Method Validation in Pharmaceutical Analysis

Edited by
J. Ermer and J. H. McB. Miller

Method Validation in Pharmaceutical Analysis. A Guide to Best Practice. Joachim Ermer, John H. McB. Miller (Eds.)

ISBN: 3-527-31255-2

Related Titles from Wiley-VCH:

M. S. Lee

LC/MS Applications in Drug Development

2002
ISBN: 0-471-40520-5

M. Stoeppler, W. R. Wolf, P. J. Jenks (Eds.)

Reference Materials for Chemical Analysis

Certification, Availability, and Proper Usage

2001
ISBN: 3-527-30162-3

J. M. Miller, J. B. Crowther (Eds.)

Analytical Chemistry in a GMP Environment

A Practical Guide

2000
ISBN: 0-471-31431-5

Method Validation in Pharmaceutical Analysis

A Guide to Best Practice

Edited by
Joachim Ermer, John H. McB. Miller

WILEY-VCH Verlag GmbH & Co. KGaA

Edited by

Dr. Joachim Ermer
sanofi-aventis
Industriepark Höchst
Build. G875
65926 Frankfurt
Germany

Dr. John H. McB. Miller
European Directorate for the Quality
of Medicines (EDQM)
16, Rue Auguste Himly
67000 Strasbourg
France

Library of Congress Card No.
applied for

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library.

Bibliographic information published by Die Deutsche Bibliothek
Die Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data is available in the Internet at <http://dnb.ddb.de>.

Printed in the Federal Republic of Germany.
Printed on acid-free paper.

Typesetting Kühn & Weyh, Satz und Medien, Freiburg
Printing betz-druck GmbH, Darmstadt
Bookbinding Litges & Dopf Buchbinderei GmbH, Heppenheim

ISBN-13: 978-3-527-31255-9
ISBN-10: 3-527-31255-2

Preface

A number of articles and guidelines already exist dealing with the validation of analytical methods. However, the editors consider that none of the texts completely covers all aspects pertinent to analytical validation for, in particular, methods in pharmaceutical analysis. The editors have attempted, with the authors of the relevant chapters, to bring all these elements together in one book that will be useful to both analysts in the pharmaceutical industry (and beyond) as well as to assessors at the registration authorities for medicines.

Methods used in pharmaceutical analysis must be sufficiently accurate, specific, sensitive and precise to conform to the regulatory requirements as set out in the relevant guidelines of "The International Conference of Technical Requirements for the Registration of Pharmaceutical for Human Use " (ICH), which are applied by the licensing authorities and by some pharmacopoeias. The chapters in Part I deal specifically with the fundamentals of the different validation parameters, giving special emphasis to practical examples and recommendations. It is not intended to replace statistical textbooks but the editors have attempted to provide sufficient background information, illustrated by practical examples to aid the reader in understanding and choosing the relevant parameters and acceptance criteria to be considered for the application of any one analytical procedure to a particular purpose.

Contributions to Part II of this book deal with the life-cycle approach to validation starting with the qualification of equipment employed, the adaptation of ICH guidelines to the early stages of drug development, the relation between analytical variability and specification acceptance criteria, the continual assessment of the performance of the methods when in regular use, the transfer of analytical procedures, and out-of-specification results. There are also chapters dealing with the validation of pharmacopoeial methods and future perspectives for validation.

December 2004

John H. McB. Miller
Joachim Ermer

Method Validation in Pharmaceutical Analysis. A Guide to Best Practice. Joachim Ermer, John H. McB. Miller (Eds.)

ISBN: 3-527-31255-2

Contents

Method Validation in Pharmaceutical Analysis. A Guide to Best Practice. Joachim Ermer, John H. McB. Miller (Eds.)

ISBN: 3-527-31255-2

List of Contributors

Dr. Martin Bloch
Analytical Research and Development
Novartis
WSJ-360.1104
4002 Basel
Switzerland

Mark Broughton
Head of QC Analytics
Holmes Chapel
Aventis
London Road, Holmes Chapel Crewe,
Cheshire CW4 8BE
UK

Dr. Christopher Burgess
Burgess Consultancy
'Rose Rae', The Lendings, Startforth,
Barnard Castle, Co,
Durham DL12 9AB
United Kingdom

Mr. Ray Cox
Retired from:
Abbott Laboratories
Manager Corporate Compendia and
Reference Standards
1222 Pigeon Creek Rd
Greeneville, TN 37743
USA

Dr. Joachim Ermer
Director of Analytical Processes and
Technology
Global Analytical Development,
QO TSS Aventis
Industriepark Höchst
Build. G875
65926 Frankfurt am Main
Germany

Dr. Gerd Kleinschmidt
Head of Laboratory (New Projects and
Technologies) Global Pharmaceutical
Development Analytical Sciences,
GDPAnSc Aventis
Industriepark Höchst
Build. H790
65926 Frankfurt am Main
Germany

Dr. John H. McB. Miller
Head of the Division III (Laboratory)
European Directorate for the Quality of
Medicines (EDQM)
16, rue Auguste Himly
67000 Strasbourg
France

Dr. David Rudd
Glaxo Smithkline
Building 5
Park Road,
Ware Hertfordshire SG12 0DP
United Kingdom

Method Validation in Pharmaceutical Analysis. A Guide to Best Practice. Joachim Ermer, John H. McB. Miller (Eds.)

ISBN: 3-527-31255-2

Prof. Hermann Wätzig
Technical University Braunschweig
Institut für Pharmazeutische Chemie
Beethovenstr. 55
38106 Braunschweig
Germany

Part I:
Fundamentals of Validation in Pharmaceutical Analysis

Method Validation in Pharmaceutical Analysis. A Guide to Best Practice. Joachim Ermer, John H. McB. Miller (Eds.)

ISBN: 3-527-31255-2

1
Analytical Validation within the Pharmaceutical Environment

Joachim Ermer

Validation is, of course, a basic requirement to ensure quality and reliability of the results for all analytical applications [8]. However, in comparison with analytical chemistry, in pharmaceutical analysis, some special aspects and conditions exist that need to be taken into consideration. For example, the analytical procedures (apart from pharmacopoeial monographs) are often in-house developments and applications. Therefore, the degree of knowledge and expertise is initially much larger compared with standard methods. The same can be assumed for the samples analysed. The matrix (placebo) in pharmaceutical analysis is usually constant and well known and the ranges where the sample under analysis can be expected are usually well defined and not very large. Evaluation (of batches, stability investigations, etc.) is based on the results of various procedures or control tests, thus their performances can complement each other. Acceptance limits of the specification are fixed values, often based on tradition, as in the case of assay of an active ingredient, or they may be based on specific toxicological studies, which take large safety factors into account, as for impurities. Last, but not least, validation in pharmaceutical analysis has its own regulations. These few – by far from exhaustive – remarks should make it obvious that these special considerations will have an impact on the way validation in pharmaceutical analysis is performed.

The first part of this book focusses on the fundamentals of validation in pharmaceutical analysis, the 'environmental' framework as well as the *implications* for experimental design and suitable calculations. Of course, the basic principles of validation are the same for any analytical procedure, regardless of its field of application. However, the discussions and recommendations focus on pharmaceutical applications, so the reader needs to adjust these to suit his or her purpose, if different. Nevertheless – as validation should never be regarded as simply working through a checklist – this is also required in the case of pharmaceutical analysis, but perhaps to a lesser extent, compared with other areas of application.

Method Validation in Pharmaceutical Analysis. A Guide to Best Practice. Joachim Ermer, John H. McB. Miller (Eds.)

ISBN: 3-527-31255-2

1.1 Regulatory Requirements

"The object of validation of an analytical procedure is to demonstrate that it is suitable for its intended purpose" [1a], determined by means of well-documented experimental studies. Accuracy and reliability of the analytical results is crucial for ensuring quality, safety and efficacy of pharmaceuticals. For this reason, regulatory requirements have been published for many years [1–7].

The International Conference on the Harmonisation of Technical Requirements for the Registration of Pharmaceuticals for Human Use (ICH) was initiated in 1990, as a forum for a constructive dialogue between regulatory authorities and industry, in order to harmonise the submission requirements for new pharmaceuticals between Europe, the United States of America and Japan. One of the first topics within the Quality section was analytical validation and the ICH was very helpful in harmonising terms and definitions [1a] as well as determining the basic requirements [1b]. Of course, due to the nature of the harmonisation process, there are some compromises and inconsistencies. In Table 1-1, the required validation characteristics for the various types of analytical procedures are shown.

Table 1-1: Validation characteristics normally evaluated for the different types of test procedures [1a] and the minimum number of determinations required [1b]

		Analytical procedure			
Validation characteristic	**Minimum number**	**Identity**	**Impurities**		**Assay**[1]
			Quantitative	**Limit**	
1. Specificity [2]	Not applicable	Yes	Yes	Yes	Yes
2. Linearity	5	No	Yes	No	Yes
3. Range	Not applicable	No	Yes	No	Yes
4. Accuracy	9 (e.g. 3 × 3)	No	Yes	No	Yes
5. Precision					
Repeatability	6 or 9 (e.g. 3 × 3)	No	Yes	No	Yes
Intermediate precision/ Reproducibility [3]	(2 series)[4]	No	Yes	No	Yes
6. Detection limit	Approach dependent	No	No [5]	Yes	No
7. Quantitation limit		No	Yes	No	No

Yes / No normally evaluated / not evaluated

1 including dissolution, content/potency

2 lack of specificity of one analytical procedure could be compensated by other supporting analytical procedure(s)

3 reproducibility not needed for submission

4 no number given in [1b], logical conclusion

5 may be needed in some cases

Two guidelines on validation were issued by the US Food and Drug Administration (FDA), one for the applicant [2], the other for inspectors and reviewers [3]. The first one is also intended to ensure that the analytical procedure can be applied in an FDA laboratory and therefore requires a detailed description of the procedure, reference materials, as well as a discussion of the potential impurities, etc. The second guideline focuses on reversed-phase chromatography and provides a lot of details with regard to critical methodological issues, as well as some indication of acceptability of results. A revised draft of the first guideline was published in 2000 [4]. According to the title "*Analytical procedures and methods validation*", it also includes the content and format of the analytical procedures, the requirements for reference standards and various types of analytical technique. Therefore, this guidance is more comprehensive than the ICH Guidelines, but is rather too focussed on providing 'instrument output/raw data'. As this is an inspection and documentation issue, it should be separated from the validation. A very detailed discussion is provided in the Canadian guideline [7] with respect to requirements and particularly acceptance criteria. Although this allows some orientation, the given acceptance criteria were sometimes rather too ambiguous, for example, the intermediate precision / reproducibility of less than 1% for drug substances (see Section 2.1.3.2 and Fig. 2.1-12).

So why is it still important to discuss validation?

First of all, the ICH guidelines should be regarded as the basis and philosophical background to analytical validation, not as a checklist. *"It is the responsibility of the applicant to choose the validation procedure and protocol most suitable for their product"* [1b]. It will be shown in the next sections that suitability is strongly connected with the requirements and design of the given analytical procedure. As this obviously varies, at least with the type of procedure, it must be reflected in the analytical validation. This includes the identification of the performance parameters relevant for the given procedure, the definition of appropriate acceptance criteria and the appropriate design of the validation studies. In order to achieve this, the analyst must be aware of the fundamental meaning of these performance parameters, as well as the calculations and tests and their relationship to the specific application. The former is discussed in detail in Chapter 2, the latter in the following sections. A lack of knowledge or (perhaps) a wrong understanding of 'efficiency' will lead to validation results that address the real performance of the analytical procedure only partly or insufficiently. This is, at the very least a waste of work, because the results are meaningless. Unfortunately, this can also be found rather too frequently in publications, although to a varying extent for the different validation characteristics. Such common insufficiencies are discussed in the respective sections of Chapter 2.

1.2 Integrated and Continuous Validation

Validation should not be regarded as a singular activity [4], but should always be understood with respect to the life cycle of the analytical procedure. Starting with the method development or optimisation, the performance of the analytical proce-

dure should be matched to the requirements in an *iterative process*. Some validation characteristics, such as specificity (selective separation) or robustness, are more important in this stage (see Section 2.7). However, this depends on the type of procedure. In the case of a complex sample preparation, or cleaning methods (see Section 2.3.4), precision and accuracy may play an important role in the optimisation process. One should also be aware that the validation requested for submission, i. e. a demonstration of the *general* suitability of the respective analytical procedure – can only be considered as a basis. The user of any method has to guarantee that it will stay consistently in a validated status, also referred to as the life-cycle concept of analytical validation [9]. In this process, an increasing amount of information can be compiled.

This does not necessarily mean that additional work always needs to be done. During the actual application of the methods, a lot of data is generated, but often left unused ('data graveyard'). In order to make rational and efficient use of these data, they must be transformed to information (i.e., processed and condensed into performance parameters). When enough reliable information is compiled, it can be further processed to gain knowledge that eventually enables us to achieve a better understanding and control of the analytical procedure (see also Section 2.1.4 and Chapter 9). The whole process is well known as an *'information pyramid'* (Fig. 1-1). This knowledge can also be used to improve analytical procedures, for example, by changing from the traditional 'daily' calibration in an LC assay to a quantitation using 'predetermined' calibration parameters (comparable to a specific absorbance in spectrophotometry), with advantages both in efficiency and reduced analytical variability [10].

Transfers of analytical procedures to another site of the company or to a contract laboratory – quite common nowadays – often result in a challenging robustness test, especially if not appropriately addressed in the validation. Acceptance criteria for a successful transfer may be derived from the validation itself, or from the same principles as for calculations and tests in validation, because here the performance of the analytical procedure is also addressed (see Chapter 7). On the other hand, comparative studies will provide quite reliable performance data of the analytical procedure (see Section 2.1.3.2).

Besides this 'horizontal' integration, analytical validation also needs to be included in the whole system of *Analytical Quality Assurance* (AQA) [8], i.e., 'vertical' integration. This involves all (internal and external) measures which will ensure the quality and reliability of the analytical data, such as an equipment qualification program (see Chapter 4), appropriate system suitability tests (see Section 2.8), good documentation and review practices, operator training, control charts (see Chapter 9), etc.

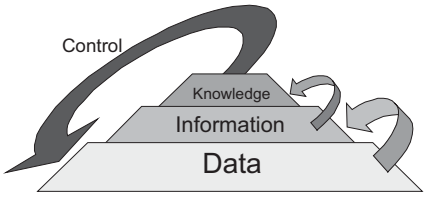


Figure 1-1 Information pyramid.

1.3 General Planning and Design of Validation Studies

Performance is strongly connected with the requirements and design of the given analytical procedure (see Section 1.4.1). As this obviously varies, it must be reflected in the planning and design of the analytical validation. Consequently, a checklist approach is not appropriate. In order to ensure thorough planning, i.e., to identify the relevant performance parameters, to define appropriate acceptance criteria and then to design the studies accordingly, *validation protocols* should be prepared. In addition to this 'good science' reason, protocols can also be regarded as a general GMP requirement and are common practice also the in case of process validation, cleaning validation, equipment qualification, transfer, etc.

The analyst may be faced with the problem of the iterative nature of the method development / validation process. However, here one may distinguish between performance parameters (and the corresponding validation characteristics) of the final analytical procedure and those obtained or derived from different method conditions, such as specificity and robustness. The former can be addressed (before starting the experimental studies, following usual practice) in the protocol, the latter can be referred to in the validation report and/or protocol (see Chapter 5).

