Introduction to method validation

VAM
valid analytical measurement

dti
National Measurement System

Introduction to method validation

What is method validation?

Method validation provides documented objective evidence that a method measures what it is intended to measure, with acceptable performance parameters such as bias and precision. It is a continuation of method development and aims to assess and, if necessary, optimise method performance in a way that meets your customer's needs in a cost effective manner.

Definition of validation[1]

> *"**Confirmation**, by the examination and the provision of **objective evidence**, that the particular requirements for a **specific intended use** are fulfilled."*

Validation has three parts and when applied to method validation, these translate as:

1. The specific intended use is the analytical requirement which is set by the problem that the analysis is intended to solve.
2. The objective evidence is usually in the form of data from planned experiments, from which the appropriate method performance parameters are calculated.
3. The confirmation is taken as a satisfactory comparison of the method performance parameters with what is required, i.e. evidence that the method is fit for purpose.

Why is method validation necessary?

Method validation is an essential part of good measurement practice, because valid data can only be produced when the strengths and weaknesses of a method are understood. For the same reasons method validation is an important requirement of laboratory accreditation to ISO/IEC 17025:2005[1].

When do you need to validate methods?

Validation usually begins during the method development stage, when some performance parameters are evaluated approximately to determine whether the method's capabilities are in line with the levels required. If a previously validated method has not been used for a period of time, its performance should be checked (verified) before it is reinstated. If the scope of the method is altered, e.g. the method is to be applied to different sample types or analyte levels than it was originally validated for, the performance must be validated for the new type of sample. It is important to remember that you must verify the performance of all test methods before they are used for the analysis of customer samples, including published and standard methods which have been validated by others.

How do you validate methods?

Method validation should always be a planned activity. The first stage is to examine the problem. Look at the reasons behind carrying out the analysis and find out what it is that the customer hopes to establish from having the work carried out. It should then be possible to decide which method performance parameters are relevant to the work and the target values that are required for each parameter. A set of experiments can then be designed which can be used to evaluate the performance of the chosen method.

When planning a method validation study it is also important to consider how the experimental data will be analysed and the statistical tests that will be used to assess fitness for purpose.

The validation plan should comprise details of the material that is going to be analysed to assess each of the performance parameters, the number of replicate measurements required and the statistical analysis that will be carried out to evaluate the data. Several parameters may be examined in one set of experiments in which case the order in which things are done can be important.

Once the plan is formalised, experiments can be carried out to produce data to allow the method performance parameters to be evaluated. The resulting data are compared with the target values to determine if the method is fit for purpose.

How do you decide 'fitness for purpose'?

Once the method performance data have been generated and collated, an assessment can be made about whether the required performance target values have been met. If they have been achieved then the method can be declared fit for purpose and considered validated. If the target values are not achieved further development of the method will be necessary, followed by reassessment against the target values.

[1] ISO/IEC 17025:2005. General requirements for the competence of testing and calibration laboratories

The validation process

Method validation builds on the information obtained during method development. The process of validation is summarised in Figure 1.

