1 Introduction

1.1 Historical perspective

The development of standard methods of analysis has been a prime objective of the Analytical Division of the Royal Society of Chemistry and its precursors, the Society of Public Analysts and the Society for Analytical Chemistry, since the earliest of days and the results of this work have been recorded in the pages of *The Analyst* since its inception in 1876. An ‘Analytical Investigation Scheme’ was proposed by A. Chaston Chapman in 1902. This later evolved into the Standing Committee on Uniformity of Analytical Methods and was charged with developing standard chemicals and securing comparative analyses of these standard materials.

In 1935, the Committee was renamed the Analytical Methods Committee (AMC) but the main analytical work was carried out by sub-committees composed of analysts with specialised knowledge of the particular application area. The earliest topics selected for study were milk products, essential oils, soap and the determination of metals in food colourants. Later applications included the determination of fluorine, crude fibre, total solids in tomato products, trade effluents and trace elements, and vitamins in animal feeding stuffs. These later topics led to the publication of standard methods in a separate booklet. All standard and recommended methods were collated and published in a volume entitled *Bibliography of Standard, Tentative and Recommended or Recognised Methods of Analysis* in 1951. This bibliography was expanded to include full details of the method under the title *Official, Standardised and Recommended Methods of Analysis* in 1976 with a second edition in 1983 and a third edition in 1994.

The work of the AMC has continued largely unchanged over the years with new sub-committees being formed as required and existing ones being disbanded as their work was completed. In 1995, the Council of the Analytical Division set in place a strategic review of the AMC in view of the changing need for approved analytical methods and the need to develop future direction for the AMC as it moves into the next millennium.

The aim of the AMC was reaffirmed to be participation in national and international efforts to establish a comprehensive framework for the appropriate quality in chemical measurements, which is to be realised by achieving five objectives:
The development, revision and promulgation of validated, standardised and official methods of analysis.

The development and establishment of suitable performance criteria for methods and analytical instrumentation/systems.

The use and development of appropriate statistical procedures.

The identification and promulgation of best analytical practices including sampling, equipment, instrumentation and materials.

The generation of validated compositional data of natural products for interpretative purposes.

1.2 Overview of the handbook

The objective for any analytical procedure is to enable consistent and reliable data of the appropriate quality to be generated by laboratories. Such procedures should be sufficiently well-defined and robust to ensure the best use of resources and to minimise the possibility of expensive large-scale collaborative trials yielding unsatisfactory results through lack of application of best practices. As part of achieving the objectives of the AMC it was felt that such a handbook would enable a consistency of approach to the work of the sub-committees.

Recently, major developments in statistical methods have been made particularly in the areas of collaborative studies and method validation and robustness testing. In addition, analytical method development and validation have assumed a new importance. However, this handbook is not intended to be a list of statistical procedures but rather a framework of approaches and an indication of where detailed statistical methods may be found. Whilst it is recognised that much of the information required is available in the scientific literature, it is scattered and not in a readily accessible format. In addition, many of the requirements are written in the language of the statistician and it was felt that a clear concise collation was needed which has been specifically written for the practising analytical chemist. This garnering of existing information is intended to provide an indication of current best practices in these areas. Where examples are given the intent is to illustrate important points of principle and best practice.

This handbook will be brief and pragmatic where possible. Inevitably, this will lead to contentious selections in parts. Consistency of a disciplined approach, however, is deemed more expedient than always espousing total scientific rigour.

1.3 Purpose and scope

The AMC identified the following four main objectives that this handbook should try to satisfy:

- Provision of a unified and disciplined framework that covers all aspects of the validation process from sample and method selection to full collaborative trial.
• Compilation of a selected bibliography of more detailed and specialist works to be used when appropriate and incorporating the work of the Statistical Sub-committee.

• Guidance in the use of the selected statistical procedures for the comparison of methods where circumstances and resources do not permit the meeting of the requirements of the IUPAC protocol.

• Illustration, by way of worked examples, of the main statistical procedures for the calculation, display and reporting of the results.

Analytical chemists are by nature innovators and seekers of improvement. In the development area these qualities are invaluable in optimising method performance. Alas far too often, this desire for continuous improvement spills over into the interpretation of methods for quality control. Here we require consistency of application and rigorous control of processes and procedures. These aspects are anathema for many practitioners of the 'art of chemical analysis'.

Whilst this may be sustainable (albeit undesirable) for some applications within a single laboratory, discipline becomes a necessity when methods have to be transferred reliably between laboratories in an organisation. When the scope of operation encompasses different organisations, national boundaries, etc., a uniformity of approach is essential if comparable results are to be obtained.

