurement minus a true value of the measurand(13).

Repeatability conditions. conditions where independent test results are obtained with the same method on identical test items in the same laboratory by the same operator

using the same equipment within short intervals of time(11).

Uncertainty of measurement: parameter, associated with the result of a measurement, that characterises the dispersion of the values that could reasonably be attributed to the measurand(14).

Note 1. The parameter may be, for example, a standard deviation (or a given multiple of it), or the half-width of an interval having a stated level of confidence.

Note 2. Uncertainty of measurement comprises, in general, many components. Some of these components may be evaluated from the statistical distribution of results of a series of measurements and can be characterised by experimental standard deviations. The other components, which can also be characterised by standard deviations, are evaluated from assumed probability distributions based on experience or other information.

Note 3. It is understood that the result of a measurement is the best estimate of the value of a measurand, and that all components of uncertainty, including those arising from systematic effects, such as components associated with corrections and reference standards, contribute to the dispersion.

Traceability: property of the result of a measurement or the value of a standard whereby it can be related to stated references, usually national or international standards, through an unbroken chain of comparisons all having stated uncertainties(13).

Reference material: material or substance one of whose property values are sufficiently homogeneous and well established to be used for the calibration of an apparatus, the assessment of a measurement method, or for assigning values to materials(13).

Certified reference material: reference material, accompanied by a certificate, one or more of whose property values are certified by a procedure which establishes its traceability to an accurate realisation of the unit in which the property values are expressed, and for which each certified value is accompanied by an uncertainty at a stated level of confidence(13).

2.2. Definitions of terms specific to this document

Internal quality control: set of procedures undertaken by laboratory staff for the continuous monitoring of operation and the results of measurements in order to decide whether results are reliable enough to be released.

Control material: material used for the purposes of internal quality control and subjected to the same or part of the same measurement procedure as that used for

test materials.

Run (analytical run): set of measurements performed under repeatability conditions.

Fitness for purpose: degree to which data produced by a measurement process enables a user to make technically and administratively correct decisions for a stated purpose

Analytical system: range of circumstances that contribute to the quality of analytical data, including equipment, reagents, procedures, test materials, personnel, environment and quality assurance measures.

3. Quality assurance practices and internal quality control

3.1. Quality assurance

Quality assurance is the essential organisational infrastructure that underlies all reliable analytical measurements. It is concerned with achieving appropriate levels in matters such as staff training and management, adequacy of the laboratory environment, safety, the storage, integrity and identity of samples, record keeping, the maintenance and calibration of instruments, and the use of technically validated and properly documented methods . Failure in any of these areas might undermine vigorous efforts elsewhere to achieve the desired quality of data. In recent years these practices have been codified and formally recognised as essential. However, the prevalence of these favourable circumstances by no means ensures the attainment of appropriate data quality unless IQC is conducted.

3.2. Choice of analytical method

It is important that laboratories restrict their choice of methods to those that have been characterised as suitable for the matrix and analyte of interest. The laboratory must possess documentation describing the performance characteristics of the method, estimated under appropriate conditions.

The use of a method does not in itself guarantee the achievement of its established performance characteristics. There is, for a given method, only the potential to achieve a certain standard of reliability when the method is applied under a particular set of circumstances. It is this collection of circumstances, known as the "analytical system", that is therefore responsible for the accuracy of analytical data. Hence it is important to monitor the analytical system in order to achieve fitness for purpose. This is the aim of the IQC measures undertaken in a laboratory.

3.3. Internal quality control and proficiency tests

Proficiency testing is a periodic assessment of the performance of individual laboratories and groups of laboratories that is achieved by the distribution by an independent testing body of typical materials for unsupervised analysis by the participants(2). Although important, participation in proficiency testing schemes is not a substitute for IQC measures, or *vice versa*.

Proficiency testing schemes can be regarded as a routine, but relatively infrequent, check on analytical errors. Without the support of a well□developed IQC system, the value of participation in a proficiency test is negligible. Probably the main beneficial effect of proficiency tests is that of encouraging participants to install effective quality control systems. It has been shown that laboratories with effective IQC systems performed better in a proficiency testing scheme(15).

4. Internal quality control procedures

4.1. Introduction

Internal quality control involves the practical steps undertaken to ensure that errors in analytical data are of a magnitude appropriate for the use to which the data will be put. The practice of IQC depends on the use of two strategies, the analysis of reference materials to monitor trueness and statistical control, and duplication to monitor precision.