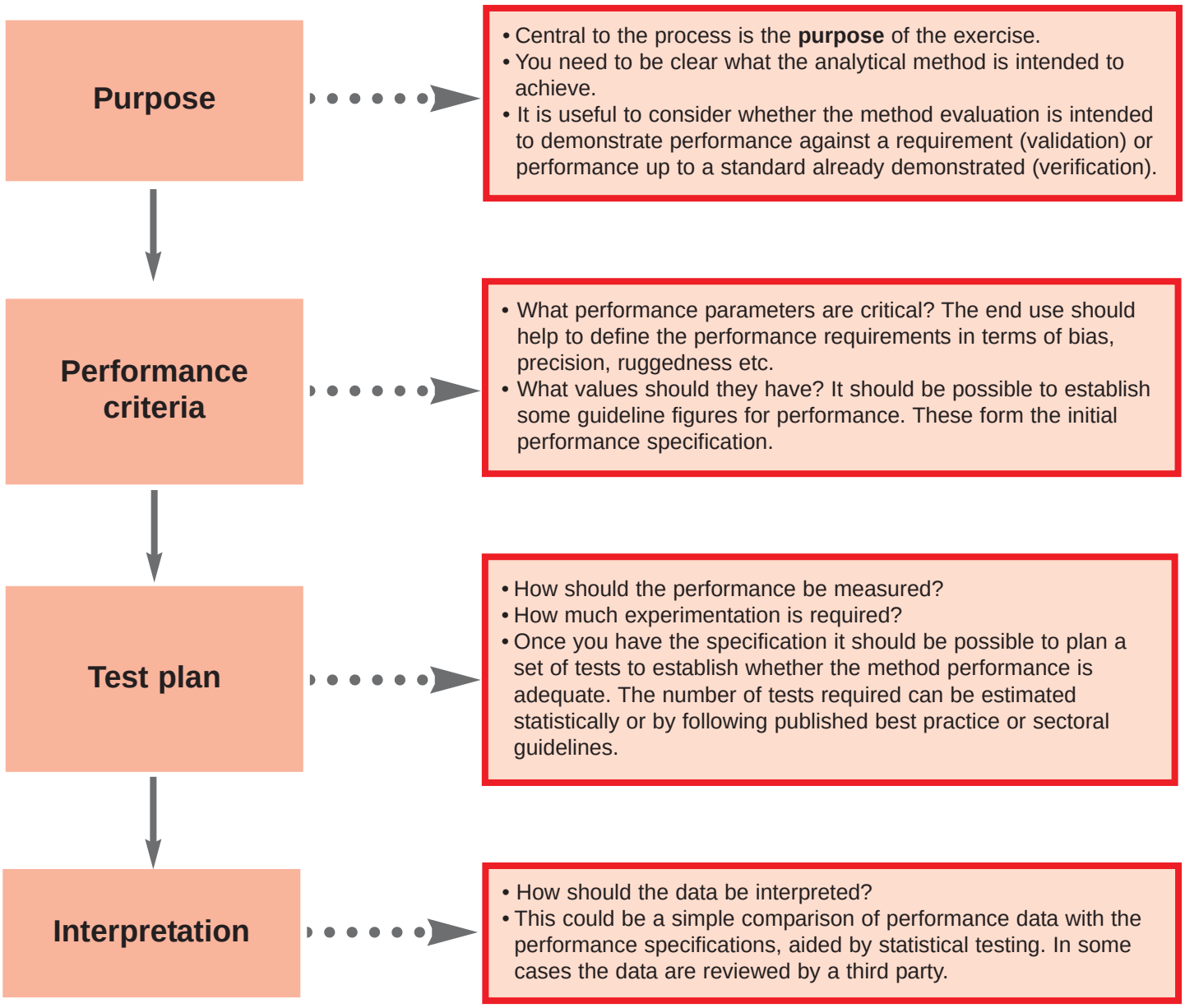

Figure 1: The validation process

Performance parameters

The key performance parameters that require attention during validation vary from one analytical requirement to another and from method to method, but some commonly important parameters are listed in the table below.

Parameter	Type of analysis			
	Qualitative	Major component	Trace analysis	Physical properties
Precision		✓	✓	✓
Specificity/selectivity	✓	✓	✓	✓
Bias		✓	✓	✓
Ruggedness	✓	✓	✓	✓
Linearity/working range		✓	✓	✓
Limit of detection	✓		✓	
Limit of quantitation			✓	

Different method performance parameters will be important in different situations. For example, accuracy (precision and bias) will be important for determining absolute values of properties or analyte concentration. For trace work, limits of detection and of quantitation are important but these parameters are less important if the analyte is present at higher concentrations. When planning calibration strategies it is important to know the range over which the response is linear. Ruggedness studies will indicate which parameters need to be controlled in order to preserve performance. The amount of validation required will also depend on the source of the test method and the extent of any previous validation studies. You are likely to have to carry out a significant amount of work to validate a method that you have developed yourself, compared to verifying the performance of a published test method that has already been validated by an interlaboratory study.

Precision

Definition of precision[2]

"The closeness of agreement between independent test results obtained under stipulated conditions."

Precision is a measure of the spread of results, i.e. how close a group of results are to each other. Precision gives no indication as to how close results are to the true value. The precision of an analytical method is evaluated by making repeat independent measurements on identical samples. From the observed spread of the results a value for the precision of the method can be calculated. It is usually expressed as a standard deviation or as a percent relative standard deviation. The magnitude of the precision is influenced by the size of random errors.

It is not necessary to know the exact concentration of the analyte in the sample used in a precision study. However, the material must be sufficiently stable and homogeneous and should be similar to the test sample in terms of the analyte concentration and matrix. The conditions under which the measurements are made determine the type of precision estimate obtained.