Of course, the *extent and depth* of the validation studies, as well as acceptance criteria, should be defined in relation to the required performance ('importance') and the 'environment' of the respective analytical procedure, such as the stages of development (see Chapter 5), or the stages of manufacturing / synthesis. Important or critical procedures (within the context of validation) can be expected to have tighter specification limits. In these cases, such as the assay of active or of critical impurities, it is recommended to address the validation characteristics *separately* (for example, precision with authentic samples and accuracy with spiked samples), in order to increase the power of the results. In other cases, such as the determination of other ingredients or of impurities or water sufficiently below specification limits, several validation characteristics, for example, precision, linearity, and accuracy (quantitation) limit in dependence on the range, see Section 2.6.4) can be investigated simultaneously, using the same spiked samples.

The ICH Guidelines [1a,b] are mainly focused on chromatographic procedures, as can be seen in the methodology guideline [1b]. Therefore, they should be regarded more as a guide to the *philosophy of validation* – i.e., used to identify relevant performance parameters of the given analytical procedure – than as a 'holy grail'. If the special conditions or techniques are not covered in the ICH guideline, the validation approach must then be adapted accordingly (see Chapter 11). The FDA Guidance [4], and the Technical Guide of the European Pharmacopoeia (EP) [11], as well as Chapter 8 also provide details for specific analytical techniques.

1.3.1 Always Look on the 'Routine' Side of Validation

Curiously, one aspect often neglected during validation is its primary objective, i.e., to obtain the real *performance of the routine application* of the analytical procedure. As far as possible, all steps of the procedure should be performed as described in the control test. Of course, this cannot always achieved, but at least the analyst should always be aware of such differences, in order to evaluate the results properly.

What does this mean in practice?

For example, precision should preferably be investigated using *authentic* samples, because only in this case is the sample preparation identical to the routine application. It is also important to apply the intended calibration mode exactly as described in the analytical procedure. Sometimes the latter is not even mentioned in the literature. Precision is reported only from repeated injections of the same solution, ignoring the whole sample preparation. This is certainly not representative for the (routine) variability of the analytical procedure (see Section 2.1.2). Investigating pure solutions is usually of very limited practical use, for example, in the case of cleaning methods (see Section 2.3.4) or quantitation limit (see Section 2.6), or may even lead to wrong conclusions, as the following examples will show.

The minor (impurity) enantiomer of a chiral active ingredient was analysed by chiral LC using an immobilised enzyme column (Chiral-CBH 5 µm, 100 × 4 mm, ChromTech). The quantitation should be carried out by area normalisation (100%-method, 100%-standard), which would require a linear response function and a negligible intercept for both active and impurity enantiomer (see also Section 2.4.1). The experimental linearity investigation of dilutions of the active, revealed a clear deviation from a linear response function (Fig. 1-2). However, when the design was adjusted to simulate the conditions of the routine application, i.e., spiking the impurity enantiomer to the nominal concentration of the active, an acceptable linear relationship was found. Although a slight trend remained in the results, the recoveries between 99 and 105% can be regarded as acceptable for the intended purpose. A possible explanation for such behaviour might be that the interaction between the enantiomers and the binding centres of the immobilised enzyme (cellobiohydrolase, hydrolysing crystalline cellulose) is concentration dependent. Maintaining the nominal test concentration in the case of the spiked samples, the sum of both enantiomers is kept constant and consequently so are the conditions for interactions. In this case, the linearity of the active enantiomer cannot be investigated separately and the validity of the 100% method must be demonstrated by obtaining an acceptable recovery.

Stress samples

Another area where the primary focus of validation is often ignored is the use of stress test samples (see also Section 2.2). At least some of the applied conditions [1g] will result in degradation products without any relevance for the intended storage condition of the drug product. Therefore, such samples should be used with reasonable judgement for method development and validation. It is the primary objective of a suitable (impurity) procedure (and consequently its validation) to address degra-

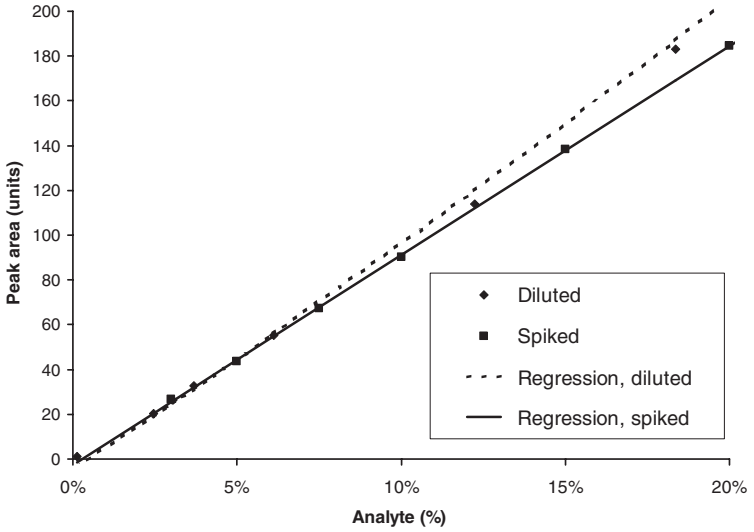


Figure 1-2 Linearity investigation of an enantiomeric LC determination. The diamonds and the squares represent dilutions of the active enantiomer and spikings of the impurity enantiomer to the active enantiomer, respectively. An obvious deviation from a linear function is observed in the case of the dilutions (broken line, polynomial to 3rd order), in contrast to the impurity in the presence of the active enantiomer (solid line, linear regression). The concentration on the x-axis is given with reference to the nominal test concentration of the active enantiomer.

dants "likely to be present" [1b], rather than a 'last resort'. However, it is also reasonable to allow for some 'buffer' [12].

Sometimes, applying artificial conditions cannot be avoided, in order to approach validation parameters, as in recovery investigations (see Section 2.3.2) or in dissolution, where no homogeneous samples are available. In the latter case, the assay part of the analytical procedure may be investigated separately. However, possible influences on the results due to the different application conditions need to be taken into account in the evaluation process as well as in the definition of acceptance criteria.

1.4
Evaluation and Acceptance Criteria

1.4.1
What does Suitability Mean?

The suitability of an analytical procedure is primarily determined by the requirements of the given test item, and secondly by its design (which is normally more flexible). Usually, the (minimum) requirements are defined by the acceptance limits of the specification (often termed traditionally as 'specification limits', but according to ICH [1e], the term 'specification' defines a *"list of tests, references to analytical proce-*

dures, and appropriate acceptance criteria"). For some applications, the requirements are explicitly defined in the ICH Guidelines. For example, the reporting level for unknown degradants in drug products is set to 0.1% and 0.05% for a maximum daily intake of less and more than 1 g active, respectively [1d] (Table 2.6-1). In the case of cleaning validation, the maximum acceptable amount of cross-contamination can be calculated based on the batch sizes and doses of the previous and subsequent product, the toxicological or pharmacological activity and/or the safety factors, and the so called specific residual cleaning limit (SRCL) [13]. Consequently, the corresponding test procedure must be able to quantify impurities or residual substance at this concentration with an appropriate level of precision and accuracy (see Section 2.3.4).

With respect to stability studies, the analytical variability must be appropriate to detect a (not acceptable) change in the tested property of the batch. This is illustrated in Figure 1-3 for determination of the content of active ingredient. The intrinsic degradation of 1.0% within 36 months can be reliably detected by an assay with a true variability of 0.5% (Fig. 1-3A), but not by one with 2.0% variability (Fig. 1-3B). Generally, acceptance limits of the specification (SL) have to enclose (at least) both the analytical and the manufacturing variability (see Chapter 6). Rearranging the equation describing this relationship (Eq. 6-12), the maximum permitted analytical variability can be calculated from the acceptance limits of the specification (Eq.1-1).

$$RSD_{\max}(\%) = \frac{|(BL-SL)| * \sqrt{n_{assay}}}{t(P,df)} \tag{1-1}$$

SL: Acceptance limits of the specification for active (% label claim).

BL: Basic limits, 100% – maximum variation of the manufacturing process (in %). In case of shelf-life limits, the lower basic limit will additionally include the maximum acceptable decrease in the content.

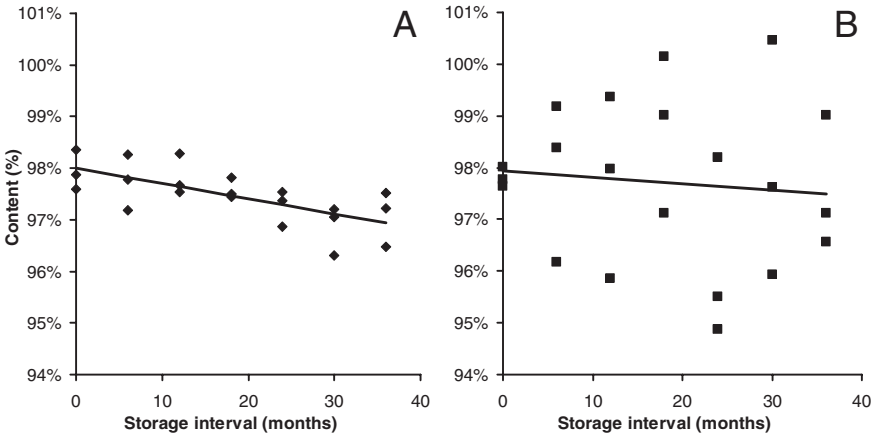


Figure 1-3 Illustration of the requirements for assay of an active ingredient during a stability study. The three individual results per storage interval were simulated based on a 1% decrease of content within 36 months and a normally distributed error of 0.5% (A) and 2.0% (B) using Eq. (2.1-3). The slope of the regression line in B is not significant.

n_{assay}: Number of repeated, independent determinations in routine analyses, insofar as the mean is the reportable result, i.e., is compared to the acceptance limits. If each individual determination is defined as the reportable result, n=1 has to be used.

$t(P,df)$: Student t-factor for the defined level of statistical confidence (usually 95%) and the degrees of freedom in the respective precision study.

The same basic considerations of the relationship between content limits and analytical variability [14] were applied to the system precision (injection repeatability) requirements of the EP [15] (see Section 2.8.3.8). The method capability index (see Section 10.5, Eq. 10-5) is based on similar considerations. However, here the normal distribution is used to describe the range required for the analytical variability (see Section 2.1.1). Consequently, the method capability index must be applied to single determinations (or to means if the standard deviation of means is used) and requires a very reliable standard deviation, whereas Eq.(1-1) can take a variable number of determinations directly into account, as well as the reliability of the experimental standard deviation (by means of the Student t-factor).

Of course, the precision acceptance limit thus obtained will be the minimum requirement. If a tighter control is needed, or if a lower variability is expected for the given type of method (analytical state of the art, see Section 2.1.3), the acceptance limits should be adjusted. A further adjustment may be required if there is a larger difference between repeatability and intermediate precision, i.e., if there is a larger inter-serial contribution (Eq. (2.1-10), Section 2.1.3.2). In such a case, an increased number of determinations in the assay will only reduce the repeatability variance, but not the variance between the series (s^2_g). Therefore, the term $\sqrt{n_{assay}}$ must be transferred to the left-hand side of Eq. (1-1) and RSD_{max}(%) rearranged to $\sqrt{s_g^2 + \frac{s_r^2}{n_{assay}}}$. This term corresponds to the standard deviation of the means from the routine assay determinations.

Many other performance parameters are linked with the analytical variability. Therefore, once an acceptable precision is defined, it can serve as an orientation for other acceptance criteria (for details, see Table 1-2 and Sections 2.1–2.6). As far as possible, *normalised (percentage) parameters* should be defined as validation acceptance limits, because they can be compared across methods and therefore more easily drawn from previous experience.

As can be seen from Eq. (1-1), the number of determinations also influences the acceptable performance, as well as the intended calibration mode (see Section 2.4). In principle, the analyst is rather flexible in his/her decision, provided that the minimum requirements are fulfilled. Often, the design of the calibration is more influenced by tradition or technical restrictions (for example the capabilities of the acquisition software) than by scientific reasons. Sometimes a 'check standard' is applied, i.e., the standard prepared and used for calibration is verified by a second standard preparation, the response of which needs to be within an acceptable range of the first one (e.g. ± 1.0%). This approach is not optimal. If the 'check standard' is only used for verification, 50% of the available data are ignored. Increasing the number of determi-

nations improves the reliability of the mean (see Fig. 2.1-4A). Therefore, it would be preferable to calculate the mean from all standard preparations (after verification of their agreement), in order to reduce the variability of the standard that will be included in the result for the sample (see discussion on repeatability and intermediate precision, Section 2.1.2). Of course, if the overall variability utilising only the first standard preparation is still acceptable, the procedure will be suitable. However, the analyst must be aware of the inter-relations and their consequences in order to make an appropriate decision and evaluation. This example also highlights the importance of applying the intended calibration, exactly as described in the control test for the intermediate precision study, otherwise the obtained result will not reflect the performance of the routine analytical procedure.

1.4.2 Statistical Tests

Significance Tests

Statistical significance tests should very cautiously be (directly) applied as acceptance criteria, because they can only test for a *statistical* significance (and with respect to the actual variability). On one hand, due to the small number of data normally used in pharmaceutical analysis, large confidence intervals (see Section 2.1.1) may obscure unacceptable differences (Fig. 1-4, scenario 3, S). On the other hand, because of sometimes abnormally small variabilities in (one of) the analytical series (that, however, pose no risk for routine application), differences are identified as significant which are of no *practical relevance* (Fig. 1-4, scenario 1, S) [16]. The analyst must decide whether or not detected statistical differences are of practical relevance. In addition, when comparing independent methods for the proof of accuracy, different specificities can be expected which add a systematic bias, thus increasing the risk of the aforementioned danger. Therefore, a statistical significance test should always be applied (as acceptance criteria) in a two-tiered manner, including a measure for practical relevance. For example, in the case of comparison of results with a target value, in addition to the nominal value *t*-test (see Section 2.3.1, Eq. 2.3-2), an upper limit for the precision and a maximum acceptable difference between the mean and the target value should be defined, in order to avoid the scenario 3 illustrated in Figure 1-4 (S).

Equivalence Tests

Such measures of practical relevance are an intrinsic part of the so-called *equivalence tests* [16, 28] (see also Section 7.3.1.3). In contrast to the significance tests, where the confidence intervals of the respective parameter(s) must include the target value (Fig. 1-4, scenario 2 and 3, S), equivalence tests, must be within an *acceptable range*. This measure of practical relevance is defined by the analyst. It is obvious in Figure 1-4, that such equivalence tests are robust with respect to small (scenario 1, E), but sensitive to large (scenario 3, E) variabilities.

Absoute Acceptance Limit

Another alternative is to use *absolute acceptance limits*, derived from experience (see Section 2.1.3) or from statistical considerations, as described in Section 1.4.1 for pre-

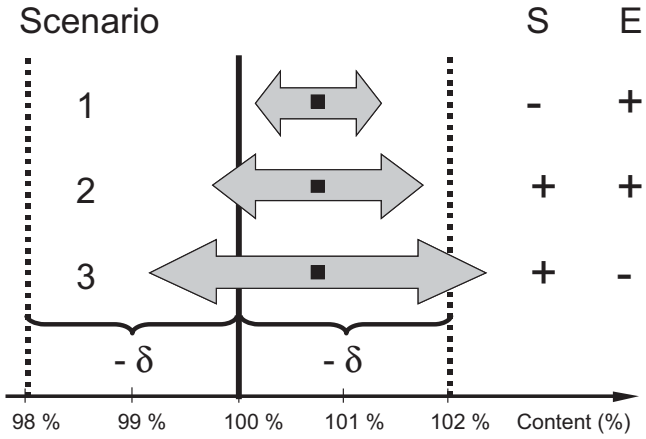


Figure 1-4 Illustration of statistical significance (S) and equivalence (E) tests for the example of a comparison between a mean and a target value of 100% (e.g., a reference or theoretical recovery). The acceptable deviation δ from the target (for the equivalence test) is symbolised by vertical dotted lines, the means, with confidence intervals indicated by double arrows. The outcome of the statistical tests for the three scenarios is indicated by '+' and '–' for 'pass' and 'fail' of the respective (H0) hypothesis, these are 'no statistical significant difference' and 'acceptable difference' for significance and equivalence test, respectively.

cision, and for a maximum acceptable difference in accuracy (see Section 2.3.5). In contrast to the equivalence tests, the actual variability of the data is neglected for the purpose of comparison (if means are used). However, usually the variability will be investigated separately.