The basic approach to IQC involves the analysis of control materials alongside the test materials under examination. The outcome of the control analyses forms the basis of a decision regarding the acceptability of the test data. Two key points are worth noting in this context.

(i) The interpretation of control data must be based on documented, objective criteria, and on statistical principles wherever possible.

(ii) The results of control analyses should be viewed primarily as indicators of the performance of the analytical system, and only secondarily as a guide to the errors associated with individual test results. Substantial changes in the apparent accuracy of control determinations can sometimes be taken to imply similar changes to data for contemporary test materials, but correction of analytical data on the basis of this premise is unacceptable.

4.2. **General Approach □ Statistical Control**

The interpretation of the results of IQC analyses depends largely on the concept of

statistical control, which corresponds with stability of operation. Statistical control implies that an IQC result *x* can be interpreted as arising independently and at random from a normal population with mean □ and variance σ^2.

Under these constraints only about 0.27% of results (*x*) would fall outside the bounds of $\mu \pm 3\sigma$. When such extreme results are encountered they are regarded as being "out□of□ control" and interpreted to mean that the analytical system has started to behave differently. Loss of control therefore implies that the data produced by the system are of unknown accuracy and hence cannot be relied upon. The analytical system therefore requires investigation and remedial action before further analysis is undertaken. Compliance with statistical control can be monitored graphically with Shewhart control charts (see Appendix 1). An equivalent numerical approach, comparing values of $z = (x - \mu)/$ □ against appropriate values of the standard normal deviate, is also possible.

4.3. Internal quality control and fitness for purpose.

For the most part, the process of IQC is based on a description in terms of the statistical parameters of an ongoing analytical system in normal operation. Control limits are therefore based on the estimated values of these parameters rather than measures derived from considerations of fitness for purpose. Control limits must be narrower than the requirements of fitness for purpose or the analysis would be futile.

The concept of statistical control is inappropriate, however, when the so-called *ad hoc* analysis is being undertaken. In *ad hoc* analysis the test materials may be unfamiliar or rarely encountered, and runs are often made up of only a few such test materials. Under these circumstances there is no statistical basis for the construction of control charts. In such an instance the analytical chemist has to use fitness for purpose criteria, historical data or consistency with the visual properties of the test material for judging the acceptability of the results obtained.

Either way, agreed methods of establishing quantitative criteria to characterise fitness for purpose would be desirable. Unfortunately, this is one of the less-developed aspects of IQC. In specific application areas guidelines may emerge by consensus. For example, in environmental studies it is usually recognised that relative uncertainties of less than ten percent in the concentration of a trace analyte are rarely of consequence. In food analysis the Horwitz curve(16) is sometimes used as a fitness for purpose criterion. Such criteria have been defined for clinical analysis(17,18). In some areas of applied geochemistry a systematic approach has given rise to fitness for purpose criteria for sampling and analytical precisions. However, it is not practicable here to give guidelines in these areas, and at present no general principles can be

advanced that would allow specific applications to be addressed.

4.4. The nature of errors

Two main categories of analytical error are recognised, namely random errors and systematic errors, which give rise to imprecision and bias respectively. The importance of categorising errors in this way lies in the fact that they have different sources, remedies and consequences for the interpretation of data.

Random errors determine the precision of measurement. They cause random positive and negative deviations of results about the underlying mean value. *Systematic errors* comprise displacement of the mean of many determinations from the true value. For the purposes of IQC two levels of systematic error are worth consideration.

(i) *Persistent bias* affects the analytical system (for a given type of test material) over a long period and affects all data. Such bias, if small in relation to random error, may be identifiable only after the analytical system has been in operation for a long time. It might be regarded as tolerable, provided it is kept within prescribed bounds.

(ii) *The run effect* is exemplified by a deviation of the analytical system during a particular run. This effect, where it is sufficiently large, will be identified by IQC at the time of occurrence as an out-of-control condition.

The conventional division of errors between the random and the systematic depends on the timescale over which the system is viewed. Run effects of unknown source can be regarded in the long-term as the manifestation of a random process. Alternatively, if a shorter-term view is taken, the same variation could be seen as a bias-like change affecting a particular run.

