Guidelines for the Development and Validation of Near-infrared Spectroscopic Methods in the Pharmaceutical Industry

Neville Broad (Pfizer), Paul Graham (Sanofi-Synthélabo, Chair), Perry Hailey (Pfizer), Allison Hardy (Eli Lilly), Steve Holland (AstraZeneca), Stephen Hughes, David Lee, Ken Prebble and Neale Salton (GlaxoSmithKline) and Paul Warren (Wyeth)

Guidelines for the Development and Validation of Near-infrared Spectroscopic Methods in the Pharmaceutical Industry

Neville Broad (Pfizer), Paul Graham (Sanofi-Synthélabo, Chair), Perry Hailey (Pfizer), Allison Hardy (Eli Lilly), Steve Holland (AstraZeneca), Stephen Hughes, David Lee, Ken Prebble and Neale Salton (GlaxoSmithKline) and Paul Warren (Wyeth)

Pharmaceutical Analytical Sciences Group, NIR Sub-Group, UK

Preface by Ken Leiper (Benson Associates, Grantham, UK)

PREFACE

In most industries new measurement technologies can be adopted provided a sound scientific rationale for the given application has been developed, proven and justified, and has been approved by internal company procedures.

However, in the highly regulated pharmaceutical industry, the competent regulatory agency, not the company, has the responsibility for approving the use of any measurement technology employed in any aspect of material or product release. Therefore, once the scientific criteria for a given application have been established and satisfied internally as above, the company must seek regulatory approval prior to routine implementation.

Obviously the regulatory agency involved must act independently, but there are practical difficulties to be overcome as expertise in the application of near-infrared (NIR) spectroscopy currently lies primarily in the user community rather than in the agency.

The lack of agreed "generic" validation guidelines for NIR has therefore been a major barrier within and across individual regulatory agencies in an increasingly globalized industry. Fortunately, the need to be in a position to address such issues using an unbiased scientific approach had been anticipated by the Pharmaceutical Analytical Sciences Group (PASG) (http://www.pasg.org.uk).

PASG is a forum for analytical scientists engaged in the management and practice of analytical science in chemistry and pharmacy disciplines within research, development and manufacturing functions of the research-based pharmaceutical industry, operating within the UK with the following aims and objectives:

- to act as a vehicle of communication for pharmaceutical analytical matters throughout the UK research-based industry;
- to act as a unified voice representing analytical science issues providing a focus through the Association of the British Pharmaceutical Industry (ABPI) to international regulatory agencies;
- to enhance the awareness of analytical science in education.

PASG co-ordinates its activities through specialist sub-groups to investigate, produce and publish best practices relating to analytical or technology policy with the objective of promoting good science and influencing the regulatory framework.

In this instance the PASG Near Infrared Sub-Group has collated and critically reviewed the experience of member companies in the successful development, validation, approval and implementation of NIR spectroscopy applications. This led to the authoring of comprehensive guidelines, which were subsequently circulated for expert peer review in the UK, Europe and the USA (see Acknowledgements). The resulting guidelines constitute this article.

The guidelines accurately reflect current best practice for the development and validation of reliable and robust NIR methods that have been approved by regulatory agencies to date, and so also establish the requirements for ensuring that NIR method development and validation will continue to be suitable for approval by regulatory agencies in future submissions. They are pitched at a level equivalent to similar guidelines on other topics produced in recent years by The International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) (http://www.ifpma.org/ich1.html). This is a unique project that brings together the regulatory authorities of Europe, Japan and the USA and experts from the pharmaceutical industry in the three regions to discuss scientific and technical aspects of product registration. The purpose of ICH is to make recommendations on ways to achieve greater harmonization in the interpretation and application of technical guidelines and requirements for product registration in order to reduce or obviate the need to duplicate the testing carried out during the research and development of new medicines. ICH guidelines Q2A and Q2B address traditional method validation requirements but do not address the unique and specific requirements for NIR method validation. These PASG guidelines cover the particular NIR requirements whilst remaining complementary to ICH Q2A and Q2B.

In the future, NIR offers the potential to move forward from traditional concepts of qualitative (e.g., identification) and quantitative (e.g. assay) methods to provide control of products and processes through qualification approaches and conformity concepts as mentioned in Section 1.1 of the guidelines. This evolution in the application of NIR methodology will be driven by novel measurement concepts that address the real needs of process control and improvements in efficiency and quality. Ultimately this in turn will support regulatory approval for shifts from conventional laboratory-based end-product testing towards material release based on process measurements made in the production area and within the process time envelope. The challenge for these guidelines will be to maintain them such that they continue to address the validation requirements of these evolving applications of NIR methodology.

1   INTRODUCTION

1.1   Background and purpose

The production of these Validation Guidelines stems from the recognition that an ICH approach to validation may not always be applicable to new technologies such as NIR spectroscopy.[1] Paradoxically in some cases an ICH approach may be suitable. A key aspect to resolving this paradox is in understanding new terminologies and how they relate to those described within ICH.[2]

The publication of the European Pharmacopoeia monograph on NIR[3] set the scene for pharmaceutical identity testing but provided limited guidance for the user in terms of developing an application. These guidelines attempt to go several steps further by providing the user, and the regulator, with a definitive guide to best practice for both qualitative and quantitative NIR method development, validation and application. The guidelines have been produced by the PASG NIR Sub-Group and have been reviewed by industrial and academic experts in pharmaceutical, statistical and chemometric disciplines.

NIR offers many advantages over pharmacopoeial methods by providing not only chemical but also physical information. This high-quality information can be obtained rapidly with little or no sample preparation – in stark contrast to many pharmacopoeial methods. The technique is applicable to both quantitative and qualitative applications and may be used throughout a process from input materials (actives and excipients), through intermediates to final products. The technique may be applied, in a laboratory or process environment, to individual components or to the matrix in its entirety.

Information from NIR data is generally accessed using mathematical techniques that offer an objective method of analysis. Generally a training set is developed that represents the chemical and physical characteristics of the material ("process signature"). The scope of any project is key and combined with experimental design should be used to determine appropriate samples, algorithms and pre-processing to build and validate a method for a particular application. NIR method development and validation should proceed in sequence through identification, qualification and quantification but a method can be applied at any of these three stages according to the scope defined (i.e. ensuring it is fit for purpose).

The availability of combined chemical/physical property information gives the user an understanding of the suitability of materials for a particular process and the potential to predict how well a particular material will perform. This is the essence of "conformity", where if a material's spectral signature falls within predefined statistical boundaries there

will be a high degree of confidence that the material will conform to specification. The conformity approach has not been addressed in this initial document on design, development and validation of NIR procedures, but it is at least as important as the qualitative and quantitative approaches covered herein, and needs to be addressed in subsequent updates/additions to this document. With NIR instrumentation being compatible with the process environment, measurement can be moved away from the laboratory, opening up the possibility of parametric release.

Some of these aspects are not covered in detail here, and future updates/editions to this document should address the following:

- development and validation of NIR methods for in-process/at-line/online applications;
- extension of the focus of the text from instruments with dispersive optics to those with other optical configurations, e.g. diode array, AOTF, FT etc.

These guidelines represent a distillation of ideas and best practices from contributors and their intended purpose is to provide a framework against which NIR methodologies can be developed and validated to a standard that will be approvable by regulatory agencies.

1.2 Overview

The guidelines covered in this document present a discussion of the characteristics for consideration during the design, development and validation of NIR methods included as part of registration applications for pharmaceutical compounds and preparations. Section 1 serves as a collection of terms, and their definitions. The terms and definitions are meant to bridge the differences that often exist between various compendia and regulators. Sections 2 and 3 provide direction on design, development and validation and routine use, i.e. what aspects and parameters need to be addressed and to what standards. More detailed and practical direction on how to accomplish validation has been covered by the work of various groups,[4,5] including the NIR Centre of Excellence at the University of London, School of Pharmacy.

The initial design requirements for the method should not be overlooked. It is necessary to define the purpose and scope of the intended method at the outset – essentially a "design qualification" stage – in order that the development and validation will be appropriate to provide a sound method for routine use. By this means, the application boundaries (e.g. compositional and process ranges) of the method are clearly established and scientifically justified, and will engender successful routine use.

The objective of the validation of a NIR method, in common with any analytical procedure, is to demonstrate that it is suitable for its intended purpose. A summary of the characteristics applicable to NIR identification, qualification and quantification approaches/procedures is provided in Table 1, and provides a link to ICH standard validation parameters.

Other NIR procedures/methods (e.g. the "conformity" approach) may be considered in future additions to this document.

The effective sample size in NIR methodology is generally significantly smaller than in conventional methods, and is often less than unit dose size. This is due not so much to the sample presentation accessories but to the area of the sample illuminated by the NIR beam. It must therefore be borne in mind that NIR is capable of detecting apparent heterogeneity, at least on a "micro" scale, and appropriate measures taken to accommodate this. For dose uniformity applications, this characteristic may provide a usable advantage, but in most applications some means of averaging the measurement over a larger sample area should be found. This may include transporting or spinning the sample through the NIR beam during spectral scanning.

1.3 Types of near-infrared procedures to be validated

The discussion of the validation of NIR methods/procedures is directed to the three most common types:

- Identification tests.
- Qualification tests for assurance of grade or fitness for intended use.
- Quantification procedures for particular ingredients in a material, whether they are the active moieties or impurities in samples of drug substance or drug product or other selected component(s) in the drug product.