If validation software is used, it must be flexible enough to meet these precautions [28].

Of course, statistical significance tests also have their merits, if properly applied. Even if a small variability does not pose a practical risk, when the suitability of a procedure is investigated, it may be assumed that such data are not representative for the usual (routine) application of the analytical procedure. This is an important consideration when the true parameter (standard deviation, mean) is the investigational objective, for example, the true precision of an analytical procedure, or if a reference standard is characterised. In collaborative trials, significance tests such as outlier tests are often defined as intermediary acceptance criteria for checking the quality of the data [17–19]. Deviating (i.e., unrepresentative) results (laboratories) are removed before proceeding to the next step, in which results are combined.

1.5 Key Points

- Validation should address the performance of the analytical procedure under conditions of routine use.
- Suitability is strongly connected with both the requirements and the design of the individual analytical procedure.
- Consequently, the analyst has to identify relevant parameters which reflect the routine performance of the given analytical procedure, to design the experimental studies accordingly and to define acceptance criteria for the results generated.
- Absolute, preferably normalised parameters should be selected as acceptance criteria. These can be defined from (regulatory) requirements, statistical considerations, or experience. Statistical significance tests should be applied with caution, they do not take into consideration the practical relevance.
- Validation must not be regarded as a singular event. The analyst is responsible for the continued maintenance of the validated status of an analytical procedure.

Acknowledgements

Some of the examples presented in my chapters, as well as the experience gained, are based on the work of, and discussion with, many colleagues in Aventis. Their important input is greatfully acknowledged, but I will abstain from an attempt to list them, both because of space as well as the danger of forgetting some of them. I would like to acknowledge in particular, John Landy, Heiko Meier, and Eva Piepenbrock.

Table 1-2 Examples of performance parameters in analytical validation. The acceptance criteria given are for *orientation purposes only*. They refer mainly to LC/GC procedures and have to be adjusted according to the requirements and the type of the individual test procedure. For details, see the respective sections in Chapter 2.

Validation characteristics parameter / calculations[1]	Type of analytical procedure / application	Acceptance criteria (orientational!)	Conditions / comment
Precision			
System precision (injection repeatability)	Assay	< 1%	Mainly reflection of the instrument (injection) variability, if sufficiently above QL
	Assay (DS)	According to EP [15]	Dependence on *n* and upper specification limit
	Assay (DP)	< 2% (USP)	Usually not sufficiently discriminative
	Impurities	< 2–5%	The smaller the concentration, the greater the influence of the detection/integration error
Repeatability			Preferably, authentic samples should be used
	Assay	Calculation from specification limits (Eq. 1-1)	Minimum requirement to achieve compatibility with specification limits
		< ≈ 1 – 2% (< 2 * TSD)	Dependent on type of DP (sample / preparation)
	Impurities	At QL: calculation from specification limits (Eq. 1-1, BL=QL)	Minimum requirement
		< ≈ 10 – 20%	Dependent on concentration level, preferably linked to QL
Intermediate precision / reproducibility			Analysis of variances (ANOVA)
Overall repeatability	Assay	< ≈ 1.5* TSD	More reliable due to increased number of determinations
Int. prec. / reproducibility		< ≈ 3–4* TSD	Dependent on type of DP (sample / preparation)
Specificity (quantitatively)			
Comparison with an independent procedure		see Accuracy	
Resolution factor	Chromatographic separations	> ≈ 2 (large difference in size)	For baseline-separated peaks, dependent on size difference, tailing and elution order
		> ≈ 1 (similar size)	
Peak-to-valley ratio	Chromatographic separations	> ≈ 0.25	For partly separated peaks

Table 1-2 Continued.

Validation characteristics parameter / calculations[1]	Type of analytical procedure / application	Acceptance criteria (orientational!)	Conditions / comment
Accuracy			
Comparison with an independent procedure or with a reference			Likely different specificities have to be taken into account
Difference between the means / to the reference	Assay (DS, n=6)	< ≈ 1 – 2%	Acceptable precision (of the most variable procedure) in the given concentration range
	Assay (DP, n=6)	< ≈ 2%	
	Impurities (n=6)	< ≈ 10 – 20%	
t-test		No significant difference between the means (95% level of significance)	Statistical significance test (see 1.4.2), only if specificities of the two procedures are the same or can be corrected.
Equivalence test	Assay (DS, n=6)	± ≈ 2%	Two-fold acceptable precision in the given concentration range; in contrast to simple comparison, the variability needs to be included (see 1.4.2)
	Assay (DP, n=6)	± ≈ 3%	
	Impurities (n=6)	± ≈ 10 – 20%	
Recovery			Spiking of known amounts of analyte into the respective matrix
Percent recovery			Concentration range < factor 10
Range of recovery mean	Assay (DP, n=9)	≈ 98–102%	Acceptable precision in the given concentration range
	Impurities (n=9)	≈ 80/90 – 110/120%	
Statistical evaluation		95% confidence interval of the mean includes 100%	Statistical significance test (see 1.4.2)
		95% confidence interval within 96 – 104%	Statistical equivalence test, definition of a practical acceptable deviation (see 1.4.2)
Relative standard deviation	Assay (DP, n=9)	< ≈ 2%	Weighing effects: small concentrations have a larger influence on the result
	Impurities (n=9)	< ≈ 10 – 20%	
Individual recoveries		No systematic trend	Graphical presentation strongly recommended.
Range of individual recoveries	Assay (DP, n=9)	≈ 97 –103%	Corresponds to ≈ 6*TSD
	Impurities (n=9)	≈ 70 – 130%	Dependent on smallest concentration
Recovery function (unweighted linear regression)			
Amount added vs. amount found			

Table 1-2 Continued.

Validation characteristics parameter / calculations[1]	Type of analytical procedure / application	Acceptance criteria (orientational!)	Conditions / comment
Slope	Assay (DP, n=9)	≈ 0.98 – 1.02	Larger weight of higher concentrations
	Impurities (n=9)	≈ 0.9 – 1.1	
Confidence interval of the slope		95% CI includes 1	Statistical significance test (see 1.4.2)
	Assay (DP, *n*=9)	95% CI within 0.96 – 1.04	Statistical equivalence test, definition of a practical acceptable deviation (see 1.4.2)
	Impurities (*n*=9)	95% CI within 0.90 – 1.1	
Further parameter see Linearity			
Linearity			Verification of the intended calibration model
Unweighted linear regression	Single-point calibration		Concentration range < factor 10 (constant variability over the whole range required)
	Multiple point calibration		
Residual plot		Random scatter, no systematic trend	
	Assay	± ≈ 2% around zero	Corresponds to ± 3*TSD, at higher concentrations
	Impurities	± ≈ 10 – 20% around zero	
Sensitivity plot		No systematic trend	If intercept negligible Corresponds to ≈ ± 3*TSD, at lower concentrations (larger weight)
	Assay	± ≈ 3% around the mean	
	Impurities	± ≈ 10 – 20% around the mean	
Numerical parameters are only meaningful after verification/demonstration of a linear function			
Residual standard deviation	Assay (DS)	< ≈ 1 – 1.5%	Acceptable precision in the given concentration range
	Assay (DP, spiking)	< ≈ 2 – 3%	
	Impurities	< ≈ 10 – 20%	
Coefficient of correlation		No suitable for quantitative measure of linearity! Relation to the experimental variability depends on number of values and concentration range.	
Statistical linearity tests		Only recommended in case of indication or assumption of non-linearity (statistical significance vs. practical relevance).	

Table 1-2 Continued.

Validation characteristics parameter / calculations[1]	Type of analytical procedure / application	Acceptance criteria (orientational!)	Conditions / comment
Mandel test		No significant better fit by quadratic regression	Statistical significance test (see 1.4.2)
Significance of the quadratic coefficient		95% CI includes zero	Statistical significance test (see 1.4.2)
ANOVA lack of fit		Measurement variability (pure error) > than deviation from linear regression line	Statistical significance test (see 1.4.2), requires replicate determinations for each concentration level
Absence of a constant systematic error			Required for single point calibration (external standard) and 100% method
Intercept as % signal at working or target concentration	Assay Impurities	< ≈ 1 – 1.5% < ≈ 10 – 20%	Acceptable precision in the given concentration range, avoid large extrapolation
Statistical evaluation of the intercept		95% CI includes zero	Statistical significance test (see 1.4.2)
		95% CI within –2% and + 2%	Statistical equivalence test, definition of a practical acceptable deviation (see 1.4.2)
Deviation between single-point and multiple-point calibration line within the working range		Maximum deviation < ≈ 1%	Error from the calibration model should be less than TSD
Weighted linear regression	Multiple point calibration	If quantitation is required over a larger concentration range (> factor 10–20), when variances are not constant. In case of constant matrix and negligible intercept, a single-point calibration is also appropriate.	
Residual plot (absolute)		Random scatter around zero	Deviations are concentration dependent (wedge-shaped)
Residual plot (relative)		Random scatter around zero, no systematic trend	Deviations dependent on the precision of the respective concentration
Non-linear regression	Non-linear calibration		Non-linear response function
Residual plot	Assay Impurities	Random scatter, no systematic trend ± ≈ 1–2% around zero ± ≈ 10 – 20% around zero	Corresponds to ≈ ± 3*TSD, at higher concentrations

Table 1-2 Continued.

Validation characteristics parameter / calculations[1]	Type of analytical procedure / application	Acceptance criteria (orientational!)	Conditions / comment
Detection and Quantitation limit		If required, DL corresponds to QL/3	Be aware of the high variability of the actual QL. Usually a general QL is required, valid for any future application.
Establishment from reporting thresholds	Unknown impurities (DS)	0.03% or 0.05%	> 1g or < 1g daily dose
	Unknown degradants (DP)	0.05% or 0.1%	> 2g or < 2g daily dose
Establishment from specification limits	Impurities, cleaning methods	50% SL	
Calculation from specification limits	Impurities, cleaning methods	According to Eq. (2.6-3)	Minimum requirement
Acceptable precision	Impurities, cleaning methods	RSD < ≈ 10 – 20%	
'Intermediate QL'	Impurities, cleaning methods	QL_{max} or Eq. (2.6-2)	Repeated determinations of QL

Abbreviations:
TSD = target standard deviation, average repeatability of a sufficient number of determinations/series, estimation for the true repeatability of the given analytical procedure or type of drug product (sample complexity/preparation)
DS = drug substance
DP = drug product
CI = confidence interval
SL = acceptance limit of the specification
D/QL = detection / quantitation limit
1: Only parameters and calculations are listed for which general acceptance criteria can be given. As some of the parameters are alternative possibilities, the analyst has to choose the parameters/tests most suitable for his/her purposes.

2
Performance Parameters, Calculations and Tests

The following sections discuss parameters and calculations, which describe the performance of analytical procedures according to the ICH validation characteristics. The selection and discussion of these parameters and calculations reflect the experience of the authors and is primarily based on practical considerations. Their relevance will vary with the individual analytical application; some are also suitable for addressing questions other than validation itself. It is not intended to replace statistical textbooks, but the authors have tried to provide sufficient background information – always with the practical analytical application in mind – in order to make it easier for the reader to decide which parameters and tests are relevant and useful in his/her specific case. Precision is discussed first, because many other performance parameters are linked to analytical variability.

2.1
Precision

Joachim Ermer

ICH

"The precision of an analytical procedure expresses the closeness of agreement (degree of scatter) between a series of measurements obtained from multiple sampling of the same homogeneous sample under the prescribed conditions. Precision may be considered at three levels; repeatability, intermediate precision and reproducibility." [1a]

Precision should be obtained preferably using authentic samples. As parameters, the standard deviation, the relative standard deviation (coefficient of variation) and the confidence interval should be calculated for each level of precision.

Repeatability expresses the analytical variability under the same operating conditions over a short interval of time (within-assay, intra-assay). At least nine determinations covering the specified range or six determinations at 100% test concentration should be performed.

Method Validation in Pharmaceutical Analysis. A Guide to Best Practice. Joachim Ermer, John H. McB. Miller (Eds.)

ISBN: 3-527-31255-2

Intermediate precision includes the influence of additional random effects within laboratories, according to the intended use of the procedure, for example, different days, analysts or equipment, etc.

Reproducibility, i.e., the precision between laboratories (collaborative or interlaboratory studies), is not required for submission, but can be taken into account for standardisation of analytical procedures.

Before discussing the precision levels in detail, some fundamentals concerning the distribution of data are recalled. This is deemed to be very important for a correct understanding and evaluation of the following sections. For practical applications, a good understanding of the acceptable and achievable precision ranges is crucial. The section concludes with the description of some approaches used to obtain precision results.

2.1.1 Parameters Describing the Distribution of Analytical Data

2.1.1.1 Normal Distribution

'*Measurements are inherently variable*' [16], i.e., the analytical data obtained scatter around the true value. The distribution of data can be visualised by histograms, i.e., plotting the frequency of the data within constant intervals (classes) throughout the whole data range observed. Such histograms can be generated using Microsoft Excel® (Tools/Data Analysis/Histogram; the Analysis ToolPak can be installed by means of Tools/Add-Ins). Usually, the number of classes corresponds approximately to the square root of the number of data. Figure 2.1-1 shows clearly that a large number of data is required to obtain a clear picture. The data were obtained by recording the absorbance of a drug substance test solution at 291 nm for 60 minutes with a sampling rate of 10/s. Of course, over such a long time, an instrumental drift cannot be avoided. From 15 minutes onwards, the drift in the absorbance values was constant. Various time segments were further investigated and for the drift between 35 and 60 minutes the lowest residual standard deviation of the regression line was observed. The data were corrected accordingly, i.e., the corrected data represent the scattering of the absorbance values around the regression line of the drift. The mean and standard deviation of these 15 000 data were calculated to be 692 and 0.1774 mAU, respectively. The very small relative standard deviation of 0.026 % represents only the detection variability of the spectrophotometer.

The usually assumed normal distribution, in physico-chemical analysis, could be confirmed for the data sets in the example, but even with 15 000 data the theoretical distribution cannot be achieved (Fig. 2.1-2). The normal distribution or Gaussian curve is bell-shaped and symmetrically centred around the mean (true value) for which the highest frequency is expected. The probability of measured data decreases with the distance from the true value and can be calculated with the probability density function (Eq. 2.1-1).

Probability density function:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \quad \text{Excel®: f(x)=NORMDIST(x;}\mu\text{;}\sigma\text{;FALSE)} \qquad (2.1\text{-}1)$$

μ and σ denote the true (population) mean and standard deviation, and replacing in Excel® 'FALSE' by 'TRUE' will give the cumulative function.

An analytical measurement can be regarded as a *random sampling* of data from the corresponding (normal) distribution of all possible data. This is illustrated in Figure 2.1-2, where randomly selected subsets of six subsequent data from the 15 000 absorbance values are presented.

But how do I know that my analysis results are normally distributed?