The statistical model used for IQC in this document is as follows[1]. The value of a measurement (x) in a particular run is given by:

- x = true value + persistent bias + run effect + random error (+ gross error).

The variance of x (σ_x^2) in the absence of gross errors is given by:

$$\sigma_x^2 = \sigma_0^2 + \sigma_1^2$$

where

σ_0^2= variance of the random error (within run) and

σ_1^2= variance of the run effect.

The variances of the true value and the persistent bias are both zero. An analytical system in control is fully described by σ_0^2, σ_1^2 and the value of the persistent bias.

Gross errors are implied when the analytical system does not comply with such a description.

5. IQC and withing run precision

5.1. Precision and duplication

A limited control of within-run precision is achieved by the duplication within a run of measurements made on test materials. The objective is to ensure that the differences between paired results are consistent with or better than the level implied by the value of □o used by a laboratory for IQC purposes[2]. Such a test alerts the user to the possibility of poor within-run precision and provides additional information to help in interpreting control charts. The method is especially useful in *ad hoc* analysis, where attention is centred on a single run and information obtained from control materials is unlikely to be completely satisfactory.

As a general approach all of the test materials, or a random selection from them, are analysed in duplicate. The absolute differences $|d| = |x_1 - x_2|$ between duplicated analytical results x_1 and x_2 are tested against an upper control limit based on an appropriate value of σ_O. However, if the test materials in the run have a wide range of concentration of analyte, no single value of σ_O can be assumed(19).

Duplicates for IQC must reflect as far as possible the full range of variation present in the run. They must not be analysed as adjacent members of the run, otherwise they will reveal only the smallest possible measure of analytical variability. The best placing of duplicates is at random within each run. Moreover the duplication required for IQC requires the complete and independent analysis (preferably blind) of separate test portions of the test material. A duplication of the instrumental measurement of a single test solution would be ineffective because the variations introduced by the preliminary chemical treatment of the test material would be absent.

5.2. Interpretation of duplicate data

5.2.1. *Narrow concentration range.* In the simplest situation the test materials comprising the run have a small range of analyte concentrations so that a common within-run standard deviation σ_0 can be applied.

A value of this parameter must be estimated to provide a control limit. The upper 95% bound of $|d|$ is $2\sqrt{2}\sigma_O$ and on average only about three in a thousand results should exceed $3\sqrt{2}\sigma_O$.A group of *n* duplicated results can be interpreted in several ways.

For example, the standardised difference

$z_d = d/\sqrt{2}\sigma_O$

should have a normal distribution with zero mean and unit standard deviation. The sum of a group of *n* such results would have a standard deviation of $\sqrt{n}$ so only about three runs in a thousand would produce a value of $|\sum z_d| > 3\sqrt{n}$. Alternatively a group of *n* values of z_d from a run can be combined to form $\sum z_d^2$ and the result interpreted as a sample from a chi-squared distribution with *n* degrees of freedom, (x_n^2). Some caution is needed in the use of this statistic, however, as it is sensitive to outlying results.

5.2.2. *Wide concentration range.* If the test materials comprising a run have a wide range of analyte concentrations, no common standard of precision (σ_O) can be assumed. In such an instance σ_O must be expressed as a functional relationship with concentration. The value of concentration for a particular material is taken to be $(x_1 + x_2)/2$, and an appropriate value of σ_O obtained from the functional relationship, the parameters of which have to be estimated in advance.

6. **Control materials in IQC**

6.1. Introduction

Control materials are characterised substances that are inserted into the run alongside the test materials and subjected to exactly the same treatment. A control material must contain an appropriate concentration of the analyte, and a value of that concentration must be assigned to the material. Control materials act as surrogates for the test materials and must therefore be representative, *i.e.*, they should be subject to the same potential sources of error. To be fully representative, a control material must have the same matrix in terms of bulk composition, including minor constituents that may have a bearing on accuracy. It should also be in a similar physical form, *i.e.*, state of comminution, as the test materials. There are other essential characteristics of a control material. It must be adequately stable over the period of interest. It must be possible to divide the control material into effectively identical portions for analysis. It is often required in large amounts to allow its use over an extended period.

Reference materials in IQC are used in combination with control charts that allow both persistent bias and run effects to be addressed (Appendix 1). Persistent bias is evident as a significant deviation of the centre line from the assigned value. The variation in the run effect is predictable in terms of a standard deviation when the system is under statistical control, and that standard deviation is used to define action limits and warning limits at appropriate distances from the true value.