This discipline does not come easily, as it requires a control framework. The framework may be considered irksome and unnecessary by some analytical chemists, particularly those from a research environment. It is hoped to persuade those who doubt its necessity that the successful deployment of a method and its wide application rely heavily on such an approach and that flair and technical excellence alone are insufficient.

The foundations for the confidence in an analytical result require that:

• the sample is representative and homogeneous;
• the method selected is based upon sound scientific principles and has been shown to be robust and reliable for the sample type under test;
• the instrumentation used has been qualified and calibrated;
• a person who is both competent and adequately trained has carried out the analysis;
• the integrity of the calculation used to arrive at the result is correct and statistically sound.

This guide is concerned with establishing a control framework for the development and validation of laboratory-based analytical methods. Many of these methods will be employed in generating data that could have profound legal or commercial impacts. The validity of analytical results should be established beyond reasonable doubt.

Validation of an analytical method is not a single event. It is a journey with a defined itinerary and stopping places as well as a final destination.

The goal is a method that satisfies the original intent. A disciplined route is
Valid Analytical Methods and Procedures

Figure 1 ISO 'V' model adapted for analytical method validation

required which maps out the validation journey, more frequently called the validation process.

The ISO 'V' model for system development life cycle in computer software validation is a structured description of such a process. In this instance, the basic 'V' model has been adapted for analytical method validation and is shown in Figure 1.

Like all models, there are underlying assumptions. The main ones for analytical method validation include the areas of equipment qualification and the integrity of the calibration model chosen. If the raw analytical data are produced by equipment that has not been calibrated or not shown to perform reliably under the conditions of use, measurement integrity may be severely compromised. Equally, if the calibration model and its associated calculation methods chosen do not adequately describe the data generated then it is inappropriate to use it. These two areas are considered in some detail in Chapter 8.

Each layer of the ISO 'V' model is dependent upon the layer below and represents stages in the process. Broadly speaking, the boxes in the left-hand portion of the 'V' model represent the aims and objectives of the validation. The boxes in the right-hand portion of the 'V' model contain the processes and procedures that must be carried out successfully and be properly documented to demonstrate that these specified aims and objectives have been met. At the fulcrum of the model is the development process itself.

At each level of the model there is a horizontal correspondence between the two boxes. Verification of the matching of these pairs provides a method of closing the loop at each level.

For example, at the highest level, conformance to the user requirements specification may be verified through data generated in house, through limited laboratory trials or through use of the full IUPAC harmonised protocol. What is critical here is the confirmation of the original user requirements under appropriate performance conditions (Figure 2).
Confirmation of suitability for use with all sample matrices
- in house
- within organisations
- between organisations

**Figure 2** Mapping the user requirements specification to fitness for purpose
One useful approach to visualising these relationships is to list bullet points for each of the pairs in the manner shown below. In this way key areas are identified although there are not corresponding relationships between individual bullet points. Individual elements of the model are covered more fully in Chapter 7 where method validation is considered as a whole.

<table>
<thead>
<tr>
<th>Specified as</th>
<th>Established by</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Method applicability</td>
<td>• Selectivity/specificity</td>
</tr>
<tr>
<td>• Analytes to be quantified</td>
<td>• Linearity</td>
</tr>
<tr>
<td>• Ranges or limits specified</td>
<td>• Accuracy</td>
</tr>
<tr>
<td>• Methodology to be used</td>
<td>• Repeatability</td>
</tr>
<tr>
<td>• Sampling considerations</td>
<td>• Within-laboratory repeatability</td>
</tr>
<tr>
<td>• Matrices to be covered</td>
<td>• Reproducibility</td>
</tr>
<tr>
<td>etc.</td>
<td>• Recovery</td>
</tr>
<tr>
<td></td>
<td>• Robustness</td>
</tr>
</tbody>
</table>

Chapter 8 outlines basic aspects of data evaluation and manipulation. The important topic of linear calibration models is covered in some detail.

Recommended procedures for comparing methods and for taking a single method through to a full IUPAC collaborative trial with the harmonised protocol are covered in Chapter 9. Chapter 10 is a bibliography of recommended books and papers that should be consulted for more details in specific areas.

2 Nomenclature: Terms and Parameters

2.1 Introduction

To avoid confusion, the terms and parameters used in the validation of methods, for example, as used in Figure 3, must be clearly and unambiguously defined. This glossary contains the recommended definitions and corresponding descriptions and is based on the various standards and publications summarised in the Bibliography.¹ This is not exhaustive and it is recommended that the IUPAC 'Orange Book'² be consulted if required.