Although there are some statistical tools to test for normal distribution [16, or statistical textbooks], they are not very suitable from a practical point of view where

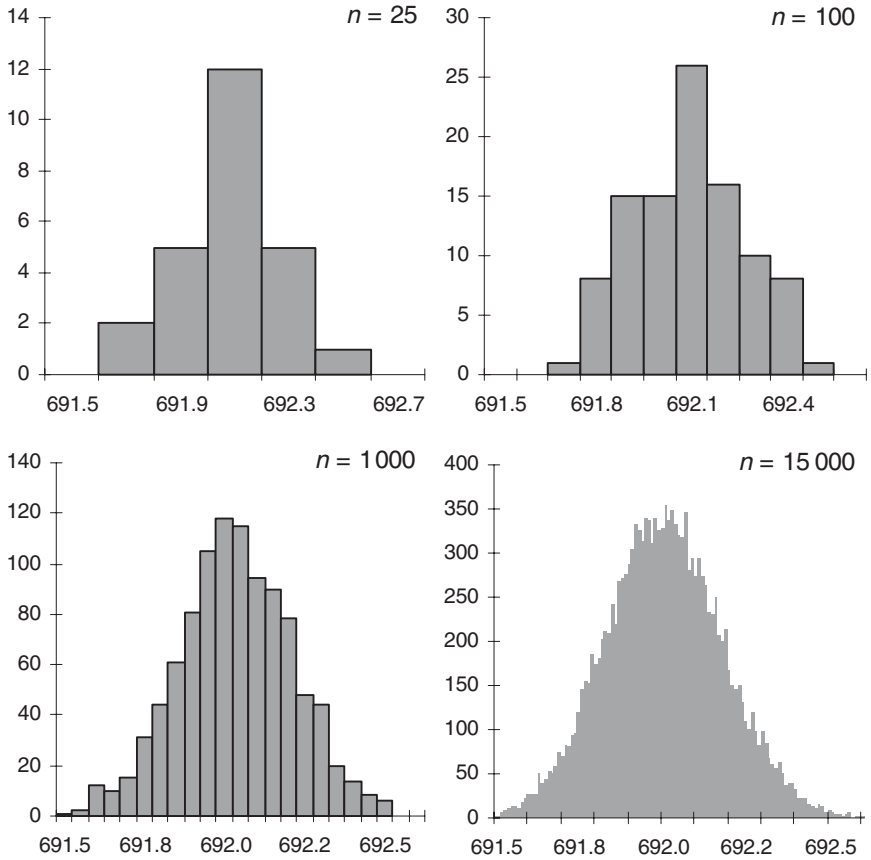


Figure 2.1-1: Histograms of 25, 100, 1000 and 15 000 data (for details see text). The y-axes display the frequency of the data within the absorbance intervals (classes) indicated on the *x*-axes. Apart from *n*=15 000 (where the number of classes is too high), each bar representing a data class is shown. The normal distribution of all four data sets was confirmed by χ^2-tests.

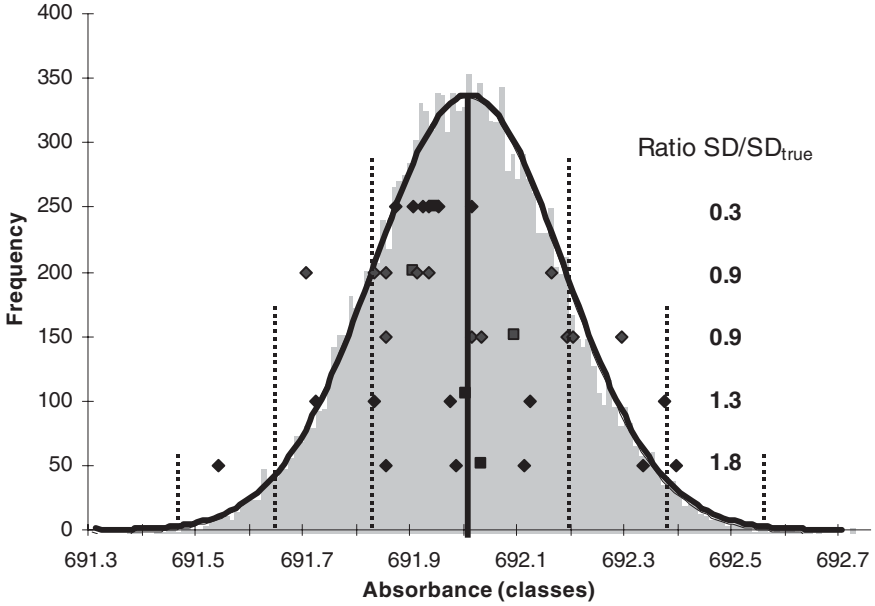


Figure 2.1-2: Histogram of 15 000 data with the theoretical Gauss curve. The intervals of 1, 2, and 3 (overall, true) standard deviations around the overall mean are indicated by dotted lines. The horizontally arranged diamonds represent random series of six subsequent data each, their means are given as squares. For each series, the standard deviation is calculated, as a ratio to the overall (true) SD.

there is only a small number of data. However, normal distribution can be assumed for the results of most physico-chemical analysis. Even if there is a minor deviation, regarding the large uncertainty of an experimental standard deviation (see Section 2.1.1.2), it will still provide a practical measure of the analytical variability. It is more important to verify the absence of systematic errors, for example, from degrading solutions, or insufficiently equilibrated systems, etc. This can be done in a very straightforward way by *visual inspection* of the data for trends. If the scale is not too large, the human eye is very sensitive in detecting trends and groupings in the experimental data. Therefore, experimental results should always be presented graphically.

Outliers

In the same way as non-random behaviour, a single datum which substantially deviates from the remaining data set, a so-called outlier, can influence both the mean and the standard deviation strongly. There are several statistical *outlier tests* available, but they suffer from the usual shortcomings of statistical significance tests (see Section 1.4.2). Most important, they cannot reveal the cause of the outlying result. The same practice as in pharmaceutical released testing, i.e., analytical results can only be invalidated if an analytical error can be assigned, should also be applied to validation studies (see Chapter 10). However, these tests may be applied as a diagnostic

tool in an investigation (although often the visual 'eyeball' test will reveal the same information). They may indicate that the data series is not representative. In such cases, the whole series should be repeated.

2.1.1.2 **Standard Deviations**

The standard deviation is an important parameter used to describe the width of the normal distribution, i.e., the degree of dispersion of the data. It corresponds to the horizontal distance between the apex and the inflection point of the Gaussian curve (Fig. 2.1-2, first pair of vertical dotted lines nearest to the mean solid line). The interval of ± 1 standard deviations around the true value includes just about two- thirds of all data belonging to this distribution. The two and three standard deviation intervals cover 95 % and 99.7 % of all data, respectively. The conclusion that almost all individual data of a population range within an interval of three standard deviations around both sides of the mean is the rationale for the so called *method capability index* (see Section 1.4 and Chapter 10).

Variance and standard deviation: $$s^2 = \frac{\sum (x_i - \bar{x})^2}{(n-1)} \quad s = \sqrt{s^2} \tag{2.1-2}$$

However, this relationship is based on the true standard deviation of the whole population (σ). Small data sets, normally available in pharmaceutical analysis and validation, will vary within the theoretical possible range of the whole population, and their calculated (sample) standard deviation s (Eq. 2.1-2) will scatter rather widely. In Figure 2.1-2, five series of six subsequent values each, randomly selected from the 15 000 absorbance data, are shown. The calculated standard deviations vary from 30 % to 180 % of the true value. Note that these data sets are based on the same normal distribution! The variation of the standard deviation is only due to random variability in the data, i.e., it is statistically caused. The smaller the number of data, the higher is the variability of the calculated standard deviation (Fig. 2.1-3). For small numbers of data, the standard deviation distribution is skewed towards higher values, because the left side is limited to zero. Additionally, a larger proportion of the results is observed below the theoretical standard deviation (63 %, 60 %, 59 %, and 56 %, for n=3, 4, 6, and 10, respectively). Using more data to calculate the standard deviation, the distribution becomes more narrow and symmetrical (Fig. 2.1-3, n=6 and 10). Standard deviations calculated from six values (five degrees of freedom) were found up to 1.6 times the true value (estimated from the upper limit of the 95 %-range of all results, i.e., ignoring the upper 2.5 % of results). This is important to note when *acceptance criteria* for experimental standard deviations are to be defined, since here the upper limit of their distribution is relevant. These experimentally obtained ranges were confirmed by large data sets simulated from a normal distribution (Table 2.1-1).

Variability limit (range): $$R = z\sqrt{2}\sigma = 1.96\sqrt{2}\sigma \approx 2.8\sigma \tag{2.1-3}$$

Often, an acceptable difference between individual determinations is of interest. The variability limit (Eq. 2.1-3) [20] describes the maximum range (or difference

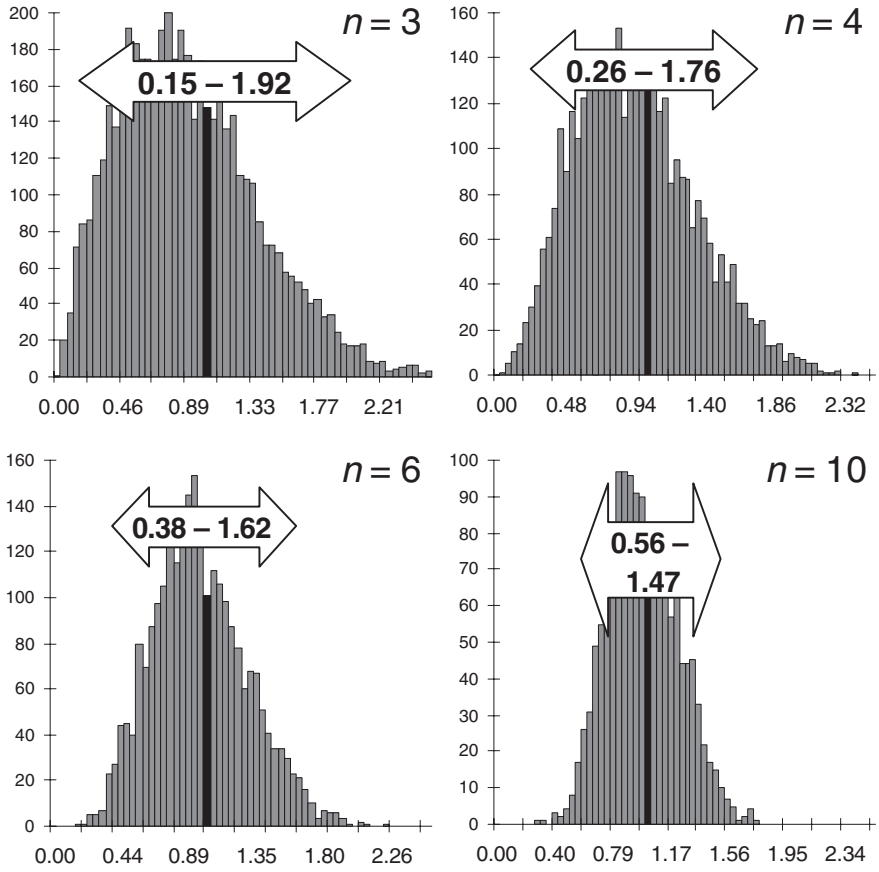


Figure 2.1-3: Distribution of standard deviations calculated from 3, 4, 6, and 10 subsequent data within the 15 000 recorded absorbance values (for details see text). In order to make a generalisation easier, the individual standard deviations are presented on the x-axes as their ratio to the overall (true) standard deviation. The y-axes represent the frequency of the standard deviations within the respective class. The ranges where 95% of all calculated values were found are indicated by double arrows, the bar containing the true standard deviation is shown in black.

between two random values) that can be statistically expected. Equation (2.1-3) is based on the true standard deviation and the normal distribution. If individual (experimental) standard deviations are used, z must be replaced by the corresponding Student-t-value. The analyst must be aware that the precision level determines the application of the variability limit, for example, with an injection repeatability, the maximum difference between two injections of the same solution is obtained, with a repeatability, the maximum range of independent sample preparations is obtained, etc.

If standard deviations are reported, it must be clearly stated to what they relate. Preferably, they should refer to single determinations. In this case, they provide information on the distribution of single data. If other calculations of the variability

Table 2.1-1 Ranges of standard deviations calculated from simulated normally distributed data sets in relation to their sample size. The normally distributed data with a true standard deviation of 1 and a mean of 100 were calculated using Equation 2.1-4.

Sample size n (df = n-1)	Lower and upper limits between the indicated percentage of 50 000 calculated standard deviations were found		
	90 %	95 %	99 %
3	0.23 – 1.73	0.16 – 1.93	0.07 – 2.29
4	0.35 – 1.61	0.27 – 1.77	0.15 – 2.07
5	0.42 – 1.54	0.35 – 1.67	0.23 – 1.92
6	0.48 – 1.49	0.41 – 1.60	0.29 – 1.82
8	0.56 – 1.42	0.49 – 1.51	0.38 – 1.69
10	0.61 – 1.37	0.55 – 1.45	0.44 – 1.61
15	0.69 – 1.30	0.63 – 1.37	0.54 – 1.49
20	0.73 – 1.26	0.69 – 1.32	0.60 – 1.42

Simulation of a normal distribution for a true mean μ and standard deviation σ (EXCEL®):

$$x = \text{NORMSINV(RAND())}\sigma + \mu \tag{2.1-4}$$

are performed, such as repeated injections for each sample preparation, the precision of the method (e.g. six times three sample preparations with duplicate injections), the standard deviation of mean results, etc., then this should be clearly described; otherwise a meaningful interpretation is not possible. Unfortunately, this is a rather frequent shortcoming in the literature.

Usually, analytical variability is reported as a relative standard deviation (RSD), i.e., divided by the respective mean. This normalisation allows a direct comparison of precisions.

An analytical procedure is always composed of many individual steps. Each of them has its own variability, and their combination results in the overall variability. In this process, the variability can only increase, also known as *error propagation* (for more details, see Section 6.2.2.1). The overall error can be calculated by the sum of all (relative) variances (uncertainty budget), also known as the bottom-up approach, to estimate *measurement uncertainty* [21]. In pharmaceutical analysis, several contributing steps are usually grouped together in the experimental design, corresponding to the precision levels (see Section 2.1.2), this is also called the top-down approach.

2.1.1.3 Confidence Intervals

The (arithmetic) mean of the measurements is an estimate of the true value of a normal distribution. The latter can be expected in a certain interval around the sample mean, the so-called confidence interval (Eq. 2.1-5). Because of the infinity of the normal distribution, data far away from the true value are theoretically possible (although with a very small probability, but this cannot predict *when* such an event will

happen), the expectation needs to be restricted to a practical range. This is represented by the error probability α, i.e., the part under the Gaussian curve, which is ignored, or the statistical confidence (or significance) level P (with P=100-α). Often, a 95 % level is chosen. The *Student-t-factor* is a correction for the (un)reliability of the experimental standard deviation obtained from a finite sample size (or strictly the underlying distribution). The term ($s/\sqrt{n}$) is also called the *standard error of the mean* and represents the variability connected to the distribution of means. Compared to the distribution of single datum, the variability of the mean is reduced, as illustrated in Figure 2.1-2 (diamonds vs. squares). The width of the confidence interval is dependent on the number of data (Fig. 2.1-4A). The more measurements that are performed, the better the mean estimates the true value. For an infinite number of data, the confidence interval of the mean approaches zero. As the confidence interval represents the range where the true value can be expected, this parameter may be useful in an out-of specification investigation (if no analytical error can be assigned) to assess whether or not a batch failure has occurred. If the whole confidence interval is within the acceptance limits of the specification, the true value can be expected to conform.