A brief description of the types of test considered in this document is provided below:

- NIR Identification tests are intended to ensure either the identity of an analyte in a sample or, more usually, the identity of the whole sample matrix, and also to ensure discrimination of the material from other materials as defined in the scope of the method. This is normally achieved by comparison of the NIR spectrum (or mathematical/chemometric transformation of it) to that of a reference library set up using approved samples of the relevant materials as defined in the scope of the method. The design, development and validation of these methods are also included in Section 2.

Table 1. Validation requirements for NIR methods.

ICH Q2A validation parameter	NIR validation parameter and interpretation	Identification	Qualification	Quantification Major components	Quantification Impurities/minor components
Specificity[a]	As ICH Q2A and Q2B	+	+	+	+
Linearity[b]	With NIR spectra the data are typically multidimensional as opposed to one-dimensional data seen with conventional analytical methodology. Therefore with NIR the equivalent of linearity is the mapping of a calibration surface/volume as opposed to the construction of a single univariate calibration line. The validation for NIR therefore involves the demonstration of correlated NIR response to samples distributed throughout the defined range of the calibration model.	−	−	+	+
Range[b]	Defined by coverage of the product, process and material variability, which needs to be accommodated in the NIR method.	−	−	+	+
Accuracy	As ICH Q2A and Q2B. Usually demonstrated for NIR methods by correlation of NIR results with analytical reference data. NIR is often constrained (e.g. particularly for intact solid dosage forms) by the non-feasibility of performing recovery experiments.	−	−	+	+
Precision					
Repeatability	As ICH Q2A and Q2B	−[c]	+[c]	+	+
Intermediate precision	As ICH Q2A and Q2B, encompassing different analysts and different days but not necessarily instruments.	−	−	+[d]	+[d]
Robustness[e]	Robustness is inherently built into an NIR method during development by correct and appropriate sample selection/presentation (see technical guidelines), but still needs to be demonstrated in a similar way to conventional methods as described in ICH guidelines Q2A and Q2B.	+	+	+	+
Detection limit	As Q2A and Q2B	−	−	−	−
Quantification limit	As Q2A and Q2B, but constrained by lowest level available in sample calibration set.	−	−	−	+

−, Signifies that this characteristic is not normally evaluated.
+, Signifies that this characteristic is normally evaluated.
[a]Lack of specificity of the NIR procedure could be compensated by other supporting analytical procedure(s).
[b]Both linearity and range of a NIR method will be dependent upon the availability of samples representing product and process variations, in contrast to the fixed range (e.g. 80–120%), applied in validation of conventional methodology.
[c]Not normally required for identification methods. For qualification methods, repeatability is addressed in order to demonstrate that the acceptance thresholds established provide reliable discrimination between acceptable and unacceptable materials; the approach is therefore conceptually different for NIR methods compared with conventional methods.
[d]In cases where reproducibility (see Section 1.6) has been performed, intermediate precision is not needed.
[e]Robustness is not listed in this table in ICH Q2A; for conventional method validation, robustness is frequently assessed after the method has been developed, and may not be built in during method development.

- NIR Qualification tests are intended either to confirm the grade or source of material, which may in turn indicate its suitability for the intended use, or to discriminate between closely related materials that are indistinguishable by simpler identification testing. Qualification is a necessary prerequisite for admitting samples to a quantitative NIR method, since it provides assurance that the material belongs to the correct population and is eligible for measurement by the quantitative calibration set up for that population. The design, development and validation of these methods are also included in Section 2.
- NIR Quantification procedures are intended to measure the concentration of an analyte in a given sample. In the context of this document, the procedure represents a quantitative measurement of the major component(s), impurities and extraneous materials (e.g. water) in the drug substance or synthetic intermediates, or of the active ingredient(s), impurities or other selected component(s) in a drug product or intermediate product. The feasibility of impurity determinations is dependent upon sensitivity, e.g. the concentration levels of the impurities to be determined and the spectral response of individual impurities relative to that of the sample matrix. The same validation characteristics may also apply to methods designed to measure other properties of the sample (e.g. particle size). The design, development and validation of these methods are included in Section 3.

1.4 Validation requirements

The purpose of the analytical procedure should be clearly understood since this will govern the validation characteristics that will need to be evaluated. Although NIR is conceptually different from conventional analytical techniques such that validation is generally achieved through the assessment of specialized chemometric parameters, these parameters can still be related to the fundamental validation characteristics required for any analytical method:

- specificity
- linearity
- range
- accuracy
- precision
  - repeatability
  - intermediate precision
- robustness
- detection limit
- quantification limit.

Each of these validation characteristics, together with other NIR and chemometrics terms, is defined in Section 1.6. Table 1 lists those validation characteristics regarded as the most important for the validation of different types of analytical procedures, as defined in the ICH Q2A Guideline "Text on Validation of Analytical Procedures". The table is therefore applicable in principle to NIR methods. This list should be considered typical for the types of NIR procedures cited, but occasional exceptions should be dealt with on a case-by-case basis. It should be noted that robustness, although not listed in the table in ICH Q2A, is included in Table 1 and should be considered at an appropriate stage in the development of an NIR procedure.

Furthermore, revalidation may be necessary in the following circumstances:

- changes in the synthesis of the drug substance;
- changes in the composition of the finished product;
- changes in the finished product manufacturing process or sources/grades of ingredients;
- changes in the analytical procedure or the NIR instrumentation.

The degree of revalidation required depends on the nature of the changes. Certain other changes may require validation as well. Some guidance on this is provided in the associated Technical Guidelines sections for qualitative and quantitative methods.

1.5 Equipment

A typical NIR application, qualitative or quantitative, will include the following stages:

- equipment selection
- equipment qualification
- sample selection and presentation
- application development
- application validation
- application maintenance and change control.

This section covers equipment selection and validation, which are relevant to both qualitative and quantitative applications. The other stages will be covered separately for qualitative and quantitative applications in Sections 2 and 3, respectively.

1.5.1 Equipment selection

Spectrophotometers for recording spectra in the NIR region consist of:

- a filter, grating or interferometer system with a wavelength range in the region of 780–2500 nm ($12\,821$–$4000\,cm^{-1}$);

- a means of collecting and measuring the intensity of the transmitted or reflected light such as a conventional cuvette sample holder, a fiber-optic probe, transmission dip cells, etc., coupled to an appropriate detector typically of lead sulfide or indium gallium arsenide;
- a means of mathematical treatment of the spectral data obtained.

Typically the spectrophotometer should be capable of achieving the following specifications:

Wavelength accuracy	Better than ± 1 nm at 1200 nm and 1600 nm Better than ± 1.5 nm at 2000 nm
Photometric noise	
High light flux (100% vs 100%)	Average RMS 1200–2200 nm: <0.3 mAbs No RMS greater than 0.8 mAbs in a 100 nm range.
Low light flux (10% vs 10%)	Average RMS 1200–2200 nm: <1 mAbs No RMS greater than 2 mAbs in a 100 nm range.
Photometric linearity	Slope of 1 ± 0.05 and intercept 0.0 ± 0.05 when plotting the observed absorbance against expected absorbance at wavelengths 1200, 1600 and 2000 nm.

For some applications a lower specification may be appropriate. The selection of the equipment should be based on the intended application, paying particular attention to the suitability and compatibility of the sampling device with the type of samples to be analyzed.

1.5.2 Equipment qualification

The purpose of equipment qualification is to demonstrate, through documented evidence, that the equipment is suitable for its intended use.

The process of equipment qualification is described in a "Position Paper on the Qualification of Analytical Equipment",[6] agreed by the PASG. The various stages of equipment qualification can be defined as follows:

Design qualification (DQ)	defining the quality parameters required of the equipment and manufacturer.
Installation qualification (IQ)	assurance that the intended equipment is received as designed and specified.
Operational qualification (OQ)	confirmation that the equipment functions as specified and operates correctly.
Performance qualification (PQ)	confirmation that the equipment consistently continues to perform as required.

The process, as applicable to NIR equipment, can be conducted as described in the following sections.

Design qualification. DQ should provide documented evidence that quality has been built into the design of the application. Ensure that the NIR spectrophotometer is selected, taking into consideration the parameters detailed in equipment selection. As a minimum this should include a listing of the manufacturer's instrument specification.

In addition, it should be confirmed that the manufacturer has demonstrated compliance with an appropriate quality system, during development and manufacture, and that the software source code is lodged with a secure third party (Escrow agreement).

Installation qualification. Unpack the equipment and check against the order, ensuring that any predelivery qualification checks were made and recorded. Carry out a reconciliation of the equipment delivered against the delivery note and record instrument identification for each module, where appropriate, including serial numbers, firmware or software revision numbers.

Ensure that the environmental location and facilities are appropriate. Assemble the equipment and ensure that it powers up.

Operational qualification. Calibrate the equipment before commencing the OQ. Carry out qualification tests appropriate to the intended use of the NIR spectrophotometer. The following tests are a guide:

Wavelength accuracy	Verify the wavelength accuracy of the spectrophotometer using a suitable standard, for example NIST SRM 1920 at around 1200, 1600 and 2000 nm. The results should be within ± 1 nm at 1200 and 1600 nm and ± 1.5 nm at 2000 nm.
Wavelength repeatability[3]	Verify the wavelength repeatability of the spectrophotometer using a suitable standard, for example polystyrene or rare-earth oxides. The standard deviation of the wavelengths is consistent with the manufacturer's specification.