2.2 Terms

2.2.1 Analyte

Component or group of components of which the presence/absence or mass fraction/concentration is to be determined in the test sample.
2.2.2 Analysis

The method used in the detection, identification and/or determination of the analyte in a sample.

2.2.3 Laboratory sample

The sample or sub-sample(s) of the bulk of the material under consideration sent to or received by the laboratory.

2.2.4 Test sample

A representative quantity of material, obtained from the laboratory sample which is representative for the composition of the laboratory sample.
2.2.5 Test portion
The representative quantity of material of proper size for the measurement of the concentration or other property of interest, removed from the test sample, weighed and used in a single determination.

2.2.6 Observed value
The result of a single performance of the analysis procedure/method, starting with one test portion and ending with one observed value or test result. Note that the observed value may be the average of several measured values on the test portion (2.2.5) via the test solution (2.2.8) or aliquots (2.2.9).

2.2.7 Test result
The result of a complete test (frequently a combination of observed values).

2.2.8 Test solution
The solution resulting from dissolving the test portion and treating it according to the analytical procedure. The test solution may be used directly to determine the presence/absence or the mass fraction or mass concentration of the analyte without attributable sampling error. Alternatively, an aliquot (2.2.9) may be used.

2.2.9 Aliquot
A known volume fraction of the test solution (2.2.8) used directly to determine the presence/absence or the mass fraction/concentration of the analyte without attributable sampling error.

2.2.10 Detection
The determination of the presence of the analyte as a chemical entity.

2.2.11 Determination (quantification)
The determination of the absolute quantity of the analyte (mass, volume, mole) or the relative amount of the analyte (mass fraction, mass concentration) in the test sample.
2.2.12 Content mass fraction

The fraction of the analyte in the test sample. The mass fraction is a dimensionless number. However, the mass fraction is usually reported as a quotient of two mass-units or mass-volume.

<table>
<thead>
<tr>
<th>Value</th>
<th>Mass fraction (SI units)</th>
<th>Non SI units</th>
</tr>
</thead>
<tbody>
<tr>
<td>$10^{-2}$</td>
<td>% (m/m or m/v)</td>
<td>ppm, parts per million</td>
</tr>
<tr>
<td>$10^{-3}$</td>
<td>mg g$^{-1}$, mg mL$^{-1}$, g kg$^{-1}$, g L$^{-1}$</td>
<td>ppm, parts per million</td>
</tr>
<tr>
<td>$10^{-4}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$10^{-5}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$10^{-6}$</td>
<td>μg g$^{-1}$, μg mL$^{-1}$, mg kg$^{-1}$, mg L$^{-1}$</td>
<td>ppm, parts per million</td>
</tr>
<tr>
<td>$10^{-7}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$10^{-8}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$10^{-9}$</td>
<td>ng g$^{-1}$, ng mL$^{-1}$, μg kg$^{-1}$, μg L$^{-1}$</td>
<td>ppb, parts per billion</td>
</tr>
<tr>
<td>$10^{-10}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$10^{-11}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$10^{-12}$</td>
<td>pg g$^{-1}$, pg mL$^{-1}$, ng kg$^{-1}$, ng L$^{-1}$</td>
<td>ppt, parts per trillion</td>
</tr>
</tbody>
</table>

2.2.13 Mass concentration

The concentration expressed as the mass of the analyte in the test solution divided by the volume of the test solution. The term mass fraction should be used if the amount of the analyte is related to the mass of the sample.

2.1.14 Matrix

All components of the test sample excluding the analyte.

2.3 Parameters

2.3.1 Standard deviation(s)

A measure of the spread in the observed values as a result of random errors (2.3.12). These observed values all have the same expected value. The equation to be used is

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{X})^2} = \sqrt{\frac{1}{n-1} \left[ \sum_{i=1}^{n} x_i^2 - \frac{1}{n} \left( \sum_{i=1}^{n} x_i \right)^2 \right]}$$

where $x_i$ = individual measured value, $\bar{X}$ = mean measured value, $n$ = number of measurements.

Equation (1) applies to the calculation of $s_r$ (2.3.7), $s_{R_\infty}$ (2.3.8) and $s_R$ (2.3.9) under the measurement conditions specified therein.
2.3.2 Relative standard deviation(s) (RSD)

The standard deviation(s) expressed as a percentage of the mean value. The relative standard deviation is defined as:

$$RSD = \frac{s}{X} \times 100\%$$  \hspace{1cm} (2)

2.3.3 Detection limit

The calculated amount of the analyte in the sample, which according to the calibration line, corresponds to a signal equal to three times the standard deviation of 20 representative blank samples. A blank sample is a sample which does not contain the analyte.