Confidence intervals:
of a mean

$$CL(P)_{\overline{x}} = \bar{x} \pm s \frac{t(P,df)}{\sqrt{n}} \tag{2.1-5}$$

of a standard deviation:

$$CL(P)_{s,\,lower} = s \sqrt{\frac{df}{\chi^2(1-P,df)}} \tag{2.1-6}$$

$$CL(P)_{s,\,upper} = s \sqrt{\frac{df}{\chi^2(P,df)}} \tag{2.1-7}$$

$t(P,df)$ = Student-t-value for the statistical confidence P (usually 95 %) and the degrees of freedom. Excel®: t= TINV(α, df); α = 1-P

$\chi^2(P,df)$ = Chi-squared value for the statistical confidence P (usually 95 %) and the degrees of freedom. Excel®: χ^2 = CHIINV(α, df); α = 1-P

This behaviour is the same with respect to the experimental and the true standard deviation (Eq. 2.1-6, 7), but the uncertainty of standard deviations is much larger than those of means (Fig. 2.1-4B). Whereas the true value can be expected (with a 95 % confidence) to be less than 1.05 (experimental) standard deviations away from the mean in the case of n=6 data, σ may deviate up to 2.09 standard deviations from the calculated one. In contrast to the confidence intervals of the mean that are symmetrical (as there is the same probability for measurements to deviate above and below the true value), the confidence intervals of standard deviations are non-symmetrical, because they are limited to the lower side by zero. The confidence intervals in Figure 2.1-4B describe the possible range of the true value with respect to the (experimentally) estimated parameter for the given set of data, whereas the distributions in Figure 2.1-3 and Table 2.1-1 represent the range of standard deviations obtained for independent, repeated data sets with respect to the true standard deviation.

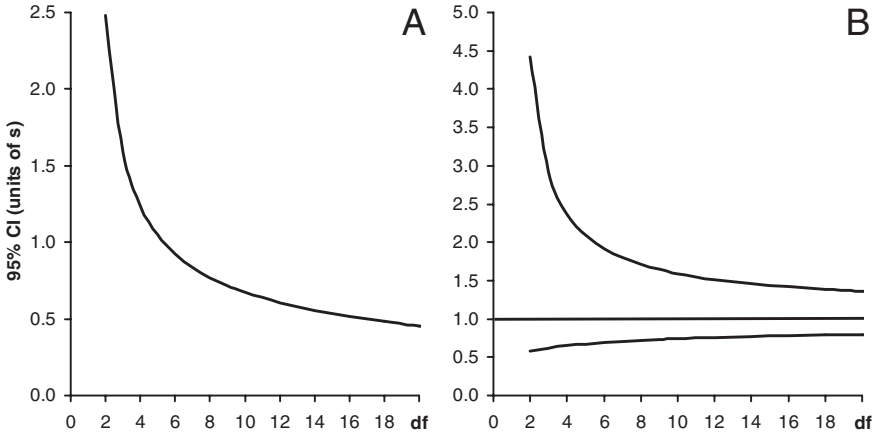


Figure 2.1-4: 95% confidence intervals of the mean (A) and of the standard deviation (B) as a function of the degrees of freedom (df) on which the calculation is based. The confidence intervals (CI) are displayed in units of the standard deviation. For one data set (one run or series) the degree of freedom is $n-1$, for several independent data sets k, the degree of freedom corresponds to $k*(n-1)$.

From Figure 2.1-4B it is also obvious that a *standard deviation calculated from three values* ($df=2$) (unfortunately not an exception in validation literature) is rather *meaningless* as σ can be expected up to 4.4fold of the calculated standard deviation! However, if several sets of data can be combined (pooled), the overall degrees of freedom and thus the reliability are increased. In such a case, only the overall standard deviation should be reported. A prerequisite for such pooling of data is that all data sets must have similar σ (if means are looked at, they must also have the same true mean; for verification, see the discussion on precision level in Section 2.1.2). Interestingly, a confidence interval is mentioned in the ICH guideline (although it is not clearly stated whether with respect to the standard deviation or to the mean) [1b]. However, the author is not aware of any publication on pharmaceutical validation which reports it. Following the standard approach with six or more determinations for a standard deviation, the confidence interval will not provide much additional information, but the benefit could be to cause people to hesitate before reporting standard deviations from three determinations only.

Significance Tests

Confidence intervals are also the basis of statistical tests. In the case of *significance tests*, the test hypothesis (H0) assumes, for example, no difference (zero) between two mean results. This is fulfilled (or strictly, the hypothesis cannot be rejected), when the two confidence intervals overlap. However, as the confidence intervals become tighter with increasing number of determinations, (theoretically) any difference – however small – can be shown to be significant. For example, assuming a standard deviation of 0.5, a difference of 0.5 is significant with nine determinations, but even a difference of 0.1 will become significant when there are 200 values. Of course, this is (usually) not of (practical) interest (see Accuracy, Section 2.3.1).

2.1.1.4 Robust Parameters

The above-described parameters are based on the assumption of a normal distribution. If this prerequisite is not fulfilled, or disturbed, for example by a single deviating result (outlier), the calculated parameters are directly impacted. This influence is decreased by the application of robust parameters that are not based on specific assumptions [22, 23]. The analogue to the arithmetic mean is the median, i.e., the middle value in an ordered sequence of results. A comparison between mean and median may provide information about a possible disturbance in the data. However, it is often a very complex matter to estimate confidence intervals or variabilities for robust parameters.

Another alternative to estimate description parameters of any distribution is the (thousand fold) repeated calculation from an experimental set of data (re-sampling) to achieve a simulated distribution, the so-called bootstrap [24], or the estimation of variability from the noise of a single measurement using a probability theory named the 'function of mutual information' (FUMI) [25]. However, these techniques are beyond the scope of this book, and the reader is referred to specialised literature.

2.1.2 Precision Levels

Regarding an analytical procedure, each of the steps will contribute to the overall variability (see also Fig. 10-4). Therefore, the overall uncertainty can be estimated by summing up each of the contributing variabilities, the so-called bottom-up approach [21, 26]. However, this approach is quite complex because each and every step has not only to be taken into account, but also its variability must be known or determined.

Alternatively, the other approach (top-down) usually applied in pharmaceutical analysis combines groups of contributions obtained experimentally, i.e., the precision levels. Such a holistic approach is easier to apply, because each of the individual contributing

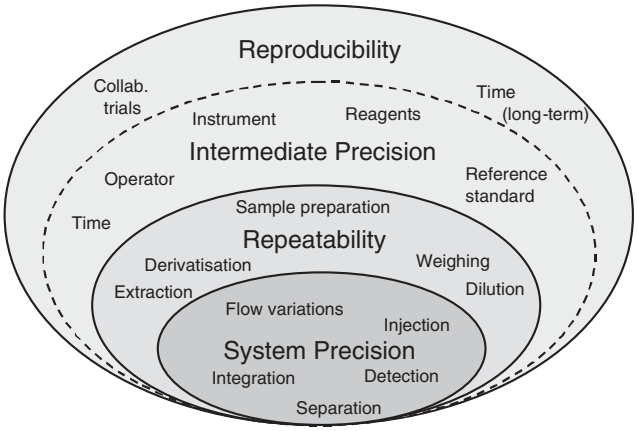


Figure 2.1-5: Illustration of the various precision levels with (some of) their contributions.

steps does not need to be known specifically. However, this may lead to misinterpretations and wrong conclusions, if the analyst is not aware of the correct level/contributions. Basically, short-term and long-term contributions can be distinguished, with system precision and repeatability belonging to the former, intermediate precision and reproducibility to the latter. Each of the levels includes the lower ones (Fig. 2.1-5).

2.1.2.1 System or Instrument Precision

The variability of the measurement itself is addressed in system precision, also termed instrument/injection precision, or injection repeatability (although the latter term is didactically not well chosen, because it may easily be confused with the (real) repeatability, see below). Although in LC the contribution from the injection system is normally the dominating one (at least at higher concentrations, see below), there are additional contributions from the pump (short-term flow constancy, relevant for peak area measurements), the separation process, and the detection/integration. (Consequently, the term 'system precision' is the best to describe this level.) The variability contribution due to short-term flow fluctuations can be separated from the overall system variance by analysing substance mixtures and subtracting the variance of the relative peak area, the so-called Maldener test, originally proposed for equipment qualification [27]. Using ten injections of an about equal mixture of methyl-, ethyl-, propyl- and butyl-esters of 4-hydroxybenzoic acid, precisions of the relative peak area between 0.04 and 0.12 % were obtained, corresponding to an error contribution of the pump of between 5 and 22 % (on a variance basis). The smaller the overall system precision, the larger is that contribution.

Variance and standard deviation from duplicates (differences):

$$s_d^2 = \frac{\sum (x_{i,1} - x_{i,2})^2}{2k} \quad s_d = \sqrt{s_d^2} \tag{2.1-8}$$

$\underline{k}$ = number of samples or batches analysed in duplicates

System precision is obtained by repeated analysis of the same sample (solution) and can be calculated using Eq.(2.1-2) for a larger number of analyses (at least five), or according to Eq.(2.1-8) from a sufficient number of duplicates. Although unfortunately not described in the ICH guidelines, system precision provides valuable information about the variability of the analytical system, mainly the instrument. Therefore, it is an important parameter for equipment qualification (see Chapter 4) and for System Suitability Tests (see Section 2.8). However, in order to reflect mainly the performance of the instrument, for these applications the analyte concentration needs to be sufficiently above the quantitation limit (at least 100 times), otherwise the contributions of the detection/integration errors will increase (Fig. 2.1-7 and Section 2.1.3.1, also Table 2.8-13).

2.1.2.2 Repeatability

This short-term variability includes, in addition to the system precision, the contributions from the sample preparation, such as weighing, aliquoting, dilution, extraction, homogenisation, etc. Therefore, it is essential to apply the *whole analytical procedure*

(as described in the control test), rather merely to injecting the same sample solution six times. This is also the reason for using *authentic samples* [1b], because only then can the analytical procedure be performed exactly as in the routine application. There may be exceptions, but these should be demonstrated or cautiously justified. For example, analysing degradants near the quantitation limit, where the variance contribution of the sample preparation can be neglected, injection precision and repeatability are identical (Figs. 2.1-7 and 2.1-8). For some applications, where precision can be regarded as less critical, such as in early development (see Chapter 5), or if the variability demands only a small part of specification range (less than approximately 10 %), or if the expected content of impurities is far away from the specification limit, artificially prepared (spiked) samples may be used, allowing several validation characteristics (linearity, precision and accuracy) to be addressed simultaneously.

Repeatability can be calculated using Eq.(2.1-2) from a larger number of repeatedly prepared samples (at least 6), or according to Eq.(2.1-8) from a sufficient number of duplicate sample preparations. Calculations should not be performed with smaller number of data due to the large uncertainty involved (Fig. 2.1-4B). The true standard deviation may be up to 4.4 times greater than a result obtained from three determinations!

2.1.2.3 **Intermediate Precision and Reproducibility**

Intermediate precision includes the influence of additional random effects according to the intended use of the procedure in the same laboratory and can be regarded as an (initial) estimate for the long-term variability. Relevant factors, such as operator, instrument, and days should be varied. Intermediate precision is obtained from several independent series of applications of the (whole) analytical procedure to (preferably) authentic, identical samples. In case of relative techniques, the preparation and analysis of the reference standard is an important variability contribution. Therefore, it is not appropriate to determine intermediate precision from the peak area of the sample alone (analysed on different days or even several concentrations only), as is sometimes reported in validation literature. Apart from ignoring the contribution of the reference standard, any signal shift of the instrument will be falsely interpreted as random variability. In order to reflect the expected routine variability properly, the calibration must be performed exactly as described in the control test.

Reproducibility, according to the ICH definition is obtained varying further factors *between laboratories* and is particularly important in the assessment of 'official' compendial methods or if the method is applied at different sites. However, understood in the long-term perspective, both intermediate precision and reproducibility approach each other, at least in the same company. Reproducibility from collaborative trials can be expected to include additional contributions due to a probably larger difference of knowledge, experience, equipment, etc. among the participating laboratories.

Analysis of variances

It is very important to address intermediate precision/reproducibility appropriately as it is an estimate for the variability (and robustness) to be expected in long-term applications, such as in stability testing. According to ICH, standard deviations

should be calculated for each level of precision. They may be calculated by means of an *analysis of variances* (ANOVA) [20]. In a (one-way) ANOVA, the overall variability is separated into the contributions within and between the series, allowing the assessment of the most sensitive part of the analytical procedure as well as its robustness (or ruggedness according to USP [5]). However, only a positive robustness statement is possible. When there is unacceptable difference between the precision levels (which does not necessarily mean significant differences between the series means (see 1.4.2)), the cause needs to be identified by investigation of the effect of the various factors individually (see Section 2.7). The intermediate precision/reproducibility is calculated from the overall variance (Eq. 2.1-11), i.e. the sum of the variances within (Eq. 2.1-9) and between (Eq. 2.1-10) the series. The latter corresponds to the additional variability caused by the factors that were varied in the experimental design (operator, equipment, time, laboratory etc.) of the various series. In case of a numerically negative term, s_g^2 is set to zero, because practically, variability can only increase.

Analysis of variances (one-way):

$$\text{Intra-serial variance: } s_r^2 = \frac{\sum\left((n_j-1)s_j^2\right)}{\sum n_j - k} \text{ or } s_r^2 = \frac{\sum\left(s_j^2\right)}{k} \text{ (with equal } n\text{)} \qquad (2.1\text{-}9)$$

Inter-serial variance:

$$s_g^2 = \left(\frac{\sum\left(n_j\bar{x}_j^2\right)\sum n_j - \left(\sum\left(n_j\bar{x}_j\right)\right)^2}{(k-1)\sum n_j} - s_r^2\right)\frac{(k-1)\sum n_j}{\left(\sum n_j\right)^2 - \sum\left(n_j^2\right)}$$

$$\text{or } s_g^2 = \frac{\sum\left(\bar{x}_j^2\right) - \frac{\left(\sum \bar{x}_j\right)^2}{k}}{k-1} - \frac{s_r^2}{n} = s_{\bar{x}}^2 - \frac{s_r^2}{n} \text{ (with equal } n\text{)} \qquad (2.1\text{-}10)$$

$$s_R^2 = s_r^2 + s_g^2 \quad \text{if } s_g^2 < 0\text{: } s_R^2 = s_r^2 \quad s_R = \sqrt{s_R^2} \qquad (2.1\text{-}11)$$

$n_j, s_j, \bar{x}_j$ = Number of determinations, standard deviation, and mean of series j
k = Number of series (for the given batch)
$s_{\bar{x}}$ = Standard deviation of the means

In a strict statistical sense, the homogeneity of the variances s_j^2 and the absence of a significant difference between the means $\bar{x}_j$ need to be tested, which may pose the already discussed problems of statistical significance and practical relevance (see Section 1.4.2). This is especially true for types of procedures where the variability contribution from the reference standard (or any other long-term factor) is known to be larger than the repeatability, as in the case of content determination of injection solutions (see Section 2.1.3.2, Fig. 2.1-9). A recommended pragmatic solution consists in defining *absolute upper limits* for the various precision levels [28]. The difference between the means will directly influence the inter-group variance (Eq. 2.1-10) and con-

sequently the intermediate precision/reproducibility (Eq. 2.1-11). It can be controlled by setting limits for this precision. Alternatively, a maximum absolute difference between the (most deviating) means can be established as a direct acceptance criterion.

Another possibility is to calculate intermediate precision/reproducibility by simply using Eq. (2.1-2) for the pooled data from all series. The justification for combining the series should be based again preferably on compliance to absolute acceptance criteria, as previously discussed.