Response repeatability	Verify the response repeatability of the spectrophotometer using a suitable standard, for example reflective thermoplastic resins doped with carbon black. The standard deviation of the maxima response is consistent with the manufacturer's specification.
Photometric linearity	Verify the photometric linearity of the spectrophotometer by using a set of transmittance or reflectance standards (e.g. Spectralon™ or carbon black). Plot the observed response against the expected response. The slope of the line should be 1 ± 0.05 and the intercept 0.0 ± 0.05.
Photometric noise	Determine the photometric noise using a suitable reflective standard, for example white reflective ceramic tiles or reflective thermoplastic resins. Scan the reflection standard in accordance with the spectrophotometer manufacturer's recommendation and calculate the photometric noise, either peak to peak, or for a given wavelength, the standard deviation of the responses. The photometric noise is consistent with the manufacturer's specification.

Performance qualification. Carry out ongoing performance checks using some of the tests applied during OQ as applicable to the intended use of the spectrophotometer. This may be before use or on a regular program of checks, or combination of the two, as appropriate. Performance checks must be performed after maintenance or lamp changes. The tests must be scheduled and results fully documented.

1.5.3 Change control

All changes must be controlled by an appropriate change control system.

Hardware. Any changes to the system hardware (i.e. spectrometer and computer system) arising from either maintenance or modifications should be reviewed against the original IQ/OQ/PQ criteria and appropriate action and testing completed to ensure the instrumentation operates in an equivalent or improved manner. Examples include lamp change, sample introduction, presentation device, and change in location or environment.

Software. Any changes to the software (including changes in version number) should be reviewed against the original IQ/OQ/PQ and computer validation criteria. Appropriate action and testing should be completed to ensure the instrumentation operates in an equivalent or improved manner.

1.6 Glossary

The definition of the analytical terms Accuracy, Precision, Repeatability, Intermediate Precision, Specificity, Detection Limit, Quantification Limit, Linearity and Range are defined in the ICH Guidance Document, Q2A Text on Validation of Analytical Procedures (http://www.ifpma.org/ich5.html).

Calibration set. Sample set incorporating all chemical and physical variation normally encountered during routine manufacture, used to generate and optimize a regression model. The samples must cover the range required by the method.

Calibration test set. A subdivision of the calibration set used to internally assess and verify the calibration model.

Identification. A spectral match value (SMV) obtained against a single reference spectrum. The reference spectrum may be derived from the mean of a number of spectra.

Mahalonobis distance. A multidimensional vector that describes the distance of a point from the mean of a multidimensional ellipse.

Multiple linear regression (MLR). An inverse calibration method in which a small number of variables is used to develop a regression model. An example is where a full wavelength instrument is used in a feasibility study but where a filter instrument (limited wavelength(s)) is to be employed.

NIST SRM. National Institute of Standards and Technology Standard Reference Material.

Principal component analysis (PCA). A mathematical manipulation of a data matrix where it is possible to describe the variation in the data with a smaller number of orthogonal components.

Principal component regression (PCR)/partial least squares (PLS) regression. PCR and PLS are inverse calibration methods where it is possible to calibrate for the desired component without selecting variables and accounting for

sources of variation within the data. The methods are factor based and use different methods to select the factors.

Qualification. An SMV obtained against a distribution derived from a number of spectra.

Standard normal variate (SNV). A data pretreatment tool that standardizes NIR spectra using the mean and standard deviation of the spectra.

Spectral match value (SMV). An identification method to match NIR spectra based upon a comparison of spectral vectors.

Soft independent modeling by class analogy (SIMCA). A classification technique based upon PCA.

Validation set. A sample set, separate from the calibration set, used to give an independent assessment of the accuracy and precision of the calibration model.

1.7 References

These references are not intended to be an exhaustive list of chemometric publications but rather as a guide for future reading.

1.7.1 Books

K.R. Beebe, R.J. Pell and M.B. Seasholtz, 'Chemometrics – A Practical Guide', J. Wiley & Sons, Chichester (1998).

H. Martens and T. Naes, 'Multivariate Calibration', J. Wiley & Sons, Chichester (1991).

E.R. Malinowski, 'Factor Analysis in Chemistry', J. Wiley & Sons, New York (1991).

1.7.2 Useful reference journals

Journal of Chemometrics, J. Wiley and Sons.

Chemometrics and Intelligent Laboratory Systems, Elsevier.

Journal of Near Infrared Spectroscopy, NIR Publications.

Applied Spectroscopy, Society for Applied Spectroscopy.

1.7.3 Useful papers

ICH Quality Guidelines: Q2A, Text on Validation of Analytical Procedures; Q2B, Validation of Analytical Procedures: Methodology (http://www.ifpma.org/ich5.html) (1996).

'Near-infrared Spectrometry, in Methods of Analysis 2.2.40' in "European Pharmacopoeia", 3rd edition, Council of Europe, Strasbourg, 43–44 (1997).

W. Plugge and C. Van Der Vlies, *J. Pharm. Biomed. Anal.*, **11**, 435 (1993).

A.C. Moffat, A.D. Trafford, R.D. Jee and P. Graham, 'Meeting the ICH Guidelines on Validation as Exemplified by a Near-infrared Reflectance Assay of Paracetamol in Intact Tablets', *The Analyst*, **125**, 1341 (2000).

M. Freeman, M. Leng, D. Morrison and R.P. Munden, 'Position Paper on the Qualification of Analytical Equipment', *Pharmaceut. Technol. Eur.*, **November**, 40 (1995).

W.L. Yoon, R.D. Jee and A.C. Moffat, 'Optimisation of Sample Presentation for the Near-infrared Spectra of Excipients', *The Analyst*, **123**, 1029 (1998).

2 TECHNICAL GUIDELINES FOR QUALITATIVE METHODS

2.1 Introduction to qualitative analysis

NIR spectroscopy can be used for both *identification* and *qualification*. The selection of samples and the subsequent degree of library development will depend on the complexity of the application. These guidance notes apply to both identification and qualification.

- Identification. This is typically used where the chemical identity only of the material is required.
- Qualification. This is usually performed after chemical identification has been ascertained, and measures how well a sample under test fits in with a model of the material. This model is derived from samples chosen to represent the natural acceptable variability of that material. These variations may include moisture, particle size, solvents and other chemical and physical properties.

For both identification and qualification, discrimination between materials in the library must be shown.

A typical qualitative NIR application will include the following stages:

- equipment selection (Section 1.5.1)
- equipment qualification (Section 1.5.2)
- feasibility study
- sample selection and presentation
- library development
- library validation
- library maintenance and change control.

2.2 Feasibility study

Prior to method development it is recommended that an initial feasibility study is performed, for example to find

optimal presentation requirement (rapid and reproducible), amount of sample and minimum number of scans for efficient analysis. A prior knowledge of the composition and molecular structure of the library samples is always beneficial to ensure NIR is the most suitable choice of analysis.

2.3 Sample authentication, collection and measurement

The selection of samples is critical to the success of the qualitative application. A strategy for selection and measurement is required, considering the points given below (see Sections 2.3.1 and 2.4.2–2.4.4). Two sets of samples will be required: one for construction of the library and an independent one for validation purposes.

All samples used to build and validate the database must have some level of authentication. The extent of this will depend on the specific application (e.g. supplier's Certificate of Analysis, internal identity tests, selected chemical/physical tests, full specification tests in house). Consideration should also be made for multiple sourcing of materials in building and validation of databases.

Different batches should be collected over a period of time to cover changes in constituent concentration, supplier, process changes or variations in storage conditions. It is possible to use retained samples providing this is justified with respect to their chemical/physical stability during storage.

The number of batches required to train the system will depend on the complexity of the analysis and must collectively describe the typical variation of the substance being analysed. Associated with these batches should be samples which are known to be outside of specification and/or different but closely related batches. This should be defined by the user as part of the method validation. Identification will normally require a small number of different batches. A larger number will be required for qualification.

2.3.1 Sample measurement/presentation

There are many accessories available for the presentation of NIR samples (e.g. cups, vials, fiber-optic probes and custom-made sampling accessories). The choice of presentation will depend on the users' requirements and should be defined in the DQ stage. Each of these will have its advantages and limitations.

Sample presentation is a potential source of variation during sample measurement and should be as consistent and reproducible as possible. Potential variation should be assessed during the method robustness experiments and

clearly documented.[7] Care should be taken to ensure the sample presentation device is adequately cleaned between measurements.

The effective sample size in NIR methodology is generally significantly smaller than in conventional methods, and is often less than unit dose size. This is due not so much to the sample presentation accessories but to the area of the sample illuminated by the NIR beam. It must therefore be borne in mind that NIR is capable of detecting apparent heterogeneity, at least on a "micro" scale, and appropriate measures taken to accommodate this. For dose uniformity applications, this characteristic may provide a usable advantage, but in most applications some means of averaging the measurement over a larger sample area should be found. This may include transporting or spinning the sample through the NIR beam during spectral scanning.

Measurement by transmission.
Liquids and solutions   Common presentations include fixed-pathlength NIR transparent cells or fiber-optic pairs. The sample can be measured using a suitable pathlength (generally 0.5–4 mm). Care should be taken to avoid the presence of air bubbles and the sample temperature should be kept constant whenever possible. A reference scan of air is required for liquid samples. A solvent reference can be used for solution samples.

Solids   Samples of tablets or powder can be measured by transmission in a close-fitting template or automation tray. The template should minimize positioning errors and light scatter. An appropriate reference spectrum should be recorded.