If the recovery (2.3.21) of the analyte is less than 100\%, ideally the detection limit should be corrected for the average recovery. However, this is a contentious issue and needs to be considered carefully for each method.

2.3.4 Limit of quantification

The minimum content of the analyte in the test portion that can be quantitatively determined with a reasonable statistical confidence when applying the analytical procedure.

- Report the limit of quantification either in absolute quantities of the analyte (mass, volume or mole) relative amount of the analyte [mass fraction (2.2.12) or mass concentration; (2.2.13)].
- The amount of test portion (for example in grams) must be reported as used in the determination.

The limit of quantification is numerically equivalent to six times the standard deviation of the measured unit when applying the analytical procedure to 20 representative blank samples. For recoveries less than 100\% the limit of quantification must be corrected for the average recovery of the analyte.

2.3.5 Sensitivity

The change of the measured signal as a result of one unit change in the content of the analyte.

The change is calculated from the slope of the calibration line of the analyte.

2.3.6 Rounding off

The process of achieving agreement between an observed value and the repeatability (2.3.7) of the analytical procedure. The maximum rounding off interval is equal to the largest decimal unit determined to be smaller than half
the value of the standard deviation of the repeatability (2.3.7). See Section 8.3.1 for more details.

2.3.7 Repeatability \((r)\)
The expected maximum difference between two results obtained by repeated application of the analytical procedure to an identical test sample under identical conditions.

The measure for repeatability \((r)\) is the standard deviation \((s_r)\). For series of measurements of a sufficient size (usually not less than 6), the repeatability is defined as

\[
r = 2.8 \times s_r \text{ (confidence level 95%)}
\]

(3)

Repeatability should be obtained by the same operator with the same equipment in the same laboratory at the same time or within a short interval using the same method.

2.3.8 Within-laboratory reproducibility \((R_w)\)
The expected maximum difference between two results obtained by repeated application of the analytical procedure to an identical test sample under different conditions but in the same laboratory. The measure for the within-laboratory reproducibility \((R_w)\) is the standard deviation \((s_{R_w})\).

For series of measurements of sufficient size (usually not less than 6), the within-laboratory reproducibility is defined as

\[
R_w = 2.8 \times s_{R_w} \text{ (confidence level 95%)}
\]

(4)

Within-laboratory reproducibility should be obtained by one or several operators with the same equipment in the same laboratory at different days using the same method.

2.3.9 Reproducibility \((R)\)
The expected maximal difference between two results obtained by repeated application of the analytical procedure to an identical test sample in different laboratories. The measure for the reproducibility \((R)\) is the standard deviation \((s_R)\).

For series of measurements of sufficient size (usually not less than 6) the reproducibility is defined as

\[
R = 2.8 \times s_R \text{ (confidence level 95%)}
\]

(5)

Between-laboratory reproducibility should be obtained by different operators
with different instrumentation in different laboratories on different days using the same method.

For a given method, the most important factors in the determination of repeatability and reproducibility are Laboratory, Time, Analyst and Instrumentation.

<table>
<thead>
<tr>
<th>Experimental condition to determine</th>
<th>Factors to vary or control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Repeatability</td>
<td>Same L, T, A, I</td>
</tr>
<tr>
<td>Within-laboratory reproducibility</td>
<td>Same L; different T; I and A may be different</td>
</tr>
<tr>
<td>Between-laboratory reproducibility</td>
<td>Different L, T, A, I</td>
</tr>
</tbody>
</table>

If it is not possible to involve additional laboratories for the determination of the between-laboratory reproducibility, then the within-laboratory reproducibility may be used to get an estimate of the between-laboratory reproducibility. The reproducibility of the method may be dependent upon the mass fraction of the analyte in the test sample. It is therefore recommended, when studying the reproducibility, to investigate whether a relation exists between concentration and reproducibility. The measurement series should be greater than 8.

2.3.10 Trueness

The closeness of agreement between the average value obtained from a large series of test results and an accepted reference value. The measure of trueness is usually expressed in terms of bias.

2.3.11 Systematic error or bias

The difference between the average observed value, obtained from a large series of observed values \((n \geq 8)\), and the true value (2.3.13) (Figure 4).