The approaches described above result in two precision levels, i.e., they combine, for intermediate precision/reproducibility, the effects of all factors that were varied. If of interest, these factors can be investigated individually by means of (fully or staggered) nested experimental design, also called multivariate or factorial design. A multifactorial design will provide the variance contributions of the individual experimental variables, as well as the interactions between them [29, 30]. Usually, the variables are combined, if none of the factors is significant [31–33].

The ICH guideline provides no guidance on the number of determinations or series for the estimation of intermediate precision/reproducibility. However, the basic relationship between the number of determinations (or strictly degrees of freedom) and the reliability of the standard deviation (Fig. 2.1-4B) should be considered. The simplest approach is to perform further repeatability studies with six determinations, varying the operator and/or equipment. In the case of two series, the intermediate precision is based on ten degrees of freedom, and the data can also be used for the determination of individual repeatabilities. Of course, each of the repeatability series must be performed *independently*, including the whole sample preparation and calibration. The more series that are performed, the more variations/combinations of factors (e.g. time, operator, equipment, reagents, etc.) can be covered and the more reliable are the results obtained. Then, the number of determinations within each series can be reduced. Examples from the literature include two operators analysing on two days using two instruments and three samples each (24 results, 16 degrees of freedom) [31, 34], two repetitions (for several batches) on seven days (14 results, seven degrees of freedom) [35], and the Japanese approach of varying six factors (by analogy with the ICH request for repeatability), such as two operators, two instruments, and two columns in a randomised way, with two repetitions each (12 results, six degrees of freedom) [36, 37]. However, it is obvious from the overall degrees of freedom, that the last two approaches do not have a large improvement in reliability. Another approach may consist in using the number of sample preparations prescribed in the control test, of course with an appropriate number of independent series, varying factors that are relevant for the routine application. The standard deviation of the means would then correspond directly to the analytical variability of the batch release procedure. In Table 2.1-2, an example is shown with four series of six determinations each for a lyophilisate sample, performed in two laboratories by different analysts and equipment. Whichever approach is chosen, for a sensible evaluation and interpretation, the precision level should be clearly distinguished and the experimental design and calculations sufficiently described in the documentation.

Table 2.1-2: Calculation of intermediate precision by means of analysis of variances [28]

	Laboratory 1		Laboratory 2	
	Analyst A	**Analyst B**	**Analyst C**	**Analyst D**
	Content (percent label claim)			
	99.84	100.21	98.27	99.41
	99.93	99.31	99.31	99.41
	99.50	99.86	98.26	99.23
	100.24	100.59	99.43	99.91
	101.30	100.54	100.01	99.13
	102.00	100.70	99.76	98.86
Mean	100.47	100.20	99.17	99.33
RSD	0.97 %	0.53 %	0.75 %	0.36 %
Cochran test for homogeneity of variances (95 % confidence level)				
		Test value	0.49	
		Critical value	0.59	
		Variances are homogeneous		
Analysis of variances (one-way)				
		Intra-serial variance	0.4768	
		Inter-serial variance	0.3292	
		Overall variance	0.8060	
		Intra-serial variance > inter-serial variance:		
		No significant difference of the means.		
		Overall mean	99.79	
		95 % Confidence interval	99.41 – 100.17	
		Overall repeatability	**0.69 %**	
		Intermediate precision	**0.90 %**	

2.1.3 Acceptable Ranges for Precisions

The minimum requirements for the analytical variability originate from the acceptance criteria of the specification (see Section 1.4) [14]. However, at least for drug products, better precisions can usually be achieved. Of course, due to the additional variance components, acceptance criteria should be defined for each level of precision separately.

For the same level of precision, some conclusions can be drawn from the distribution of standard deviations, as shown in the following, for repeatability. For the purpose of evaluation, it is important to distinguish between individual repeatabilities (s_j) and the overall repeatability (s_r). The former can vary in a certain range around the true standard deviation (depending on the number of determinations, Fig. 2.1-3 and Table 2.1-1), but for the question of acceptability, the *upper limit* of the distribution is relevant. The latter is an average (pooled) standard deviation describing the variability within the series, and therefore, due to the increased degrees of freedom

gives a better estimate of the true (repeatability) standard deviation (Fig. 2.1-4B). Therefore, this parameter is also termed the *target (repeatability) standard deviation (TSD)* [18]. For six determinations, the upper limit of the 95 % range of standard deviations is 1.6 σ (Table 2.1-1). Because there is always some uncertainty even for quite reliable TSDs (the lower 95 % limit for df =20 is 0.7), a statistical distribution range of up to 2.3 TSD can be expected for individual repeatabilities. This corresponds well with the upper 95 % confidence limit of 2.1 for a standard deviation from five degrees of freedom. Overall repeatabilities can be expected to be smaller for a factor of 0.85 to 0.8 (see Table 2.1-1).

No a priori conclusion can be drawn for the relationship between the precision levels, because of additional variance contributions.

In the pharmaceutical area, systematic investigations of experimentally obtained precisions for a variety of analytes and/or techniques are not very frequent; some of the papers having been published about 25 years ago. Some results are summarised in Table 2.1-3 (see also Table 6-4). According to the widespread utilisation of LC methods, more (current) information is available for them (see Fig. 2.1-9 and 10).

Table 2.1-3 Collection of published precision results for various analytical techniques in pharmaceutical applications.

Analytical technique	Samples / Remarks	Repeatability (%)	Reproducibility / intermediate precision (%)
Gas chromatography	4 CT, 8 drugs [19]	1.3	2.6
	7 CT, 12 drugs [38]	1.25 ± 0.54	2.41 ± 0.85 increase of 0.7 % for 10fold decrease in concentration
	Estimation from instrument precision [39]	Direct: 1.5 Headspace: 2.3	2.2 3.5
LC	11 CT, 12 analytes [40]	1.00 ± 0.35	2.50 ± 0.85 increase of 0.4 % for 10 fold decrease in concentration
	CT, cloxacillin [41]	0.6 (0.11 – 1.22)	0.8 (recalculated)
	CT, various antibiotics [41]		0.2 – 1.5 (standard deviation of means!)
	Estimation from instrument precision [39]	1.1 – 1.5 0.6 – 0.8 (automated)	1.6 – 2.2 0.9 – 1.1
UV Spectrometry	5 CT, 5 drugs [19]	1.1	2.5
	CT, 9 analytes [42]	1.21 ± 0.63	2.34 ± 1.04
	CT, prednisolone [41]	0.6 (0.02 – 1.68)	1.5 (recalculated)
	CT, cinnarizine [41]	0.8 (0.04 – 1.28)	2.0 (recalculated)
	CT, dienestrol, albendazole, methyl-prednisolone [41]	0.9, 1.3, 1.0	2.2, 2.7, 3.5 (recalculated)
	Automated [43]	1.1 – 2.8	1.2 – 3.3

Table 2.1 Continued.

Analytical technique	Samples / Remarks	Repeatability (%)	Reproducibility / intermediate precision (%)
Polarography	CT, 4 analytes [42]	1.73 ± 0.57	2.60 ± 0.88
Titration	CT, ephedrine hydrochloride [18]	0.41 (0.03 – 0.70)	0.67 (recalculated)
	CT, 3 analytes [44]	0.28 (0.02 – 0.52)	
	85 CT, 72 analytes [45]		1.6 prediction for concentration of: 100 % – 0.96 1 % – 2.8 0.01 % – 8.6
CE	Internal standards [46-48]	< 1 %	
	Mirtazapin [49]	C=100 %: 1.7 C=0.2 %: 12 – 28	
	CT, 5 participants, sodium [50]	C=5.5 %: 1.4 (0.5 – 2.6)	
	Estimation from instrument precision [39]	1.5	2.2
	Antibiotics [51]	1.5 – 2.4	2.0 – 2.8
LC	CT, penicillin impurity [52]		C~1 %: 18.2, 35.3, 24.1
	CT, oxacillin, sum of impurities [53]		C=0.45 %: 39.4 C=1.25 %: 24.4 C=1.02 %: 30.5 C=0.22 %: 7.66
	CT, dicloxacillin, sum of impurities [54]		C=0.36 %: 33.8 C=1.63 %: 20.8 C=1.83 %: 16.9 C=0.40 %: 25.0
	CT, dicloxacillin, indiv. impurity [54]		C=0.29 %: 4.2 C=0.41 %: 7.6 C=0.13 %: 15.4 C=0.33 %: 3.9
NMR	CT, disulfiram [42]		2.67 ± 0.48
	[55]	< 2 %	
	Herbicides, standards [56]	0.1 – 1.1	0.2 – 0.7
Colorimetry	CT, disulfiram [57]		2.6, 2.4
AAS	Magnesium [58]	0.6	1.1, 1.6
	CT, 4 analytes [42]		3.18 ± 0.86
	Palladium [59]	C~2 ppm: 4.7, 1.8, 4.3 C~20 ppm: 0.6	

Table 2.1 Continued.

Analytical technique	Samples / Remarks	Repeatability (%)	Reproducibility / intermediate precision (%)
TLC (densitometric detection)	Estimation from instrument precision [39]	2.1 – 2.9	3.2 – 4.3
Impactors	CT [60] salbutamol,		
	total deliverd dose	4.5 – 6.0	8.2 – 10.0
	Fine particle dose	6.3 – 7.8	8.6 –15.3
NIR (quantitative)	Caffeine tablets [61]	0.55; 0.74	0.61; 0.48
	Tolbutamide tablets [62]		1.0
Ion chromatography	CT, fluoride [42]	1.16	2.70 ± 0.66

Abbreviations: CT = collaborative trials; C = concentration

2.1.3.1 Concentration Dependency of Precision

Examining a large number of data sets from collaborative trials covering various analytes, matrices and analytical techniques over large concentration ranges, Horwitz et al. found a strikingly simple exponential relationship between the relative standard deviation among laboratories (i.e. reproducibility) to the concentration of the analyte, expressed as a mass fraction (Fig. 10-5, see also Section 10.4). It describes, in accordance with observations, that the standard deviation decreases less rapidly than the concentration, resulting in an increase in the relative standard deviation for lower concentrations. The Horwitz curve is widely used as an initial estimate of expected reproducibility and as a benchmark for the performance of laboratories in a collaborative study: *"Acceptable performance usually provides variability values within one-half to twice the value predicted by the equation from the concentration. Within-laboratory variability is expected to be one-half to two-thirds the among-laboratory variability."* [63]

Whilst excellent for use in describing the general concentration dependency of precision and providing orientation within large concentration ranges, a different behaviour is observed for limited concentration ranges when applying the same technique in pharmaceutical analysis. If sufficiently above the quantitation limit, there is only a small concentration dependency on the precision, which is more influenced by the sample composition (i.e., drug product types, Figs 2.1-9 and 2.1-10) [64]. This may be due to additional variability effects in collaborative trials, thus the reason for outlier testing and removal [19]. For in-house applications, the experience and control of the method is greater. It can also be expected that this contribution to the variability becomes larger for very small concentrations due to more complex sample preparation and matrix interferences. Also, there are also some logical inconsistencies. Take, for example, a drug substance LC assay with a test solution of 0.1mg/ml. According to the concentration fraction of the original sample of about 100 %, the corresponding reproducibility range should be between 1.0 and 4.0 %,

the repeatability between 1.0 and 1.3 %. If it is assumed that the same drug substance is formulated in an injection solution of 0.1mg/ml, which is directly injected for assay, the concentration fraction is now 0.01 %, corresponding to a reproducibility between 4.0 and 16.0 % and a repeatability between 4.0 and 5.3 %. In practice, the same variability will be found for both samples (perhaps even a bit more for the drug substance due to the weighing step).

Approaching the quantitation limit, the relative standard deviation increases much more sharply than predicted by the Horwitz function. In Figure 2.1-6, a precision investigation over a large concentration range is shown [65]. The repeatability was calculated from five to seven sample preparations of a reconstituted model drug product, consisting of Glibenclamide and tablet formulation excipients. The concentration fraction of the active at the nominal test concentration was 5.26 % (5mg active and 90 mg excipients). From the repeatability and system precision of the higher concentrations (above 10 %), the sample preparation and injection errors were estimated. The latter can be assumed to correspond directly to the system precision, the former is obtained from the difference between the variances of the two precision levels. Because both sample preparation and injection error can be assumed to remain constant (if expressed as relative values), the increasing overall variability can be attributed to the *integration error* that becomes dominant in the lower concentration range (Fig. 2.1-7). This is also the reason that injection precision and repeatability approach each other at lower concentrations. As shown in Figure 2.1-8, this is the case below 5 to 10 %, corresponding to about 100 times the quanti-

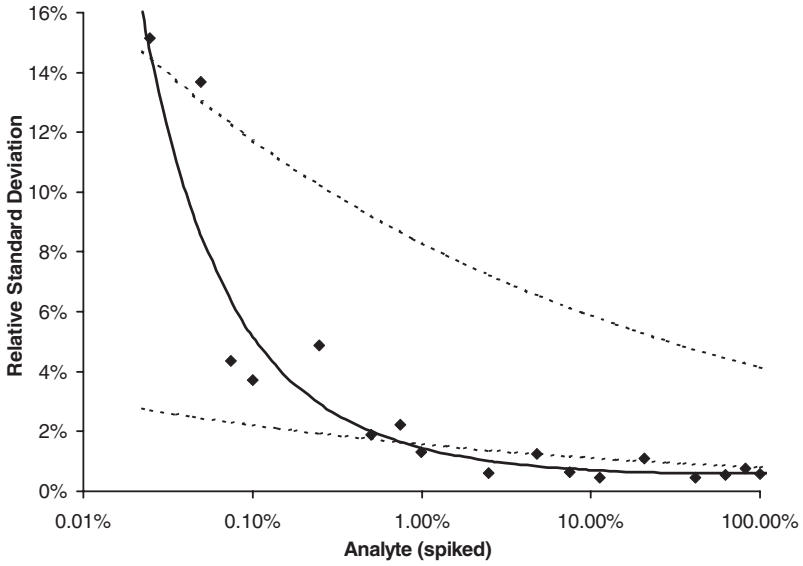


Figure 2.1-6: Repeatability investigations for an LC assay over a concentration range from 0.025 to 100% (data from [65]). The solid line illustrates the repeatability trend, obtained by quadratic regression of the logarithm of the standard deviation versus the logarithm of the concentration. The broken lines indicate the limits of the Horwitz range for repeatability [63].

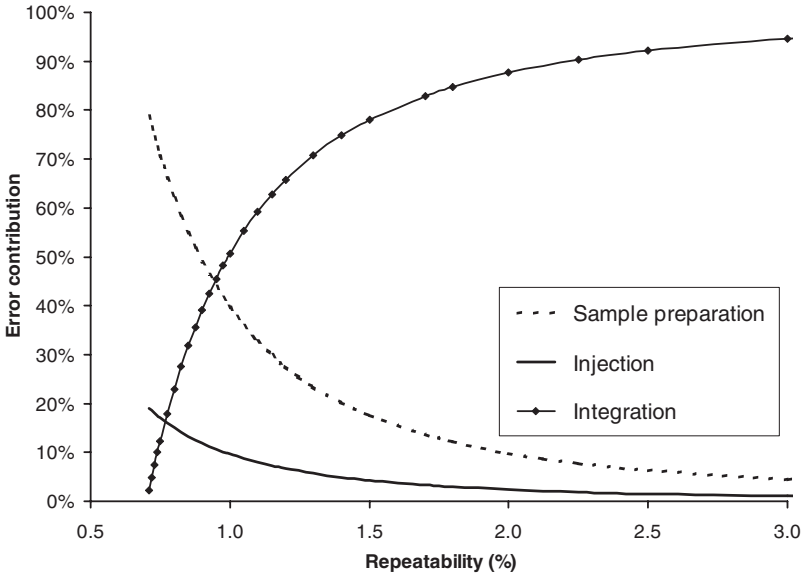


Figure 2.1-7: Error contributions to an increasing repeatability (data from [65]). The sample preparation error and injection error were estimated to 0.63% RSD and 0.31% RSD, respectively, calculated from seven data sets of five – seven samples each, between 10% and 120% of the nominal test concentration.