Measurement by diffuse reflection. Solid samples can be measured by diffuse reflection. The sample can be measured in sample cups, disposable vials or using a reflection probe either for direct scanning by insertion into the sample or indirect scanning through a packaging material. Care should be taken to ensure consistent packing against the optical surface. Consideration should be given to packing density, sample depth, probe pressure, sample cell variation and coverage of the detector window. An appropriate reflectance reference should be used (e.g. ceramic, Spectralon™).

Measurement by transflection. Liquid samples and solid suspensions may be measured by transflection, i.e. a combination of reflection and transmission. Samples can be presented using an inert diffuse reflector (e.g. dispersed titanium dioxide or a reflecting metal surface). Care should be taken to avoid the presence of air bubbles and the sample temperature should be kept constant whenever possible. An appropriate reference dependent on the presentation method should be recorded.

2.4   Library development

An overview of the activities associated with library development is shown in the flowchart in Figure 1.

A qualitative library incorporates the training data set for each material grouping, any subsequent transformation and discriminatory analysis. These will depend on the intended use of the library. A typical qualitative library development will involve the following stages:

- define the purpose of the library
- selection of samples/spectra for calibration set
- data pre-processing/transformation
- library construction
- determination of thresholds.

2.4.1   Define the purpose

It is important to define the scope of the library in terms of its intended use prior to starting development. This can be either identification alone or for identification and qualification. Consideration should be given to the chemical similarity and numbers of material groups to be discriminated.

2.4.2   Selection of samples/spectra for calibration set

Spectral data should be acquired for the calibration sets according to guidance specified in the sample selection/presentation section.

Sample variability due to the following factors may be built into the library. This is especially important for qualification libraries:

- moisture
- particle size
- residual solvents
- degradation products
- compositional change of formulated product
- other chemical/physical properties
- time (constituent/process changes and also instrument variation)
- alternative sources of material
- retained samples
- temperature (especially liquids)
- operator
- presentation (e.g. probe insertion)
- between-instrument variation
- others.

These factors, and the extent to which they are considered, depend upon the intended scope and the required discriminatory powers of the method.

2.4.3   Display data

It is important to visually examine all spectra to check for absence of anomalies or the presence of outliers. Potential outliers must be investigated and can only be excluded for valid analytical reasons, and any exclusion must be documented.
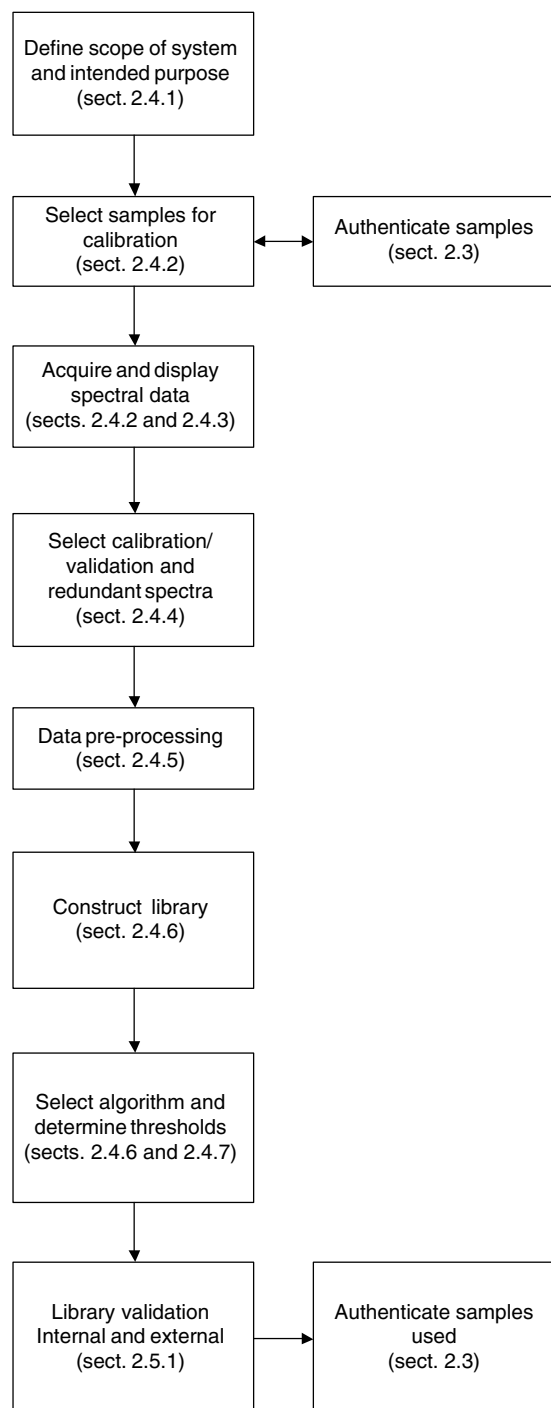

Figure 1. NIR library development activities.

2.4.4 Calibration set selection

It may be desirable to select representative samples from a larger population. In the simplest case a visual assessment may be sufficient. However for more complex situations the use of sample selection tools may be useful in determining the membership of a material group (e.g. PCA or cluster analysis).

The number of samples required for each material group will depend on the discriminant algorithm used and the complexity of the application, i.e. how accurately the group boundaries need to be determined.

2.4.5 Data pre-processing/transformation

It is often necessary to mathematically treat the data to reduce spectral complexity. For example derivatives and scatter correction algorithms may be used to reduce offsets due to physical characteristics. The use of raw data may be more applicable to qualification applications where effects due to physical form are important.

Caution should be exercised when performing any mathematical transformation as artefacts can be introduced or essential information can be lost. An understanding of the algorithm is required and in all cases the rationale for the use of the transform should be documented.

2.4.6 Library construction

The library structure may be dependent on the limitations of the software and the user's requirements. In the simplest case all materials may be incorporated into one library. Alternatively they may be split into sublibraries to ensure the required level of specificity.

All material groups in the main library must have the same mathematical transform. Transforms must be the same within each sublibrary but may be different from each other, an example of which is when performing qualification after identification, e.g. subclassification of grades of lactose from the main excipient library.

The full or a reduced instrument wavelength range may be used. A reduced range may be necessary due to the use of sampling accessories or the removal of irrelevant data (e.g. areas that exceed the dynamic range, high noise areas, etc.). Wavelength segments can also be useful to remove unwanted effects or to enhance small but important differences.

Similar to other techniques, NIR may not be able to discriminate between all material groups, e.g. close members of a homologous series. In these cases it may be necessary to merge the two groups into one and use other methods of control to determine the specific identity/quality of the material.

Many algorithms exist, for example correlation, SIMCA, Mahalanobis distance, and SMV. The choice is dependent on the user, considering the scope of the library, as mentioned in Section 2.3.1. However, it is recommended that the simplest available algorithm that can be clearly defined and gives adequate discrimination should be used. For example, for identification only, where physical factors are not to be determined, a match by wavelength correlation method using second derivative data should suffice.

2.4.7 Determination of thresholds

Initially, internal validation (see Section 2.5.1) should be performed using the software default values, or those recommended by the manufacturers.

Library thresholds can be modified following internal validation of the library, an assessment of external sample performance and a consideration of the next best match.

Once the threshold values have been set, the internal validation should be repeated to prove acceptable discrimination between different groups while maintaining acceptance of a material to its group. This may be an iterative process.

2.5 Library validation

The objective of validation of an analytical procedure is to ensure that it is suitable for its intended purpose. This purpose must be considered in determining the validation parameters required.

This section gives guidance on various aspects of validation, but each application must be considered individually. Any prospective work must be documented in a validation program. This should describe in detail the tests to be performed and the acceptance criteria to be applied.

2.5.1 Internal and external validation

Internal. For any spectral database construction, an evaluation of the performance of that library is performed. This is based on the samples selected to make up that library (i.e. demonstrate that library samples can be discriminated from each other). This is usually performed by the software. The exact procedure used can vary depending on the software, but typical steps are:

- Verification that the spectra used to create the library are identified correctly, using the chosen match method (e.g. correlation or distance);
- confirmation that the distributions for materials in the library do not overlap;
- the use of cross-validation in library construction.

External. After successful internal validation, the performance of the database is verified using authenticated samples that were not used to generate that database.

Specificity   Potential "challenges" should be presented to the database. These can be materials received on site that are similar to database members visually, chemically or by name. Consideration should be given to materials manufactured by external suppliers, but not necessarily received on site, that could be delivered in error. The extent of specificity testing is dependent on the application and the risks being controlled.

Samples of material represented by the library, but not used to create it (i.e. the independent set of samples specified in Section 2.3), must give positive identifications when analyzed.

In the case of a qualification method, it is important to specify why it is being used, and then apply an appropriate challenge, e.g. different grades of the same material, anhydrous/hydrated forms, different polymorphs.

Repeatability   This is not normally required for identification methods. For qualification methods, repeatability is addressed in order to demonstrate that the acceptance thresholds established provide reliable discrimination between acceptable and unacceptable materials; the approach is therefore conceptually different for NIR methods compared with conventional methods.

Robustness   The challenges performed under this category may vary depending on the application and sampling technique. Some of the challenges may be covered as part of the development of the library and sampling technique. This tests the effect of minor changes to normal operating conditions on the analysis. The use of experimental design may be considered to maximize the information available.