6.2. The role of certified reference materials

Certified reference materials (CRM) as defined in Section 2 (*i.e.*, with a statement of uncertainty and traceability), when available and of suitable composition, are ideal control materials in that they can be regarded for traceability purposes as ultimate standards of trueness(20). In the past CRMs were regarded as being for reference purposes only and not for routine use. A more modern approach is to treat CRMs as consumable and therefore suitable for IQC.

The use of CRMs in this way is, however, subject to a number of constraints.

(i) Despite the constantly increasing range of CRMs available, for the majority of analyses there is no closely matching CRM available.

(ii) Although the cost of CRMs is not prohibitive in relation to the total costs of analysis, it may not be possible for a laboratory with a wide range of activities to stock every relevant kind of reference material.

(iii) The concept of the reference material is not applicable to materials where either the matrix or the analyte is unstable.

(iv) CRMs are not necessarily available in sufficient amounts to provide for IQC use over extended periods.

(v) It must be remembered that not all apparently certified reference materials are of equal quality. Caution is suggested when the information on the certificate is inadequate.

If for any of the above reasons the use of a CRM is not appropriate it falls on individual laboratories or groups of laboratories to prepare their own control materials and assign traceable[3] values of analyte concentration to them. Such a material is sometimes referred to as a "house reference material" (HRM). Suggestions for preparing HRMs are listed in Section 6.3. Not all of the methods described there are applicable to all analytical situations.

6.3. Preparation of control materials

6.3.1. *Assigning a true value by analysis*. In principle a working value can be assigned to a stable reference material simply by careful analysis. However, precautions are necessary to avoid biases in the assigned value. This requires some form of independent check such as may be provided by analysis of the materials in a number of laboratories and where possible, the use of methods based on different physico-chemical principles. Lack of attention to independent validation of control materials has been shown to be a weakness in IQC systems(15).

One way of establishing a traceable assigned value in a control material is to analyse a

run comprising the candidate material and a selection of matching CRMs, with replication and randomisation. This course of action would be appropriate if only limited amounts of CRMs were available. The CRMs must be appropriate in both matrix composition and analyte concentration. The CRMs are used directly to calibrate the analytical procedure for the analysis of the control material. An appropriate analytical method is a prerequisite for this approach. It would be a dangerous approach if, say, a minor and variable fraction of the analyte were extracted for measurement. The uncertainty introduced into the assigned value must also be considered.

6.3.2. *Materials validated in proficiency testing* comprise a valuable source of control materials. Such materials would have been analysed by many laboratories using a variety of methods. In the absence of counter-indications, such as an obvious bias or unusual frequency distribution of results, the consensus of the laboratories could be regarded as a validated assigned value to which a meaningful uncertainty could be attached. (There is a possibility that the consensus could suffer from a bias of consequence, but this potential is always present in reference values.) There would be a theoretical problem of establishing the traceability of such a value, but that does not detract from the validity of the proposed procedure. The range of such materials available would be limited, but organisers of proficiency tests could ensure a copious supply by preparing batches of material in excess of the immediate requirements of the round. The normal requirements of stability would have to be demonstrable.

6.3.3. *Assigning a true value by formulation*. In favourable instances a control material can be prepared simply by mixing constituents of known purity in predetermined amounts. For example, this approach would often be satisfactory in instances where the control material is a solution. Problems are often encountered in formulation in producing solid control materials in a satisfactory physical state or in ensuring that the speciation and physical distribution of the analyte in the matrix is realistic. Moreover an adequate mixing of the constituents must be demonstrable.

6.3.4. *Spiked control materials*. "Spiking" is a way of creating a control material in which a value is assigned by a combination of formulation and analysis. This method is feasible when a test material essentially free of the analyte is available. After exhaustive analytical checks to ensure the background level is adequately low, the material is spiked with a known amount of analyte. The reference sample prepared in this way is thus of the same matrix as the test materials to be analysed and of known analyte level - the uncertainty in the assigned

concentration is limited only by the possible error in the unspiked determination. However, it may be difficult to ensure that the speciation, binding and physical form of the added analyte is the same as that of the native analyte and that the mixing is adequate.