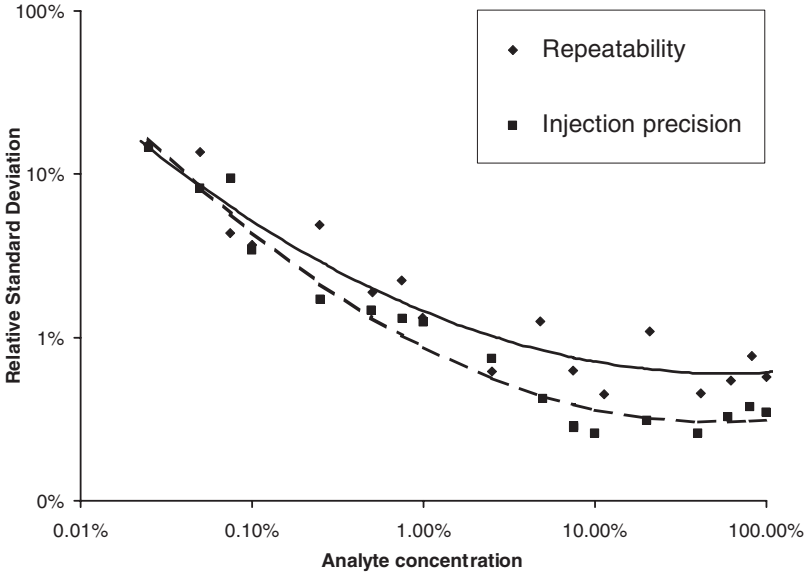


Figure 2.1-8: Investigations of injection precision and repeatability for an LC assay over a concentration range from 0.025 to 100% (data from [65]. The solid and broken lines indicate the repeatability and injection precision curves, respectively, and were obtained by quadratic regression of the logarithm of the respective standard deviations versus the logarithm of the concentration.

tation limit. In contrast, at higher concentrations, the sample preparation error often dominates (Fig. 2.1-7), in dependence on the sample (preparation) complexity, which is usually linked to a given type of drug product (see next section). Consequently, the utilisation of *internal standards* is only sensible if the injection error is the dominating error source (or in the case of matrix interferences).

Because of the direct relation between integration error and measurement noise level, an acceptable precision for impurity concentrations can be generated by linking to the quantitation limit, as proposed in [66] (Table 2.1-4).

Table 2.1-4 Acceptable precisions for impurities. The concentration ranges are normalised with respect to the quantitation limit (QL). In this context, QL must be regarded as the intrinsic quantitation limit of the repective analytical procedure (see also Section 2.6).

Concentration range	**Maximum acceptable precision [66]**		**Repeatability (concentration/QL[1]) [65]**	
	Repeatability	Reproducibility	Glimepiride-urethane	Glimepiride-3-isomer
QL to < 2xQL	25.0 %	30.0 %	17.0 % (0.8) 5.2 % (1.6)	14.1 % (0.8) 14.9 % (1.2)
2xQL to < 10xQL	15.0 %	20.0 %	5.6 % (2.5) 4.3 % (3.3) 2.4 % (8.0)	8.7 % (1.6) 4.5 % (4.1)
10xQL to < 20xQL	10.0 %	15.0 %	3.1 % (16)	2.3 % (12)
≥ 20xQL	5.0 %	10.0 %	1.3 % (32)	

1: Calculation from residual standard deviation of an unweighted linear regression in a range from ~QL to 3-5 times QL.

2.1.3.2 Precisions for LC Assay

System precision

In EP, for chromatographic assay of drug substance, a maximum permitted *system precision* is defined, in relation to the upper specification limit and the number of injections [15]. The difference between the upper specification limit and 100 % corresponds to the range available for the analytical variability, because the content of a drug substance cannot be larger than 100 % (see Section 2.8.3.8). An analytically available upper range of 2.0 %, for example, allows a relative standard deviation of 0.73 % and 0.85 %, for five and six injections, respectively. The FDA [3] and Canadian guidelines [7] recommend system precisions of less than 1.0 %. The 2.0 % acceptance criterion of the USP cannot be regarded as suitable for a system precision [14]. Results for auto sampler variability range between 0.3 and 0.7 %, but if not controlled by internal (qualification) specifications, can even be between 0.12 and 2.1 % (mean 0.76 ± 0.23 %). For system suitability tests, results between 0.5 and 1.2 % [67], or even from 0.7 to 1.0 % [39] have been reported. In Figure 2.1-9, some results for system precisions from literature are summarised [34, 98, 218–229]. The relative standard deviations range between 0.06 and 1.90 %, with an average of 0.92 %. About three-quarters of all results are less than 1.0 %. However, it must be

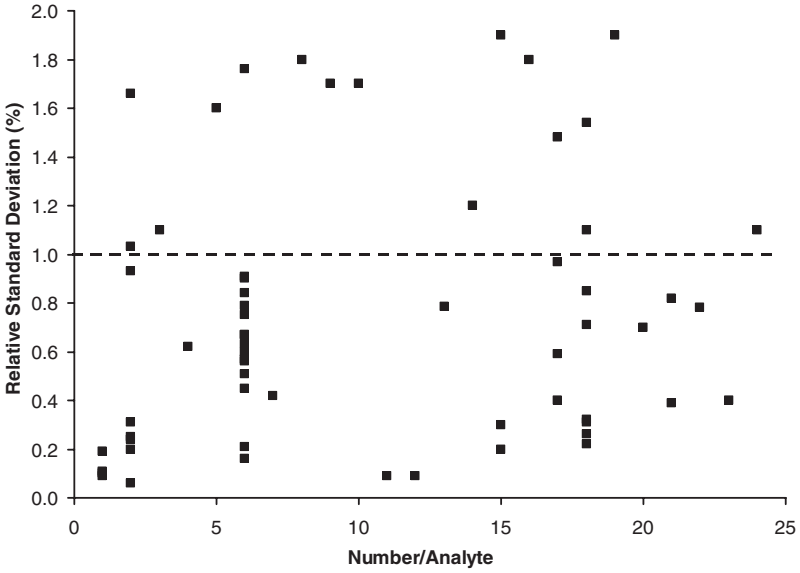


Figure 2.1-9: System precision results from literature.

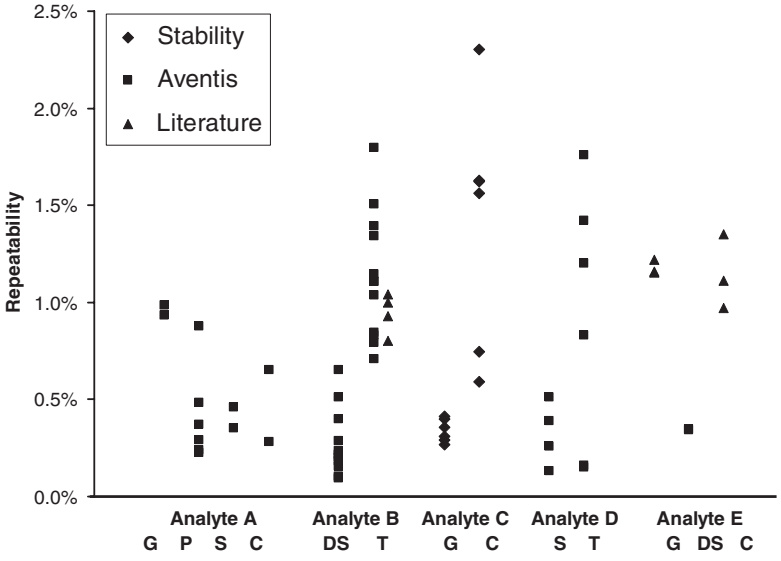


Figure 2.1-10: Repeatabilities for LC assay sorted according to the analyte. The letters correspond to the type of drug product (G = gel, P = powder, S = solution, C = cream, DS = drug substance, T = tablet). The origin of the data is indicated by the symbols and explained in the text.

taken into consideration that in the reported procedures, sometimes comparability is difficult, because insufficient information is provided about the relation of the test concentration to the quantitation limit. Therefore, the reported injection precision may partly include contributions from the integration error (see Section 2.1.3.1).

Repeatability

Results for individual repeatabilities and intermediate precision/reproducibility are shown in Figures 2.1-11 and 2.1-12, grouped according to the type of sample (drug product). The data originate from validation studies, comparative studies during method transfer, and stability investigations (see also Section 2.1.4.1) performed by Aventis, from a collaborative project to obtain precisions from stability studies (organised from the Working Group Drug Quality Control / Pharmaceutical Analytics of the German Pharmaceutical Society, DPhG [64]), and from literature. In the latter case, to keep the result recent, only papers from 1990 onwards were included [31, 34, 53, 218–261].

The results show no clear dependency on the analyte (Fig. 2.1-10), but rather on the type of drug product. For the results from Aventis and the DPhG project, the limits of the range of 95 % of all results for each subgroup (if a sufficient number of data was available) were identified and average values were calculated for these ranges (Table 2.1-5). The upper and lower 2.5 % of the results were ignored to minimise the influence of extremes. The average values can be regarded as an estimate for the true or *target standard deviation* for this group. The upper limit of the 95 % range can serve as an estimate for a maximum acceptable distribution range of (individual) repeatabilities of the respective subgroup. The results from the literature were not included in the calculation, due to the lower degree of reliability, compared with the other two sources, as already discussed for system precision. In addition, the analytical procedures were sometimes optimised for the simultaneous determination of several analytes [218, 219, 236, 254, 255, 260, 261] and may therefore not be directly comparable to procedures optimised for a single active ingredient.

The distribution of the individual repeatabilities reflect the complexity of the sample and/or its preparation. The RSDs for drug substances, solutions and lyophilisates have an average of about 0.5 – 0.6 %, the latter at the upper end of this range. This target value corresponds well to the results of 0.6 % from a collaborative trial of the EP for the LC assay of cloxacillin [41]. The LC-assay for tablets and creams is accompanied by a higher variability of approximately 0.9 –1.0 %. For all groups, the ratio between the upper limit and the average repeatability is about 2, which corresponds very well to the ratio based on the theoretical considerations given at the beginning of Section 2.1.3. For other types of drug product, the number of data and/or analytes available are not sufficient to estimate target variabilities or ranges. However, the much higher results obtained for baths, emulsions or chewing gums, confirm the dependence of the repeatability from the sample (preparation). For such kind of drug products the inhomogeneity of the sample or during sampling can also be expected to play a role. Repeatabilities obtained from literature are in principle consistent with the results from internal data collection and the DPhG project, apart from some data for solution and drug substances. This may be partly attributed to

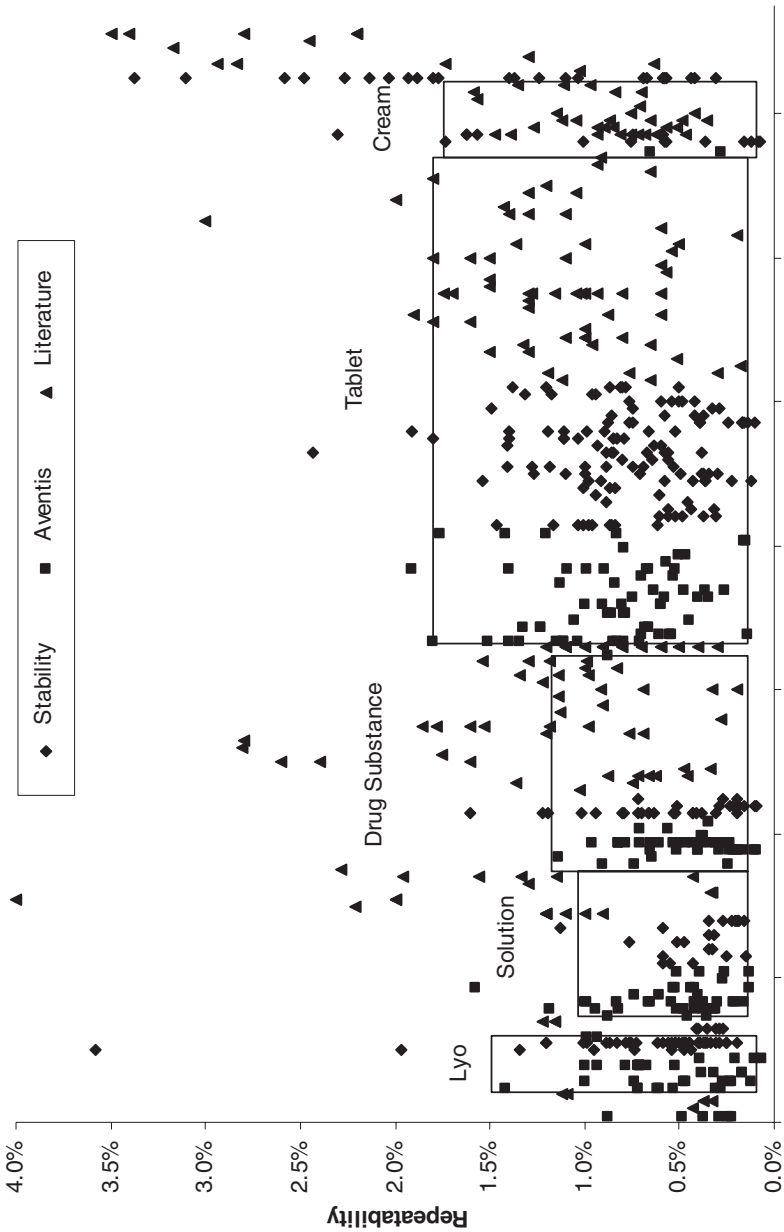


Figure 2.1-11: Repeatabilities for LC assay sorted according to the type of sample (drug product). The sequence on the x-axis corresponds to different analytes or sources. The origin of the data is indicated by the symbols and explained in the text. For sample types with sufficient results available (Aventis and Stability origin), the 95% distribution ranges are shown by rectangles.

the aforementioned reasons, but it can also be expected that some analytes / methods require larger variabilities. Therefore, the target values and distribution ranges discussed should be regarded as *orientation* for typical applications.

Intermediate precision/reproducibility

In the case of intermediate precision/reproducibility, the averages (target values) are between 1.4 and 2 times larger than for repeatability, reflecting the additional variability contributions. They range from 1.0 % for lyophilisates and drug substances to 1.1 % for solutions and 1.2 % and 1.6 % for tablets and creams, respectively. Solutions approach more closely to the variability of tablets.

Ratio

If repeatability and intermediate precision/reproducibility were obtained for the same sample, the ratio between them was calculated to estimate the factor between the precision levels (Fig. 2.1-13). In contrast to the ratio calculated from the average repeatability and reproducibility, which represent target values for the respective group of samples, these are individual ratios for the given analyte samples. A classification of these factors would allow the prediction of the long-term variability of given analytical procedures from repeatability determinations.