Repeatability represents the tightest extreme of independent replicate measurements. It describes the precision that you would expect for a set of replicate measurements made one after the other, in a single laboratory, by a single analyst on a single instrument. This type of precision study is useful for evaluating the likely variation in measurements made in the same batch of analyses.

Reproducibility represents the widest extreme of precision. It describes the variation within a set of measurements made on a sample over an extended time period, in several laboratories, by a number of different analysts on different instruments. You would expect reproducibility to reflect variation in the method from all possible sources. This type of precision represents the expected variation in results when a method is used to analyse a sample in several different laboratories.

Intermediate precision is a less widely accepted term compared to repeatability and reproducibility. It represents the variation in results obtained in a single laboratory over an extended time period. When a single laboratory uses several analysts or sets of equipment for a particular test method, intermediate precision has a great practical value. Compared to repeatability, intermediate precision is likely to give a better estimate of the precision of the method in routine use and is therefore the most appropriate precision value for setting quality control limits. There are various combinations of conditions that can lead to an estimate of intermediate precision so the conditions used in the study should always be stated. This type of precision estimate is sometimes referred to as, 'within laboratory reproducibility'.

If the method is to be applied to a range of sample types (e.g. different analyte concentrations or sample matrices) then the precision will need to be evaluated for a representative range of samples. In particular precision can vary significantly with analyte concentration.

Bias

Definition of trueness[2]

"The closeness of agreement between the average value obtained from a large set of test results and an accepted reference value."

with a note that

*"The measure of trueness is usually expressed in terms of **bias**."*

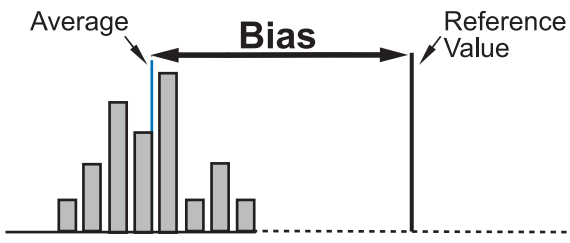

Figure 2: Bias

The bias of a method (as illustrated in Figure 2) is the difference between the average of a number of measured values and the true value. Bias is evaluated by carrying out replicate analysis of a sample with a known or accepted reference value, ideally a certified reference material (CRM), and comparing the average of the

[2] ISO 3534-1:1993 Statistics - Vocabulary and symbols - Part 1:Probability and general statistical terms

measured results with the reference value. If no suitable reference material is available, then it may be possible to evaluate bias by analysing a spiked sample (spiking involves adding a known amount of the analyte to a previously characterised sample). Bias can be expressed in a number of different ways. One way is simply to calculate the difference between the mean of the observed results and the expected result. This difference is frequently expressed as a percentage of the expected value. The difference is usually calculated so that higher observed results give a positive bias.

In some cases, it is more useful to describe bias in terms of the ratio of the mean of the observed results to the expected value. Analytical recovery is usually quoted as this ratio, multiplied by 100, to give 'percent recovery'.

As for precision, bias studies should cover the scope of the method.

Accuracy

Accuracy is defined as the closeness of agreement between a single test result and the accepted reference value. It should be noted that accuracy is the property of a single result and is influenced by both random and systematic errors. Results that are precise and unbiased are said to be accurate.

Ruggedness/robustness

Ruggedness/robustness is the degree to which a method is affected by small changes in the operating conditions. Changes in the conditions may be introduced when the method is used by different analysts or in different laboratories. When a method is intended to be used in different laboratories it should ideally be unaffected by changes in instruments, reagent supplier and environmental conditions. Ruggedness testing helps to identify those parameters which have a significant effect on the performance of a method, and how closely these identified parameters need to be controlled to avoid the performance of the method being affected. A rugged method is one whose performance is not affected by changes in the experimental parameters, within the defined control limits.

A ruggedness study involves designing experiments which deliberately introduce known changes to method parameters. The effect of changing these parameters is determined by analysing a sample under differing conditions defined by the experimental plan. The differences introduced for each parameter should reflect the likely variation which may occur during the normal operation of the method. The results obtained from a sample tested under differing conditions are evaluated statistically to determine whether any of the parameters have a significant effect on the results. The ruggedness test will identify critical parameters in the method which need to be further evaluated. These parameters will vary depending on the method being tested.