6.3.5. *Recovery Checks*. If the use of a reference material is not practicable then a limited check on bias is possible by a test of recovery. This is especially useful when analytes or matrices cannot be stabilised or when *ad hoc* analysis is executed. A test portion of the test material spiked with a known amount of the analyte and analysed alongside the original test material. The recovery of the added analyte (known as the "marginal recovery") is the difference between the two measurements divided by the amount that is added. The obvious advantages of recovery checks are that the matrix is representative and the approach is widely applicable □ most test materials can be spiked by some means. However, the recovery check suffers from the disadvantage previously noted regarding the speciation, binding and physical distribution of the analyte. Furthermore, the assumption of an equivalent recovery of the analyte added as a spike and of the native analyte may not be valid. However, it can normally be assumed that a poor performance in a recovery check is strongly indicative of a similar or worse performance for the native analyte in the test materials.

Spiking and recovery testing as an IQC method must be distinguished from the method of standard additions, which is a measurement procedure: a single spiking addition cannot be used to fulfil the roles of both measurement and IQC.

6.4. Blank determinations

Blank determinations are nearly always an essential part of the analytical process and can conveniently be effected alongside the IQC protocol. The simplest form of blank is the "reagent blank", where the analytical procedure is executed in all respects apart from the addition of the test portion. This kind of blank, in fact, tests more than the purity of the reagents. For example it is capable of detecting contamination of the analytical system originating from any source, *e.g.*, glassware and the atmosphere, and is therefore better described as a "procedural blank". In some instances, better execution of blank determinations is achieved if a simulated test material is employed. The simulant could be an actual test material known to be virtually analyte-free or a surrogate (e.g., ashless filter paper used instead of plant material). Where it can be contrived, the best type of blank is the "field blank", which is a typical matrix with zero concentration of analyte.

An inconsistent set of blanks in a run suggests sporadic contamination and may add weight to IQC evidence suggesting the rejection of the results. When an analytical

protocol prescribes the subtraction of a blank value, the blank value must be subtracted also from the results of the control materials before they are used in IQC.

6.5. **Traceability in spiking and recovery checks**

Potential problems of the traceability of reagents used for spikes and recovery checks must be guarded against. Under conditions where CRMs are not available, traceability can often be established only to the batch of analyte provided by a manufacturer. In such cases, confirmation of identity and a check on purity must be made before use. A further precaution is that the calibration standards and spike should not be traceable to the same stock solution of analyte or the same analyst. If such a common traceability existed, then the corresponding sources of error would not be detected by the IQC.

7. Recommendations

The following recommendations represent integrated approaches to IQC that are suitable for many types of analysis and applications areas. Managers of laboratory quality systems will have to adapt the recommendations to the demands of their own particular requirements. Such adaption could be implemented, for example, by adjusting the number of duplicates and control material inserted into a run, or by the inclusion of any additional measures favoured in the particular application area. The procedure finally chosen and its accompanying decision rules must be codified in an IQC protocol that is separate from the analytical system protocol.

The practical approach to quality control is determined by the frequency with which the measurement is carried out and the size and nature of each run. The following recommendations are therefore made. The use of control charts and decision rules are covered in Appendix 1.

In each of the following the order in the run in which the various materials are analysed should be randomised if possible. A failure to randomise may result in an underestimation of various components of error.

(i) *Short (e.g., n<20) frequent runs of similar materials.* Here the concentration range of the

analyte in the run is relatively small, so a common value of standard deviation can be assumed.

Insert a control material at least once per run. Plot either the individual values obtained, or

The mean value, on an appropriate control chart. Analyse in duplicate at least half of

the

Test materials, selected at random. Insert at least one blank determination.

(ii) *Longer (e.g., n>20) frequent runs of similar materials*. Again a common level of standard deviation is assumed.

Insert the control material at an approximate frequency of one per ten test materials. If the run size is likely to vary from run to run it is easier to standardise on a fixed number of insertions per run and plot the mean value on a control chart of means. Otherwise plot individual values.

Analyse in duplicate a minimum of five test materials selected at random. Insert one blank

determination per ten test materials.

(iii) *Frequent runs containing similar materials but with a wide range of analyte concentration.* Here we cannot assume that a single value of standard deviation is applicable.