Typical challenges are:

- effect of environmental conditions (e.g. temperature, humidity) on the analysis;
- effect of sample temperature on the analysis;
- position of the sample on the optical window;
- probe depth and compression/packing of material;
- effect of different sampling presentation devices;
- influence of change in instrument parts (e.g. lamps);
- changes in pre-processing and calibration algorithm parameters (e.g. derivative gap/segment, distance threshold, etc.).

2.6   Routine use

All operations concerning the use of NIR should be clearly documented.

Typically, these include:

- development and maintenance of libraries (including additions to an existing library);
- instrument calibration and maintenance;
- ongoing instrument PQ;
- routine use, including actions on failures.

Access to the system should be controlled so that only the required functions are available. This may be through password control. For example, a NIR system manager or development analyst may require full access to the software, while a routine operator may need access only to perform a routine identification.

2.6.1   Out-of-specification results

In developing spectral identification libraries, the aim is to capture the majority of natural acceptable variations in a material.

Occasionally, all these acceptable variations will not be captured in the initial training set, and there may be instances where test materials will be analyzed that are outside the model represented by the library, resulting in a "NIR non-conformance" against the NIR model. In these circumstances it is essential that the material is authenticated using appropriate alternative tests, prior to its acceptance and incorporation in the library (see Section 2.7.4). A documented procedure describing this process must be available.

The flowchart in Figure 2 gives an overview of actions resulting from a NIR identity failure.

2.7   Library maintenance

2.7.1   Database

Good information technology practices should be applied to ensure that adequate controls are in place (see Section 1.5.3):

- The current database is backed up after each change so that the information system can be recreated in the event of hardware failure or database corruption.
- A copy of each previous version of the database is available in the event of a review of the database at any future point in time.

2.7.2   Material groupings

Under normal conditions it is not recommended that materials are removed from the library even if their use is discontinued, as their presence adds to the overall robustness of the library database.
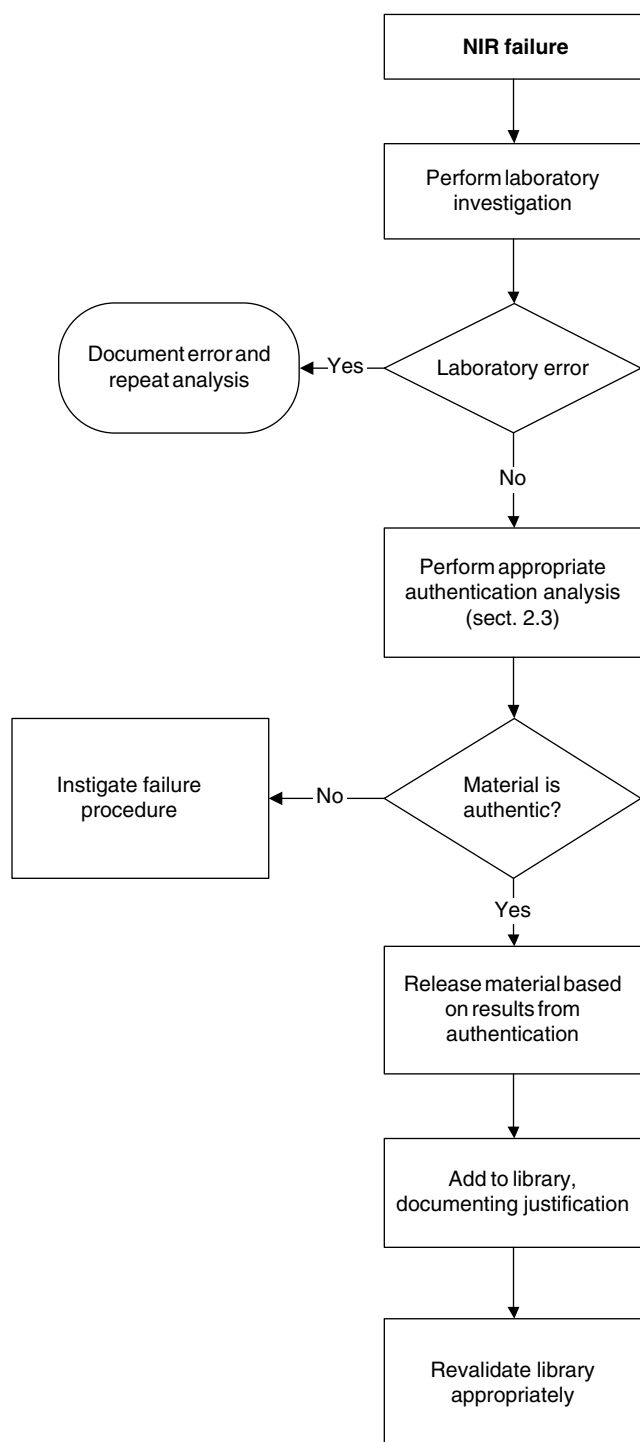
2.7.4 Material "library group" modification

On occasions it may be necessary to modify the sample set for a particular material to accommodate:

- changes in the physical properties of the material
- changes in the source of supply
- coverage of a wider range of characteristics.

In each case the authenticity of the new samples should be demonstrated using techniques other than NIR before their acceptance into the library. Where these samples are found to be acceptable, the library may be modified using the sample selection procedures detailed in Section 2.4.2. The library should then be revalidated to demonstrate continued specificity.

3 TECHNICAL GUIDELINES FOR QUANTITATIVE METHODS

3.1 Introduction to quantitative analysis

This section is complementary to the overview provided in Section 1, which presents a discussion of the characteristics that should be considered during the development and validation of NIR methods. It should also be read in conjunction with Section 2, since in NIR analysis the successful application of a qualitative method is a prior requisite in order to qualify the sample as eligible for analysis by the quantitative method. The purpose of this guideline is to provide some guidance and recommendations on how to consider the various design, development and validation characteristics for quantitative NIR methods, and to provide an indication of the data that should be presented in a registration application.

All relevant data collected during validation, and the procedures or formulas used for calculating validation characteristics, should be submitted and discussed as appropriate.

NIR methods are generally applied to detect and determine the analyte as it exists in the sample matrix (i.e. without any sample preparation), and quantitative NIR methods require calibration of the NIR spectral response against authentic reference data (e.g. obtained from gravimetric data for the input ingredients or from application of a valid reference analytical method). For this reason, the design, development and validation of a quantitative NIR method are inextricably linked and must be considered holistically. These features dictate that NIR methods be developed and validated in a conceptually different manner from conventional analytical methods, as outlined in Section 1, and the required approaches are detailed in the following sections. However, approaches other than those in this guideline may

Figure 2. Actions resulting from a NIR failure.

2.7.3 New materials addition

To add a new material to the library database, the material's sample set should be selected as detailed in the sample selection section. The library should be revalidated to demonstrate continued specificity.

be applicable and acceptable, and it is the responsibility of the applicant to choose the validation procedure and protocol most suitable for their product and process.

It is important to remember that the main objective of validation of any analytical procedure, including NIR, is to demonstrate that the procedure is suitable for its intended purpose.

In accordance with Section 1, and for the sake of clarity, this section considers the various design, development and validation characteristics in distinct sections, and the arrangement of these sections reflects the process by which a quantitative NIR analytical procedure may be designed, developed and validated.

In practice, it is still usually possible to design the experimental work on a quantitative NIR method such that the appropriate validation characteristics can be considered simultaneously to provide a sound, overall knowledge of the capabilities of the procedure, for instance specificity, linearity, range, accuracy, precision and robustness.

3.2   Feasibility study

A feasibility study should be performed to show that a quantitative analysis may be possible. Consideration should be given to analyte concentration and spectral response as an initial assessment of feasibility. Where possible, each of the individual components should be scanned separately to identify a band associated with the analyte of interest. It may be necessary to pre-process the spectra in order to assist the spectral interpretation. If individual components are not available, it may be possible to spike the analyte of interest into the matrix; this procedure may be used to investigate the linearity of the NIR response, though care should be taken to ensure adequate mixing of solids. The linearity of the NIR response may also be investigated directly by the algorithm's diagnostic tools.

The feasibility study will also give an indication of the calibration method to choose. In situations where the analyte shows a band free from interference from other constituents, then an MLR method could be attempted. However, where there is significant interference, more sophisticated methods such as PCR or PLS may be necessary. The choice of calibration method applied must be justified.

Sample handling and presentation should also be investigated. Transmission and reflection methods should be considered. It may be necessary to use fiber optics. Spectra of representative subsamples of the system should be repeatable. The contribution of sample presentation to the robustness of the measurement should be considered, e.g. repeated measurements of the presentation of the same sample should give spectra that overlay.

3.3   Sample collection

In view of the generally tight control of manufacturing processes, it may not be possible to generate manufacturing scale batches over the entire range of interest. Production samples may be augmented with batches made on development scale in order to achieve a wider compositional range. This is particularly relevant for NIR methods for intact solid dosage forms where extension of the compositional range is difficult to achieve by spiking samples. In such cases, due consideration should be given to including an appropriate balance of production and development samples in the sample sets, and it should be demonstrated that all samples are from the same statistical population.

Where compositional variation is achieved through specially manufactured batches, careful consideration should be given to the choice of compositions, and some form of structured compositional design (e.g. using a suitable experimental design software package) should preferably be undertaken. Variations in composition should preferably be established for all ingredients, not just the intended analyte(s), in order to cover matrix effects in the calibration. Intercorrelations between the analyte variations and other ingredient variations should be checked and minimized.