Specificity/selectivity

Specificity is the ability of a test method to unambiguously detect and measure the analyte of interest in the presence of other components that could be present in the sample matrix.

Selectivity and specificity are used synonymously and are rarely distinguished from each other. Specificity can be termed the ultimate in selectivity, i.e. if a process is specific it is 100% selective.

When using an analytical method, consideration should be given to the effect on the measurement of other components in the sample. Design of selectivity experiments requires background knowledge of the science of the method and the typical samples which will be analysed using the method. An ideal selectivity experiment should test the effects of all possible interferents on the typical observations, although this is rarely possible in practice.

Linearity and working range

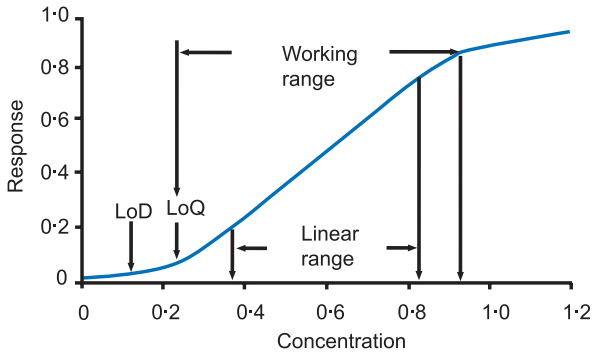

Figure 3: Linearity and working range

Linearity is the ability of the method to produce test results that are proportional to the analyte concentration within a given range (as illustrated in Figure 3).

During method validation it is necessary to establish the relationship between response and analyte concentration, this applies to both the response of the instrument used to measure the property of interest and of the test method as a whole. When using a specific principle of measurement, e.g. measurement using a particular instrument, it is important to evaluate its linearity characteristics across the concentration range of interest. This evaluation is made using pure chemical standards (i.e. at this stage it is not necessary to consider the effects of the sample matrix). In addition, the linearity of the whole test method should be studied. This is achieved through the analysis of CRMs, spiked samples or matrix matched standard solutions. The study

will reflect non-linearity due to, for example, interferences or incomplete recovery of the analyte from the sample. For both types of study the concentrations of the reference materials/standards should be evenly spaced across the range so as not to introduce bias into the experiment.

The information given by the two studies is complementary so it is worth doing both at the validation stage. The instrument study can be used for other applications, such as the same analyte in a different matrix, or different (but related) analytes.

The first stage of evaluating linearity should always be to plot the data. Having visually inspected the plot and made a subjective judgement that the method being validated is fit for purpose with regard to linearity, the data need to be supported with some statistics. The most commonly reported statistic is the correlation coefficient (*r*). This is a measure of the degree of association between the response and the analyte concentration.

The correlation coefficient is useful because it is easy to interpret; a high value (near 1, or near -1) means a good correlation. However it is easily misinterpreted because correlation and linearity are only loosely related, and spurious correlation is easy to generate with poorly chosen data points. *r* needs to be very close to 1 for reliable predictions from a calibration curve. For these reasons, the correlation coefficient should always be interpreted in conjunction with a plot of the data.

What is acceptable in terms of linearity depends on the intended use of the method. The uncertainty of the predicted values obtained from the calibration graph should be taken into account.

The working range is the interval between the upper and lower concentration of an analyte in the sample for which it has been determined that the method is suitable (i.e. the concentration range within which the results have an acceptable level of uncertainty, as illustrated in Figure 3). The upper boundary of the working range is defined by the concentration at which the sensitivity of the method decreases significantly (i.e. the point beyond which there is no longer sufficient change in response per unit change of concentration). The lower limit of the range is usually determined by the limit of quantitation, beyond which results cannot be determined with an acceptable uncertainty.

Limit of detection (LoD)

The limit of detection is the lowest concentration of an analyte in a sample that can be detected but not quantified (as illustrated in Figure 3). Terms such as minimum detectable value or detection limit are also used.

Detection limit calculations are based on statistical reasoning. The limit is set such that the probability of obtaining false positive results (i.e. declaring the analyte to be present when it is not) and false negative results (i.e. declaring the analyte to be absent when it is in fact present) is relatively low. A typical detection limit study involves carrying out replicate analysis of a blank sample or a sample containing a small amount of the analyte and calculating the standard deviation of the results. Approximate estimates of the limit of detection are typically obtained by multiplying the standard deviation by 3 or 4.65; these values are derived from statistical significance testing. Note that estimates of the limit of detection determined during method validation tend to be indicative. For typical in-house validation, approximate values are usually adequate; however, detection limits on which action depends should be rigorously checked and monitored frequently.