Insert control materials in total numbers approximately as recommended above. However, there should be at least two levels of analyte represented, one close to the median level of typical test materials, and the other approximately at the upper or lower decile as appropriate. Enter values for the two control materials on separate control charts. Duplicate a minimum of five test materials, and insert one procedural blank per ten test materials.

(iv) *Ad hoc analysis.* Here the concept of statistical control is not applicable. It is assumed, however, that the materials in the run are of a single type, *i.e.*, sufficiently similar for general conclusions on errors to be made.

Carry out duplicate analysis on all of the test materials.Carry out spiking or recovery tests or use a formulated control material, with an appropriate number of insertions (see above), and with different concentrations of analyte if appropriate. Carry out blank determinations. As no control limits are available, compare the bias and precision with fitness for purpose limits or other established criteria..

8. **Conclusions**

Internal quality control is an essential aspect of ensuring that data released from a laboratory are fit for purpose. If properly executed, quality control methods can monitor the various aspects of data quality on a run-by-run basis. In runs where performance falls outside acceptable limits, the data produced can be rejected and, after remedial action on the analytical system, the analysis can be repeated.

It must be stressed, however, that internal quality control is not foolproof even when

properly executed. Obviously it is subject to "errors of both kinds", *i.e.*, runs that are in control will occasionally be rejected and runs that are out of control occasionally accepted. Of more importance, IQC cannot usually identify sporadic gross errors or short□term disturbances in the analytical system that affect the results for individual test materials. Moreover, inferences based on IQC results are applicable only to test materials that fall within the scope of the analytical method validation. Despite these limitations, which professional experience and diligence can alleviate to a degree, internal quality control is the principal recourse available for ensuring that only data of appropriate quality are released from a laboratory. When properly executed it is very successful.

Finally, it must be appreciated that a perfunctory execution of any quality system will not guarantee the production of data of adequate quality. The correct procedures for feedback, remedial action and staff motivation must also be documented and acted upon. In other words, there must be a genuine commitment to quality within a laboratory for an internal quality control programme to succeed, *i.e.,* the IQC must be part of a total quality management system.

9. **References**

- "*Protocol for the Design, Conduct and Interpretation of Method Performance Studies*", Edited W Horwitz, *Pure Appl. Chem.,* 1988, **60**, 855- 864. *(Revision in press*)
- "*The International Harmonised Protocol for the Proficiency Testing of (Chemical) Analytical Laboratories*", Edited M Thompson and R Wood, *Pure Appl. Chem.,* 1993, **65**, 2123-2144. (Also published in *J. AOAC International*, 1993, **76**, 926-940.
- "*IFCC approved recommendations on quality control in clinical chemistry. Part 4: internal quality control*", *J. Clin. Chem. Clin. Biochem.,* 1980, **18**, 534-541.
- S Z Cekan, S B Sufi and E W Wilson, "*Internal quality control for assays of reproductive hormones: Guidelines for laboratories*". WHO, Geneva, 1993.
- M Thompson, "*Control procedures in geochemical analysis*", in R J Howarth (Ed*)*, "*Statistics and data analysis in geochemical prospecting*", Elsevier, Amsterdam, 1983.
- M Thompson, "*Data quality in applied geochemistry: the requirements and how to achieve them*", *J. Geochem. Explor.*, 1992, **44**, 3-22.
- Health and Safety Executive, "*Analytical quality in workplace air monitoring*",

London, 1991

- "*A protocol for analytical quality assurance in public analysts' laboratories*", Association of Public Analysts, 342 Coleford Road, Sheffield S9 5PH, UK, 1986.
- "*Method evaluation, quality control, proficiency testing*" *(AMIQAS PC Program),* National Institute of Occupational Health, Denmark, 1993.
- ISO 8402:1994. "*Quality assurance and quality management - vocabulary*".
- ISO 3534 -1: 1993 (E/F). "*Statistics, vocabulary and symbols - Part 1: Probability and general statistical terms*".
- ISO Guide 30:1992. "*Terms and definitions used in connections with reference materials*"
- "*International vocabulary for basic and general terms in metrology*" , 2nd Edition, 1993, ISO, Geneva.
- "*Guide to the expression of uncertainty in measurement*", ISO, Geneva, 1993.
- M Thompson and P J Lowthian, *Analyst*, 1993, **118**, 1495-1500.
- W Horwitz, L R Kamps and K W Boyer, *J. Assoc. Off. Anal. Chem.,* 1980, **63**, 1344.
- D Tonks, *Clin. Chem.,* 1963, **9**, 217-223.
- G C Fraser, P H Petersen, C Ricos and R Haeckel, "*Proposed quality specifications for the imprecision and inaccuracy of analytical systems for clinical chemistry*", *Eur. J. Clin. Chem. Clin. Biochem.,* 1992, **30**, 311-317.
- M Thompson, *Analyst*, 1988, **113**, 1579-1587.
- ISO Guide 33: 1989, "*Uses of Certified Reference Materials*", Geneva.