In all cases, the sample population should also encompass the expected variation in composition of matrix components other than the analyte of interest to determine whether the method is robust to these variations. Differences in physical properties (e.g. particle size) should also be included. Samples should be collected over a sufficient period of manufacturing time to allow for any expected process variations. If a method is being developed to analyze materials from different suppliers then these criteria should also be applied to include representative samples from each supplier.

The number of samples chosen should be sufficient to generate a calibration model of good predictive ability. The number of samples required to generate a calibration model will generally depend on the complexity of the matrix and analyte concentration.

Two sets of samples should be collected to create a calibration/calibration test set and an independent validation set.

3.4   Sample scanning

Spectra should be collected using a single, optimal method of sample presentation and instrument parameters for the system under investigation. A sufficient number of scans should be co-averaged to obtain suitable signal-to-noise levels for the quantitative application. The sample presentation should be chosen on the basis of the feasibility study.

Samples should be scanned at least in duplicate and selected from the sample set in random order. NIR spectra are affected by temperature and it may be necessary to record spectra of each sample over a range of temperatures. This will be required when the method will be implemented in poorly controlled environments. Consideration should be given to all experimental variables in the measurement possibly by use of experimental design to identify the significant factors. From these experimental design results, the contribution of each measurement variable should be investigated and either controlled or incorporated into the model.

The effective sample size in NIR methodology is generally significantly smaller than in conventional methods, and is often less than unit dose size. This is due not so much to the sample presentation accessories but to the area of the sample illuminated by the NIR beam. It must therefore be borne in mind that NIR is capable of detecting apparent heterogeneity, at least on a "micro" scale, and appropriate measures taken to accommodate this. For dose uniformity applications, this characteristic may provide a usable advantage, but in most applications some means of averaging the measurement over a larger sample area should be found. This may include transporting or spinning the sample through the NIR beam during spectral scanning.

3.5 Displaying and checking spectra

Each spectrum should be displayed over the entire wavelength range of interest and reviewed to ensure that they are suitable for use in the quantitative method. The spectra should be complete over the required wavelength range. Obvious outliers and poor duplicates can be identified visually. Spectra that feature excessive noise, unique and/or unexpected bands or "spikes" should be rejected and the sample re-scanned. If the repeat scan also shows unusual features it may be a true representation of the variation in the sample population. In this case further samples of similar composition or origin should be obtained and included in the population. Prior to the rejection of any spectra/spectrum, an investigation should be undertaken and a rejection rationale documented.

The spectra obtained during the development and validation phase of a study should be compared with the reference spectra recorded during the feasibility study to confirm that the spectra vary as expected with changing analyte concentration.

3.6 Reference data

In order to set up and validate a calibration of the NIR spectral response, quantitative reference data are needed for the analyte(s) involved and may be obtained gravimetrically, chromatographically or spectroscopically.

Wherever possible, reference data should be available for the same samples that were subjected to NIR scanning, where ideally the reference analysis is performed at the same time as the NIR scanning. Samples for the reference method should be representative of online or other process-orientated applications (e.g. taken at the same location and time).

Information on the performance of the reference method will be available from the existing validation data for that method, and should be gathered ready for use since it will influence and/or limit the performance achievable with the new NIR method which is calibrated against it. Important performance data are, for example, the linearity, accuracy and precision of the method. In this context, consideration should be given to the number of replicate determinations to be carried out and subsequently averaged to provide the reference data for the calibration and test set samples.

3.7 Sample selection – calibration and calibration test sets

The samples collected for use in the generation and optimization of the method (see Section 3.3) should be divided appropriately into calibration and calibration test sets.

- The calibration set will be used to calibrate the NIR spectral response against the reference data (i.e. to generate a regression model), and should be selected to cover the full variation in the sample set as a whole in order to ensure that the calibration (i.e. the method) covers the full range of interest for the analyte(s) and is robust to variations in excipients contents. In general in the subsequent application of the NIR method, it will be acceptable to interpolate within this range but not to extrapolate beyond it.
- The calibration test set is used as an initial test of the calibration model, and may also be used in optimizing the model as part of an interactive approach. The samples in this set should cover but not exceed the range of variations (e.g. compositions) in the calibration set, in order to present a meaningful challenge to the calibration.

In practice, this division into calibration and calibration test sets may be achieved in a number of ways:

- manual or software approaches may be used to select samples distributed across the full compositional range for the analyte of interest;
- software approaches may be used to select samples based on the degree of variation in their spectral data.

Whichever procedure is used, due consideration should be given to the need to provide an even distribution across the range. It may not always be possible to achieve even distribution, and uneven distributions (e.g. center-weighted) may be justifiable in some circumstances.

A statement should be included in the validation report to indicate what rationale was applied for sample selection.

Since pharmaceutical products are manufactured to tight tolerances, it may be necessary to augment production samples with samples from batches made on development scale in order to achieve a wider compositional range; this is particularly relevant for NIR methods for intact solid dosage forms where extension of the compositional range is difficult to achieve by spiking samples. In such cases, due consideration should be given to including an appropriate balance of production and development samples in both the calibration and calibration test sets, and wherever possible it should be demonstrated that all samples are from the same statistical population.

In certain circumstances, it may not be possible to extend the sample set to cover the desired calibration range; this might be the case if only production samples are available, or if a process-based or online method is being developed. In such situations the method will be valid only over the reduced range of the developed method, which may not extend to the specification limits for the analyte in the product. It may therefore be applicable for release of batches against tighter in-house limits but not for batches outside of this range but within the registered specification limits.

As well as considering even distribution across the range in both calibration and calibration test sets, the relative numbers of samples in each set should also be carefully apportioned. It is not possible to generalize on the number of samples required in a set since this depends on the nature of the product and the calibration model. Once calibration samples have been selected out of the whole sample set as described above, the remaining samples will constitute the calibration test set, and these should in turn be assessed for provision of satisfactory distribution across the range and adequate coverage at the extremes of the range in order to ensure a comprehensive challenge of the calibration. No calibration test set samples should fall outside the range covered by the calibration set. It should also be borne in mind that a calibration test set established in this way is constrained as it represents the residual samples after selecting out the calibration set, and therefore may not be viewed as entirely independent.

It is for this reason that a further independent set of samples – the validation set – is collected for use in formal validation of the method (see Section 3.3). This set may comprise only production batches, or include both production and development batches (to augment the range as described for calibration samples), but should cover variations up to but not exceeding the extremes of the calibrated range.

3.8   Data pre-processing

Pre-processing is a vital step in the chemometric analysis of NIR spectral data. Pre-processing can be defined as the mathematical manipulation of the NIR spectral data to enhance spectral features and/or remove or reduce unwanted sources of variation prior to the development of the calibration model. The pre-processing tools available can be broadly separated into two distinct categories depending upon whether they are applied to the sample spectra (individual wavelength) or whether they are applied to the entire data set. The selection of the optimum pre-processing tool will often necessitate iteration between the calibration model and the pre-processing step (Figure 3). It is, however, considered appropriate that the selection of the pre-processing tool is based upon an examination of the spectral data prior to any data modeling. Alternatively,
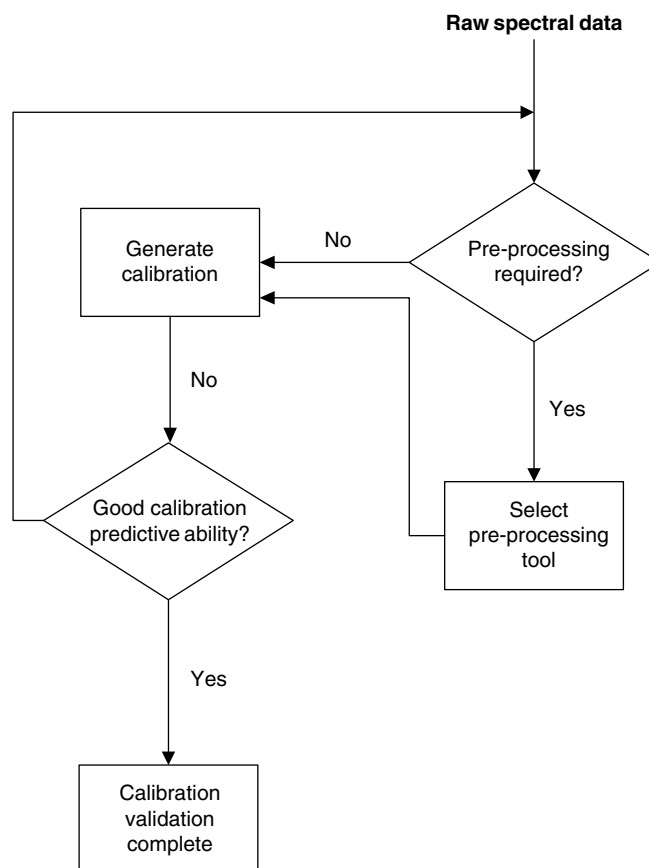

Figure 3. Data pre-processing workflow.

a number of pre-processing treatments can be evaluated in parallel and the optimal treatment chosen.

The first set of pre-processing tools are those that operate on the sample spectra. Numerous pre-processing tools exist, such as normalization, smoothing, baseline corrections and derivatives. Typical examples in NIR spectroscopy are derivatives and SNV. The selection of the best pre-processing tool should be made on known characteristics of the spectral data. An example might be the selection of derivatization to enhance spectral features in the data or the use of SNV to assist in minimizing baseline offsets. The second set of pre-processing tools are those that operate on the entire data set. Examples of this type of pre-processing are mean centering, variance scaling and autoscaling. An example of their use would be scaling experimental responses to unit variance.