Note that in some sectors, for example clinical measurements, the term *sensitivity*, is used to describe the lower limit of application of a method.

Limit of quantitation (LoQ)

The limit of quantitation is the lowest concentration of an analyte in a sample that can be determined (quantified) with acceptable uncertainty under the stated operational conditions of the method (as illustrated in Figure 3). A quantitation limit attempts to identify the concentration below which a method becomes sufficiently unreliable to make quantification suspect.

Clearly, 'acceptable' quantitation is a matter of judgement. A common convention is to use $10s_o$ for the quantitation limit, where s_o is the standard deviation of the results obtained from the replicate analysis of a blank sample or a sample containing a small amount of the analyte.

Measurement uncertainty

Definition of measurement uncertainty[3]

"A parameter associated with the result of a measurement, that characterises the dispersion of the values that could reasonably be attributed to the measurand."

Measurement uncertainty is the extent to which the quoted result might reasonably differ from the true value. The measurement uncertainty is reported as x ± y where x is the result of the measurement and ± y is the degree of uncertainty. Uncertainty is usually calculated by creating an uncertainty budget, which is a list of all known, significant factors affecting analytical results obtained by a method, together with a quantitative estimate of their associated contribution to the overall

[3] Guide to the Expression of Uncertainty in Measurement, 1st edition, ISO (1995) (ISBN 92 67 10188 9)

uncertainty. This list forms the basis of the uncertainty estimation; the individual uncertainty estimates are combined using the appropriate mathematical rules to give the uncertainty in the final result.

An estimate of measurement uncertainty can be obtained using the data produced from method validation studies. This leaflet has shown that in-house method validation studies consist primarily of the determination of method performance parameters such as precision and bias. Uncertainty estimation from these parameters requires:

- The best available estimate of method precision (e.g. an estimate of the intermediate precision);
- The best available estimate(s) of method bias and its uncertainty;
- Quantification of any uncertainties associated with effects incompletely accounted for in the precision and bias studies.

For more information about measurement uncertainty please refer to the leaflet “What is measurement uncertainty?” in this series.

Evaluating and documenting the method and its validation

Method validation is carried out to provide objective evidence that a method is suitable for a given application. A formal assessment of the validation information, against the performance requirements specified at the start of the validation process is therefore required.

You should check that the relevant performance parameters described in this leaflet have been addressed adequately and that appropriate limits have been set for routine use of the method.

As mentioned previously, the extent of validation or verification will depend on previous experience and the origin/history of the method. For example, for verification of previously validated methods all that is needed is a check that the method works satisfactorily in your laboratory using your equipment and samples. However, for the validation of a new method, consideration would need to be given to evaluating all of the parameters mentioned in the leaflet.

When selecting and validating a method, compromise regarding what is possible and cost effective, may be necessary. Any such assessment needs to be documented.

It is important to document the validated method, in the form of a standard operating procedure for example, before it is used for the analysis of test samples.

Once validation is completed the quality control procedures are agreed and implemented. You need to document the validation (and verification) data and have this signed off as fit for purpose by a senior member of staff. This should be kept on file.

A system needs to be established to manage changes to the method, which may be necessary to meet new analytical requirements or to take advantage of technological developments. Validation, documentation and approval of the changes need to be addressed.

Further help from VAM

Valid Analytical Measurement (VAM) is a DTI funded programme which aims to:

- Improve the reliability of analytical measurements made in the UK;
- Facilitate mutual recognition of analytical data across international boundaries;
- Develop a robust and transparent infrastructure aimed at achieving international comparability and traceability of chemical measurements.

Further information about this programme can be obtained from the VAM website www.vam.org.uk. The website will also provide you with up to date information on current publications, new resources and VAM events.

Resources that are directly relevant to method validation include:

- In-House Method Validation – A Guide for Chemical Laboratories, LGC (2003) (ISBN 0 948926 18 X)
- mVal – software for method validation, LGC (2003)
- MV advisor web tool – available at www.vam.org.uk/mvadvisor
- The Fitness for Purpose of Analytical Methods, 1st edition, Eurachem (1998) (ISBN 0 948926 12 0)

www.vam.org.uk

Setting standards in analytical science

1432/RR/1006