Appendix 1. Shewhart control charts

1. **Introduction**

The theory, construction and interpretation of the Shewhart chart(1) are detailed in numerous texts on process quality control and applied statistics, and in several ISO standards(2-5). There is a considerable literature on the use of the control chart in clinical chemistry(6,7). Westgard and co-workers have formulated multiple rules for the interpretation of such control charts(8), and the power of these results has been studied in detail(9-10). In this appendix only simple Shewhart charts are considered.

In IQC a Shewhart control chart is obtained when values of concentration measured on a control material in successive runs are plotted on a vertical axis against the run number on the horizontal axis. If more than one analysis of a particular control material is made in a run, either the individual results *x* or the mean value $\bar{x}$ can be used to form a control chart. The chart is completed by horizontal lines derived from the normal distribution N(μ,σ^2) that is taken to describe the random variations in the plotted values. The selected lines for control purposes are μ ± 2□ and μ ± 3□. Different values of □ are required for charts of individual values and of means. For a system in statistical control, on average about one in twenty values fall outside the μ ± 2 □ lines, called the "warning limits", and only about three in one thousand fall outside the μ ± 3□ lines, the "action limits". In practice the estimates $\bar{x}$and *s* of the parameters μ and □ are used to construct the chart. A persistent bias is indicated by a significant difference between $\bar{x}$ and the assigned value

2. Estimates of the parameters μ and σ

An analytical system under control exhibits two sources of random variation, the within-run, characterised by variance σ_0^2 and the between-run with variance σ_1^2. The two variances are typically comparable in magnitude. The standard deviation σ_x used in a chart of individual values is given by

$$\sigma_x = (\sigma_0^2 + \sigma_1^2)^{1/2}$$

whereas for a control chart of mean values the standard deviation is given by

$$\sigma_{\bar{x}} = (\sigma_0^2/n + \sigma_1^2)^{1/2}$$

where *n* is the number of control measurements in a run from which the mean is calculated. The value of *n* therefore must be constant from run to run, otherwise control limits would be impossible to define. If a fixed number of repeats of a control material per run cannot be guaranteed (*e.g.*, if the run length were variable) then charts of individual values must be used. Furthermore the equation indicates that σ_x or $\sigma_{\bar{x}}$ must be estimated with care. An attempt to base an estimate on repeat values from a single run would result in unduly narrow control limits.

Estimates must therefore include the between-run component of variance. If the use of a particular value of *n* can be assumed at the outset, then □$_{\bar{x}}$ can be estimated directly from the *m* means

$$\bar{x}_i = \sum_{j=1}^{n} x_{ij}/n$$

(i = 1,.....,m) of the n repeats in each of m successive runs.

Thus the estimate of μ is

$$\bar{x} = \sum_i \bar{x}_i/m$$

and the estimate of $\sigma_{\bar{x}}$ is

$$s_{\bar{x}} = \sqrt{\frac{\sum_i (\bar{x}_i - \bar{x})^2}{m-1}}$$

If the value of n is not predetermined, then separate estimates of $\sigma_0$and σ_1 could be obtained by one-way analysis of variance. If the mean squares within- and between-groups are MS_w and MS_b respectively, then

$\sigma_0^2$is estimated by MS_w and

σ_0^2 is estimated by $(MS_b - MS_w)/n$

Often in practice it is necessary to initiate a control chart with data collected from a small number of runs, which may be to a degree unrepresentative, as estimates of standard deviation are very variable unless large numbers of observations are used. Moreover, during the initial period, the occurrence of out-of-control conditions are more than normally likely and will produce outlying values. Such values would bias $\bar{x}$ and inflate s beyond its proper value. It is therefore advisable to recalculate $\bar{x}$ and s after a further "settling down" period. One method of obviating the effects of outliers in the calculation is to reject them after the application of Dixon's Q or Grubbs'(11) test, and then use the classical statistics given above. Alternatively, the methods of robust statistics could be applied to the data(12, 13).