2.3.12 Random error

The difference between a single observed value and the average value of a large number of observed values (at least 8), obtained by applying the same analytical procedure to the same homogeneous test sample.

2.3.13 True value

The value that describes the content and is completely defined by the circumstances under which the content has been determined.

2.3.14 Precision

A measure of the agreement between observed values obtained by repeated application of the same analytical procedure under documented conditions (2.3.7–2.3.9).
2.3.15 Accuracy
A measure of the agreement between a single analytical result and the true value.

2.3.16 Ruggedness
Ruggedness of an analytical method is the insensibility of the method for variations in the circumstance and the method variables during execution. It is important to note that statistically significant deviations are not always relevant. The purpose for which the measurement is made is a more important criterion for deciding on the ruggedness and the statistical method employed is merely a tool.

2.3.17 Noise
A phenomenon defined as fast changes in the intensity and frequency of a measured signal irrespective of the presence or absence of the analyte. The speed of change is significantly different from the normally expected detector response. A measure of noise is the measured difference between the highest and lowest value of the measured signal with no analyte present, observed in a relatively short time-span, as compared to the time-span necessary for measurement of the analyte.

2.3.18 Selectivity
A measure of the discriminating power of a given analytical procedure in differentiating between the analyte and other components in the test sample.
2.3.19 **Significant figures**

Values which contain the information consistent with either the repeatability or reproducibility of the analytical procedure. Significant values are obtained by using the described method for rounding off (Section 8.3.1).

2.3.20 **Specificity (see also Selectivity)**

The property of the analytical procedure to measure only that which is intended to be measured. The method should not respond to any other property of the analyte or other materials present.

2.3.21 **Recovery**

The fraction of the analyte determined in a blank test sample or test portion, after spiking with a known quantity of the analyte under predefined conditions.

- Recovery is expressed as a percentage.
- The part of the analytical procedure in which recovery is involved should be reported.
- The critical stages/phases relating to instability, inhomogeneity, chemical conversions, difficult extractions, etc. should be reported.
- Recovery must not be based on an internal standard unless work is undertaken to demonstrate identical behaviour under the conditions of test.

2.3.22 **Scope**

The collection of matrices and analyte(s) to which the analytical procedure is applicable.

2.3.23 **Range**

The content mass fraction interval to which the analytical procedure is applicable.

2.3.24 **Drift**

The phenomenon observed as a continuous (increasing or decreasing) change (slowly in time) of the measured signal in the absence of the analyte.
3 Samples and Sampling

3.1 Introduction

The importance of sampling in method validation and, in particular, inter-comparison of methods cannot be overemphasised. If the test portion is not representative of the original material, it will not be possible to relate the analytical result measured to that in the original material, no matter how good the analytical method is nor how carefully the analysis is performed. It is essential that the laboratory sample is taken from a homogeneous bulk sample as a collaborator who reports an outlying value may claim receipt of a defective laboratory sample. It is important to understand that sampling is always an error generating process and that although the reported result may be dependent upon the analytical method, it will always be dependent upon the sampling process.

The essential question in the inter-comparison of analytical methods is, 'If the same sample (or a set of identical aliquots of a sample) is analysed by the same method in different laboratories, are the results obtained the same within the limits of experimental error?'. It is apparent, therefore, that the selection of an appropriate sample or samples is critical to this question and that the sampling stage should be carried out by a skilled sampler with an understanding of the overall context of the analysis and trial.

Any evaluation procedure must cover the range of sample types for which the method under investigation is suitable, and details of its applicability in terms of sample matrix and concentration range must be made clear. Similarly, any restrictions in the applicability of the technique should be documented in the method.

For more details, the works listed in the Bibliography should be consulted. In particular, Crosby and Patel's *General Principles of Good Sampling Practice* and Prichard provide readily digestible guidance to current best practices in this area.

3.2 What is a sample?

The Commission on Analytical Nomenclature of the Analytical Chemistry Division of the International Union of Pure and Applied Chemistry has pointed out that confusion and ambiguity can arise around the use of the term 'sample' and recommends that its use is confined to its statistical concept. When being used to describe the material under analysis, the term should be qualified by the use of 'laboratory sample' or 'test sample', for example.

One of the best treatments of sampling terminology is given in recommendations published by IUPAC which describes the terms used in the sampling of bulk or packaged goods. In this example, the sampling procedure reduces the original consignment through lots or batches, increments, primary or gross samples, composite or aggregate samples, subsamples or secondary samples to a laboratory sample. The laboratory sample, if heterogeneous, may be further