When pre-processing data there are a number of important factors that should be taken into consideration prior to any analysis. Initial consideration should be given to any pre-processing steps performed by the instrument that are not controllable by the operator. An example of this might be an n-point smooth, and users are advised to clarify any inherent pre-processing with the software vendor. The selection of the pre-processing tool must be supported by a rationale for selection. This rationale should include a description of the effect of the pre-processing tool on the spectral data as well as its effect on the calibration model. In addition, consideration must be given as to how the selected software actually performs that pre-processing of the data. This is particularly important when selecting a derivative-pre-processing tool as different approaches to calculating derivative spectra exist.

3.9 Generation of calibration model

Calibration is the process of constructing a mathematical model to relate the response from an analytical instrument to the properties of samples. Prediction is then the process of using the developed model to predict properties of an unknown sample given the output from an analytical instrument. Both the construction and prediction stages are vital in the generation of calibration models for NIR.

In its broadest definition, there are two distinct approaches to the generation of calibration models: univariate and multivariate. Univariate calibration is traditionally most common, where a single response from an analytical instrument is related to the concentration of a single component. This is most probably not the case with NIR. Multivariate calibration is the process of relating multiple responses from an analytical instrument to the properties of a sample. The decision tree in Figure 4 is seen as a guide for the selection of

the appropriate multivariate calibration tool. This example considers only linear models and is not exhaustive as some of the models do tolerate some degree of nonlinearity.

A number of calibration techniques may be investigated in parallel to help select the optimum approach.

3.10 Validation of calibration model

Validation of the chosen calibration technique should proceed via the decision tree shown in Figure 5.

The sample set should be separated into calibration and validation sets. The calibration model may then be generated on the calibration sample set and an assessment made as to the quality of the calibration obtained. Calibration models can be generated using either internal (e.g. cross-validation) or external (calibration test set) approaches. The obtained calibration should then be tested against an independent validation set to obtain information on the predictive ability of the generated calibration model.

The accuracy and precision of the NIR method should be comparable to those of the reference method. Consideration should be given to root mean square error of calibration (RMSEC) and root mean square error of prediction (RMSEP), residuals and calibration variable factor selection. The regression coefficient (R^2) for the NIR method can be calculated but does not have the same relevance or importance as it does for traditional univariate methods, and reliance should not be placed upon it.

The validation of a multivariate calibration model is iterative in nature, but the rationale for each iteration should be clearly documented.

The application of a multivariate calibration should always include a classification stage. The intention of the classification stage is to ensure that the calibration model is applied only to samples that are consistent with the calibration model population (see Section 2).

3.11 Performance verification

NIR methods approved/adopted for use should be the subject of suitable performance verification. This may include accuracy monitoring, maintenance of the calibration model and appropriate change control.

3.11.1 Accuracy monitoring

Throughout routine application of a NIR method it is advisable to monitor the performance of the method. The first requisite for this is the routine operation of PQ checks (see Section 1.5.2) to demonstrate that the instrument is performing within specification. The continuing
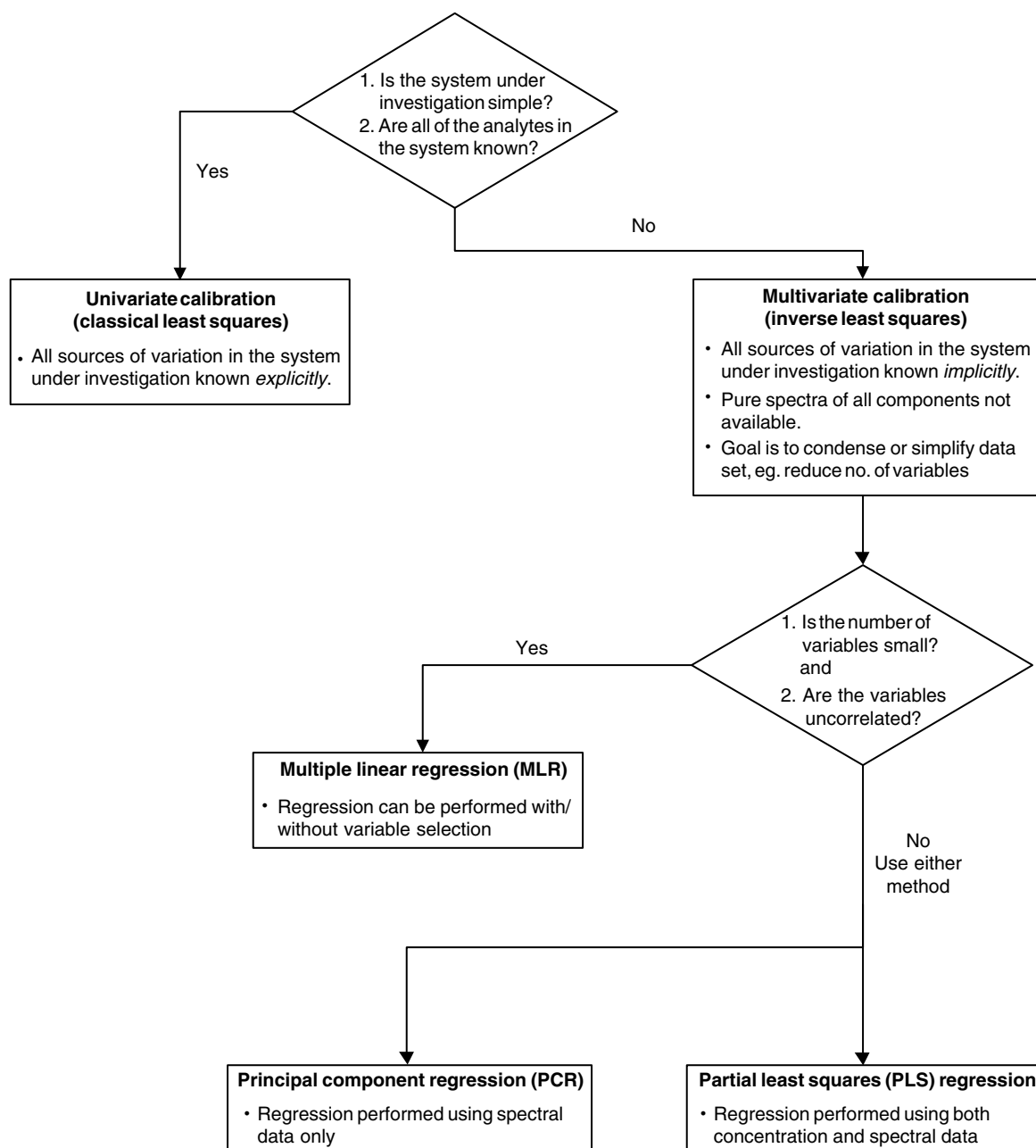
Figure 4. Multivariate calibration decision tree.

acceptable performance of the method in question may then be demonstrated by the use of a check sample and/or by comparison with data from the reference method. In both cases, acceptance criteria for agreement between the check data should be set prior to the monitoring.

If either of these procedures indicates unacceptable NIR results, corrective action will be necessary. This will involve initial investigations into the cause of the discrepancy – using established out-of-specification (OOS) investigational procedures – and may indicate that

the calibration model is not performing satisfactorily. Maintenance of the calibration model will then be required and may involve revalidation of the method.

Use of a check sample. A check sample of the product for which the method was developed may be used for this, but only if it is known that the sample is stable over time. The check sample is analyzed periodically before analysis of a new sample, and the result obtained is compared with previous results to demonstrate that the method remains accurate over time.
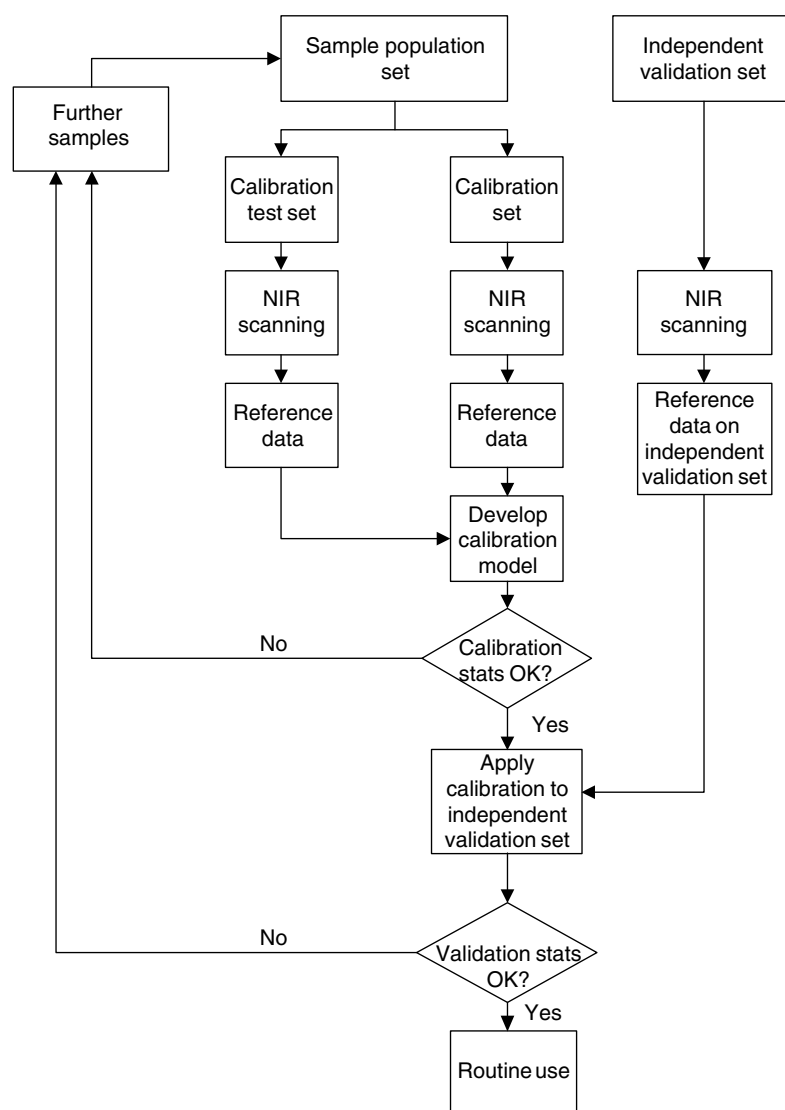
Figure 5. Calibration validation workflow.