3. **The interpretation of control charts**

The following simple rules can be applied to control charts of individual results or of means.

Single control chart. An out-of-control condition in the analytical system is signalled if any of the following occur.

(i) The current plotting value falls outside the action limits.

(ii) The current value and the previous plotting value fall outside the warning limits but within the actions limits

(iii) Nine successive plotting values fall on the same side of the mean line.

Two control charts. When two different control materials are used in each run, the respective control charts are considered simultaneously. This increases the chance of a type 1 error (rejection of a sound run) but decreases the chance of a type 2 error (acceptance of a flawed run). An out-of-control condition is indicated if any of the following occur.

(i) At least one of the plotting values falls outside the action limits.

(ii) Both of the plotting values are outside the warning limits.

(iii) The current value and the previous plotting value on the same control chart both fall outside the warning limits.

(iv) Both control charts simultaneously show that four successive plotting values on the same side of the mean line.

(v) One of the charts shows nine successive plotting values falling on the same side of the mean line.

A more thorough treatment of the control chart can be obtained by the application of the full Westgard rules, illustrated in Figure 2.

The analytical chemist should respond to an out-of-control condition by cessation of analysis pending diagnostic tests and remedial action followed by rejection of the results of the run and reanalysis of the test materials.

4. References

- W A Shewhart, "*Economic control of quality in manufactured product*", Van Nostrand, New York, 1931.
- ISO 8258:1991. "*Shewhart control charts*".
- ISO 7873:1993 "*Control charts for arithmetic means with warning limits*".
- ISO 7870:1993. "*Control charts - general guide and introduction*".
- ISO 7966:1993. "*Acceptance control charts*".
- S Levey and E R Jennings, *Am. J. Clin. Pathol.,* 1950, **20**, 1059-1066.
- A B J Nix, R J Rowlands, K W Kemp, D W Wilson and K Griffiths, *Stat. Med.*, 1987,**6**,425-440.
- J O Westgard, P L Barry and M R Hunt, *Clin. Chem.,* 1981, **27**, 493-501.
- C A Parvin, *Clin. Chem.*, 1992, **38**, 358-363.

- J Bishop and A B J Nix, *Clin. Chem.*, 1993, **39**, 1638-1649.
- W Horwitz, *Pure Appl. Chem.*, (in press).
- Analytical Methods Committee, *Analyst*, 1989, **114**, 1693-1697.
- Analytical Methods Committee, *Analyst*, 1989, **114**, 1699-1702.

Technical report from the Symposium on the 'Harmonisation of quality assurance systems for Analysis Laboratories, Washington DC, USA, 22-23 July 1993 sponsored by IUPAC, ISO et AOAC International

Prepared for publication by MICHAEL THOMPSON1 and ROGER WOOD2

1Department of Chemistry, Birkbeck College (University of London), London WC1H OPP, UK

2MAFF Food Science Laboratory, Norwich Research Park, Colney, Norwich NR4 7UQ, UK

1991-95 work group :

Chairman : M. Parkany (Switzerland) ; Membres : T. Anglov (Denmark) ; K. Bergknut (Norway and sweden) ; P. De Biève (Belgium) ; K.-G. von Boroviczény (Germany) ; J.M. Christensen (Denmark) ; T.D. Geary (South Australia) ; R. Greenhalgh (Canada) ; A.J. Head (United Kingdom) ; P.T. Holland (New Zealand) ; W. Horwitz (USA) . A. Kallner (Sweden; J. Kristiansen (Denmark) ; S.H.H. Olrichs (Netherlands) ; N. Palmer (USA) . M. Thompson (United Kingdom) ; M.J. Vernengo (Argentina) ; R. Wood (United Kingdom).

---

[1] The model could be extended if necessary to include other features of the analytical system

[2]There is no intention here of estimating the standard deviation of repeatability □*r* from the IQC data or of comparing estimates: there would usually be too few results for a satisfactory outcome. Where such an estimate is needed the formula $s_r = \sqrt{\sum d^2/2n}$ can be used.

[3]Where a CRM is not available traceability only to a reference method or to a batch of a reagent supplied by a manufacturer may be necessary.