Comparison with reference method. New samples analyzed by the NIR method may also be analyzed by the reference method and the results compared in order to demonstrate that the calibration model is still performing correctly.

A parallel testing stage can be considered once a fully validated method has been established, but is not a necessity. If adopted, it involves application of the validated method to a further set of samples, collected from ongoing production batches, and reference analytical data will need to be generated on these samples in the same way as described in Section 3.6. The NIR data for the parallel testing set should be assessed in the same way as validation data on the validation set (see Sections 3.9 and 3.10), and in addition statistical comparison of the NIR and the reference method may be carried out (e.g. by the application of

a Student *t*-test). Such statistical comparison dictates that parallel testing be continued for at least six batches.

However, it is more usual to carry out periodic checks ("skip lot testing") as a means of monitoring the method. Samples taken at defined intervals, e.g. either every *n* months or every *n*th batch, are analyzed by the reference method, as well as by the NIR method, and the results compared.

3.11.2 Maintenance of the calibration model

Maintenance of the model may be required as a result of accuracy monitoring (see Section 3.11.1) or for other reasons. Appropriate change controls should be established to cover any such maintenance. Changes that may lead to maintenance of the model include:
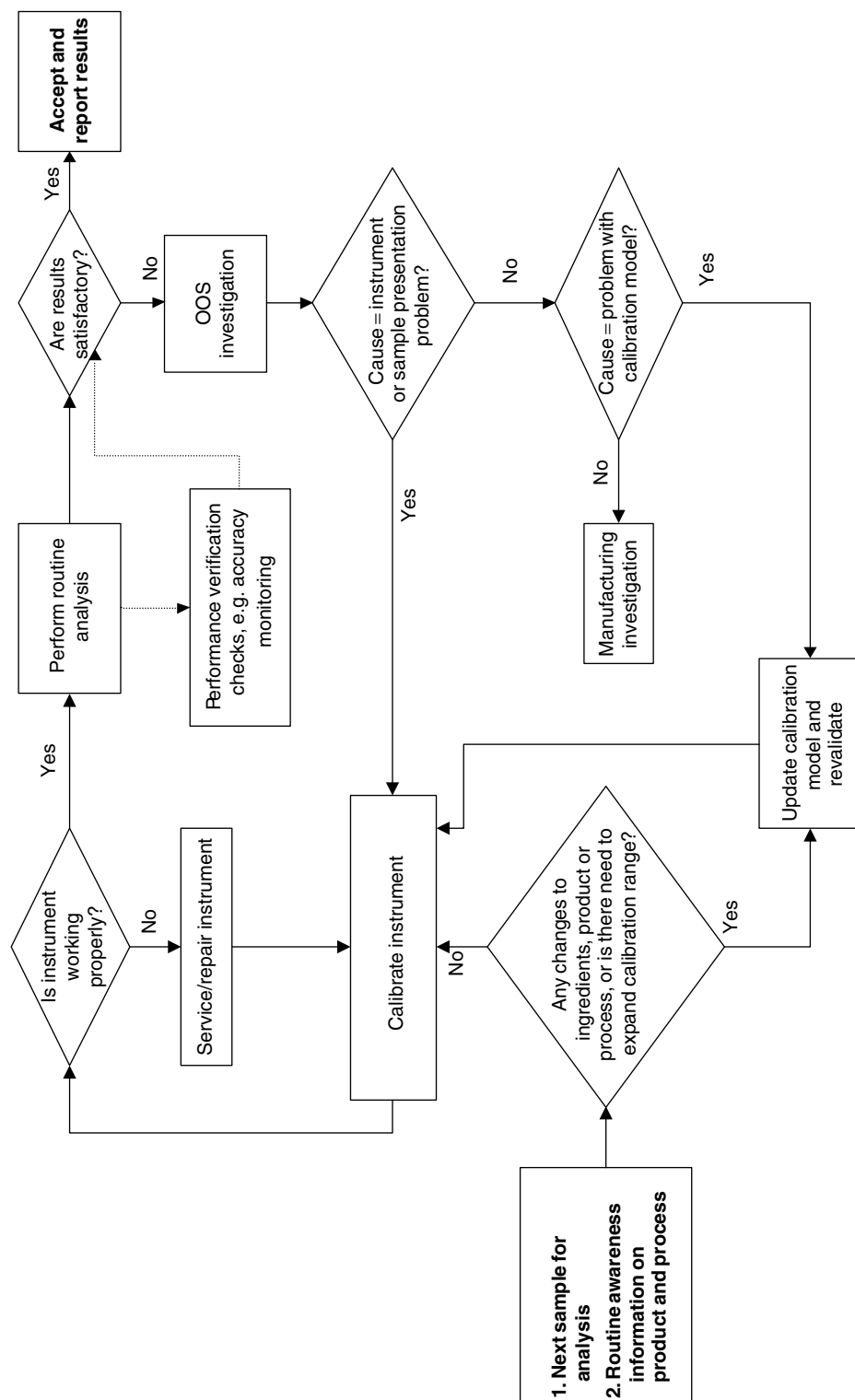
Figure 6. Maintenance of the calibration model.

- updating of the model to improve its applicability, e.g. extension of the valid range;
- different supplier or grade of an ingredient used in manufacture of the product to which the method applies;
- modifications to manufacturing process parameters (e.g. temperature, solvent, time, etc.);
- instrument modification or repair.

A decision tree to establish what actions are necessary to maintain or revalidate the model is provided in Figure 6.

3.12 Method transfer

The calibration model for a NIR method is developed, stored and applied in electronic form as part of an appropriate instrument/software package, and should not be transferred to another instrument unless procedures and criteria are applied to demonstrate that the model remains valid on the second instrument. Simpler means of transferring methods from one location to another include:

- Relocation of the original instrument complete with the calibration.
- Use of the complete calibration and validation samples sets to regenerate the method on a second instrument. In this case, the same method parameters are used as optimized on the originating instrument and the calibration is simply regenerated using spectra scanned on the second instrument.

In general, electronic calibration transfer is only recommended to another instrument of the same type and configuration. A number of calibration transfer procedures exist and should be applied as appropriate bearing in mind the availability of any necessary samples and/or standards. Procedures involve the use of various chemometric (mathematical and statistical) approaches with appropriate validation.

The need for method transfer can be planned into a NIR method at the development stage. This allows for an option whereby the calibration model is developed and validated for several instruments at the same time by including the instrument's variation in the calibration. By this means the calibration is valid on all instruments used in the development and the need for transfer between them is avoided. Furthermore, if a PLS calibration has been developed on a single instrument it can be supplemented to become a two-instrument calibration at a later time.

ABBREVIATIONS AND ACRONYMS

ABPI	Association of the British Pharmaceutical Industry
DQ	Design Qualification
ICH	The International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use
IQ	Installation Qualification
NIST SRM	National Institute of Standards and Technology Standard Reference Material
OOS	Out-of-specification
OQ	Operational Qualification
PASG	Pharmaceutical Analytical Sciences Group
PQ	Performance Qualification
RMS	Root Mean Square
RMSEC	Root Mean Square Error of Calibration
RMSEP	Root Mean Square Error of Prediction
SIMCA	Soft Independent Modeling by Class Analogy
SMV	Spectral Match Value
SNV	Standard Normal Variate

REFERENCES

1. K. Leiper, J. Vessman and C. van der Vlies, *PHARMEUROPA*, **10**(3), 468 (1998).

2. ICH Quality Guidelines Q2A, Text on Validation of Analytical Procedures; Q2B, Validation of Analytical Procedures: Methodology (http://www.ifpma.org/ich5.html) (1996).

3. 'Near-infrared Spectrometry, in Methods of Analysis 2.2.40' in "European Pharmacopoeia", 3rd edition, Council of Europe, Strasbourg, 43 (1997).

4. W. Plugge and C. Van Der Vlies, *J. Pharm. Biomed. Anal.*, **11**, 435 (1993).

5. A.C. Moffat, A.D. Trafford, R.D. Jee and P. Graham, *The Analyst*, **125**, 1341 (2000).

6. M. Freeman, M. Leng, D. Morrison and R.P. Munden, *Pharmaceut. Technol. Eur.*, **November**, 40 (1995).

7. W.L. Yoon, R.D. Jee and A.C. Moffat, *The Analyst*, **123**, 1029 (1998).