The smallest possible ratio is 1.0, i.e., no additional variability between the series is observed and both precision levels have the same standard deviation. Experimentally, this can occur even if the true ratio is larger than one, if one or several experimental repeatabilities are obtained in the upper range of the distribution, thus covering the differences between the series. The upper 90 % distribution limit of the ratios was determined to be between 2.5 and 3 for creams, tablets and drug substances and about 2 for lyophilisates. For solutions, markedly larger ratios up to 4 were found. The averages range between 1.5 and 2.1 and agree well with the ratio of the target variabilities per group. The larger the repeatability for a given group of samples, the smaller is the weight of the additional variability contributions for reproducibility, such as reference standard preparation and analysis, operator, time, etc. Consequently, the ratio is also smaller, and vice versa. From the ratio, the error contribution of repeatability to the overall variability can be directly calculated as the square of the reciprocal (variance of repeatability/variance of reproducibility). For example, in the case of solutions, the larger ratio may be explained by the simple sample preparation, resulting in a repeatability contribution of only 23 % (using the average ratio of 2.1). As a consequence, the influence of the reference standard and the other variations to the overall variability is increased, directly affecting the reproducibility. In contrast, for more complex samples such as a bath or emulsion, the repeatability dominates, resulting in small ratios. The same is true (to a lesser extent, but supported by more data) for tablets and creams where repeatability and reproducibility have about equal contributions. It should be taken into consideration that the uncertainty of the ratio is larger because it includes the uncertainty of both precision levels. Therefore, for estimating the limit of the distribution and calculating the average of the ratios, the upper 10 % were ignored.

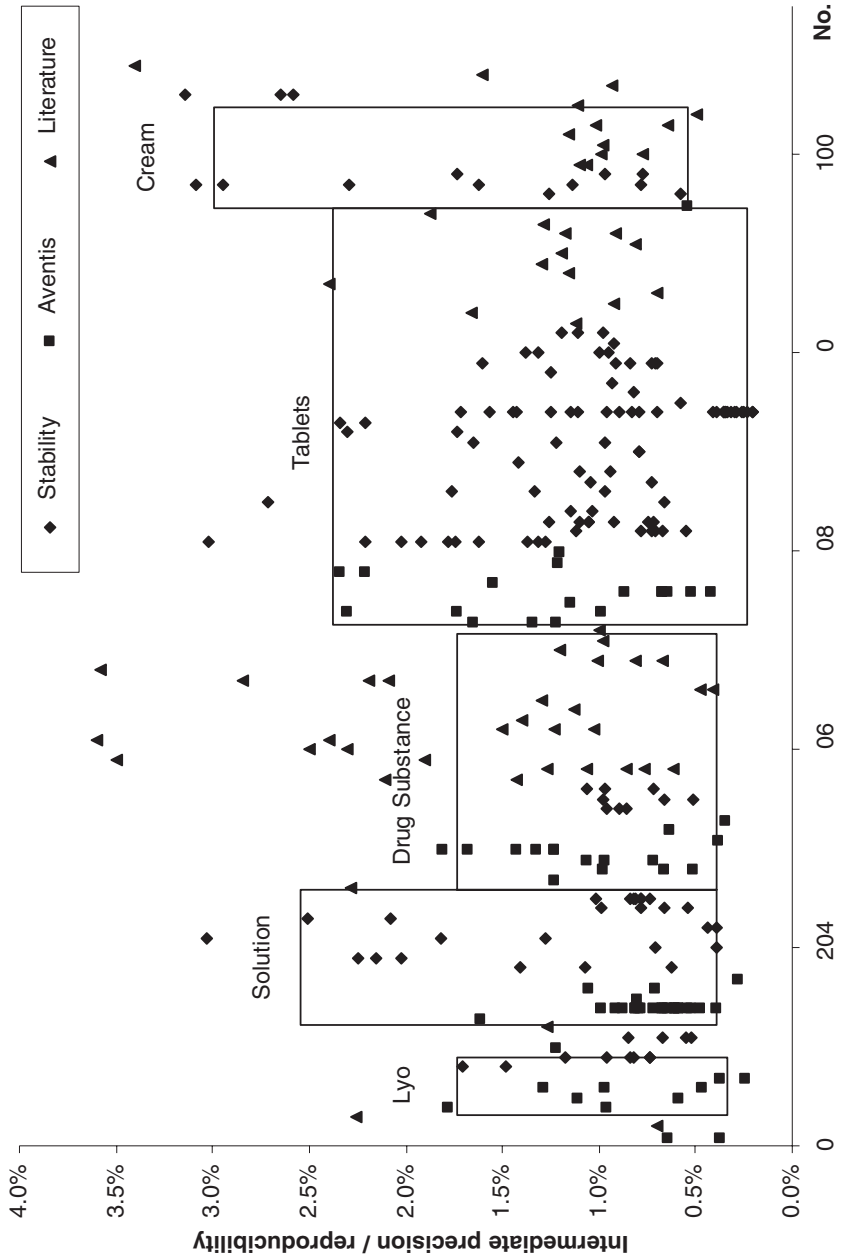


Figure 2.1-12: Intermediate precisions/reproducibilities for LC assay sorted according to the type of sample (drug product). The sequence on the x-axis corresponds to different analytes or sources. The origin of the data is indicated by the symbols and explained in the text. For sample types with sufficient results available (Aventis and Stability origin), the 95% distribution ranges are shown by rectangles.

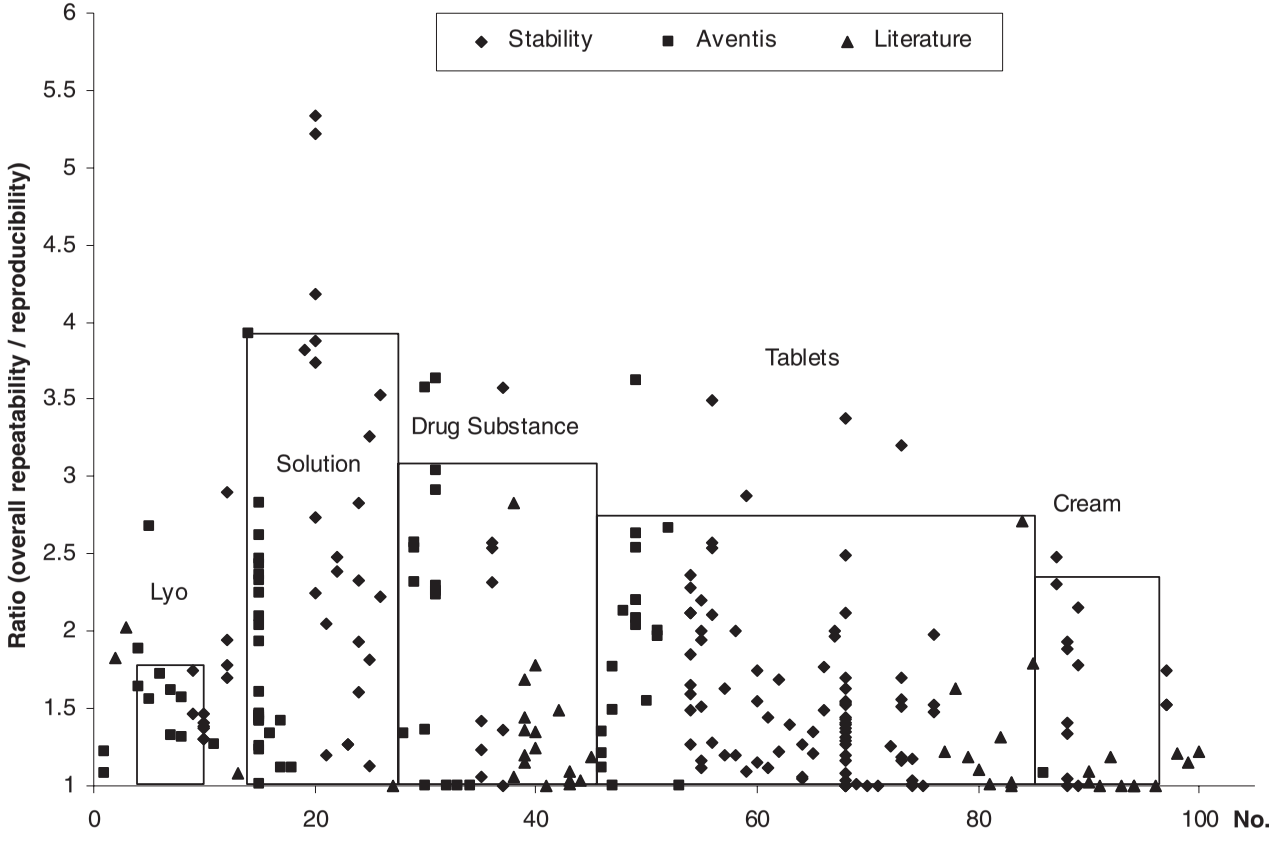


Figure 2.1-13: Ratio between reproducibility and overall repeatability for LC assay sorted according to the type of sample (drug product). The sequence on the x-axis corresponds to different analytes or sources. The origin of the data is indicated by the symbols and explained in the text. For sample types with sufficient results available (Aventis and Stability origin), the upper limit of the 90% distribution range is shown by rectangles.

Table 2.1-5 Averages and ranges for repeatability and reproducibility, originating from Aventis and the DPhG-project (see text for details).

	Repeatability (individual)				Reproducibility				Ratio between reproducibility and overall repeatability		
Sample Type	**Seq.[1]**	**No.[2]**	**Av.[3]**	**Range**	**Seq.[1]**	**No.[2]**	**Av.[3]**	**Range**	**Seq.[1]**	**Av.[3]**	**Range**
Powder	1-4	6 / 1		0.9 %[6]	1-3	2 / 1		0.6 %[6]	1–3		1.2[6]
Lyophilisate	5-11	77 / 7	0.65 %	1.4 %[5]	4-9	17 / 7	1.01 %	1.7 %[5]	4–10	1.5	1.7[7]
Gel	12-14	8 / 2		1.0 %[6]	10–12	5 / 2		1.2 %[6]	11–13		2.9[6]
Solution	15-35	61 / 13	0.54 %	1.2 %[5]	13–26	52 / 13	1.06 %	2.5 %[5]	14–27	2.1	3.9[7]
Drug Substance	36-64	79 / 10	0.55 %	1.2 %[5]	27–51	25 / 10	0.99 %	1.7 %[5]	28–45	1.8	3.0[7]
Syrup[4]	65-66	15 / 2		1.2 %[6]	52	1 / 1		1.0 %[6]			
Tablet	67-134	184 / 34	0.87 %	1.8 %[5]	53–94	108 / 32	1.19 %	2.3 %[5]	46–85	1.5	2.5[7]
Cream	135-144	18 / 3	0.95 %	1.7 %[5]	95–105	12 / 4	1.58 %	3.0 %[5]	86–96	1.5	2.3[7]
Bath	145	24 / 1		3.4 %[6]	106	3 / 1		3.2 %[6]	97		1.8[6]
Emulsion[4]	146-150	9 / 5		4.9 %[6]	107	1 / 1		0.9 %[6]	98		1.2[6]
Chewing gum[4]	151	8 / 2		14.6 %[6]	108–110	4 / 2		12.1 %[6]	99–100		1.2[6]

1: Sequence on the x-axis of Fig. 2.1-11 to 13
2: Number of data / different analytes
3: Average, pooled standard deviation of the 95 % data range or average of the 90 % range in case of the ratio (if sufficient data and analytes were available)
4: From literature only
5: Upper limit of the empirical range, including approx. 95 % of all values, in order to minimize the influence of extreme results
6: Largest experimentally obtained result
7: Upper limit of the empirical range, including approx. 90 % of all values, in order to minimize the influence of extreme results

These results are in agreement with the more general estimation of factors between the precision levels of about 1.5 per level, [39], i.e., a ratio of 2.2 for repeatability and long-term precision.

Concentration dependency
For the results from stability [64], the dependence of the variabilities from both concentration fraction (Horwitz relation) and the amount of analyte injected was investigated. Partly, a linear relationship was found, which is not surprising due to the large number of data (with 250 data, a correlation coefficient of just 0.124 becomes significant, see Section 2.4.1.2, Table 2.4-2). However, the trend was not confirmed by the limits of the distribution of individual variabilities, which were constant over the whole range.

2.1.4
Sources to Obtain and Supplement Precisions

In the previous section, acceptable ranges of precisions were discussed. Under certain conditions (the same analytical technique, sufficiently above the quantitation limit), the analytical variability seems to be mainly dependent on the sample type (preparation). However, due to specific aspects of the given analytical procedure as well as the analyte/sample, it can be assumed that each procedure has its specific target variability within this general range. Whereas the general range is important in order to define acceptance criteria for validation (provided that the minimum requirement from the specification limits is fulfilled), a reliable precision for a specific analytical procedure is essential for purposes *of internal analytical quality assuranc*e, i.e., ensuring the routine quality of the analytical results [8]. For such a life-cycle precision, validation can only provide a beginning. Knowing reliably the analytical variability can help in the detection of atypical results and can facilitate investigations of out-of specification results (see also Chapter 10), etc. Therefore, the basic data obtained in validation should be supplemented during the routine application of the analytical procedure. This does not necessarily mean an additional experimental effort, it only requires a more efficient use of the data produced. There are many sources from which to extract results for the different precision levels [10]. System precision results can be gathered from System Suitability Tests, Equation 2.1-8 can be used to calculated repeatabilities from routine batch analysis (duplicate sample preparations), and if control samples are analysed (control charts, see Chapter 9), long-term reproducibility can be calculated. Experimental studies with repeated analysis of the same samples, such as during method transfer (see Chapter 7) and stability studies, are excellent sources of both repeatability and intermediate precision/reproducibility.

2.1.4.1 **Precisions from Stability**
In stability studies, the same analytical procedure is applied over a long time. Therefore, these data provide very reliable, long-term analytical variability. A prerequisite to calculating the precision is the availability of non-rounded, individual results for

Table 2.1-6: Stability study of a film tablet, stored at 25 °C/60 %RH

Storage interval	Content (mg)		
(months)	Preparation 1	Preparation 2	Preparation 3
0	3.907	3.914	3.908
3	3.954	4.121	4.051
6	3.902	3.945	3.965
9	3.967	3.987	4.083
12	4.020	3.998	3.973
18	3.895	3.930	3.890
Unweighted linear regression			
Slope (± 95 % confidence interval)		–0.00148±0.00574	
Residual standard deviation		1.72 %	
ANOVA [28]			
Intra-serial variance		0.0021595	
Inter-serial variance		0.00259837	
Overall variance		0.00475787	
Overall mean		3.967	
Overall repeatability		1.17 %	
Reproducibility		1.74 %	

each storage interval. If repeated determinations are performed for each storage interval, both overall repeatability (Eq. 2.1-9) and reproducibility, can be calculated. When there are sufficient replicates, individual repeatabilities can be calculated. In order to increase the number of replicates, several presentations or storage temperatures of the same bulk batch can be combined, provided that they do not have any influence on the stability and as long as they were analysed in the same series, using the same reference standard preparations.

Reproducibilities are calculated either by an analysis of variances (one-way ANOVA, see Section 2.1.2.3, Eqs. 2.1-9 to 11), or – in the case of a significant decrease in content – from the residual standard deviation of the linear regression (Eq. 2.4-2) of the individual content determinations (y-values) versus the storage time (x-values). In order to normalise this parameter, it is referred to the content mean (Eq. 2.1-12).

$$\text{Reproducibility from regression:} \quad s_R = \frac{s_y}{\bar{y}} 100\,\% \tag{2.1-12}$$

In the example given in Table 2.1-6 and Figure 2.1-14, the confidence interval of the slope includes zero and is not significant. Therefore, the reproducibility can be calculated by an ANOVA. Comparing this result with the residual standard deviation of the regression, both calculation procedures result in identical reproducibilities. This could also be verified by examination of a large number of data sets [64]. Therefore, it can be concluded that the residual standard deviation of the regression also provides a suitable measure of the analytical variability. However, due to the weighting factor included in the regression and the mean content value, the content decrease should be limited to about 10 %.

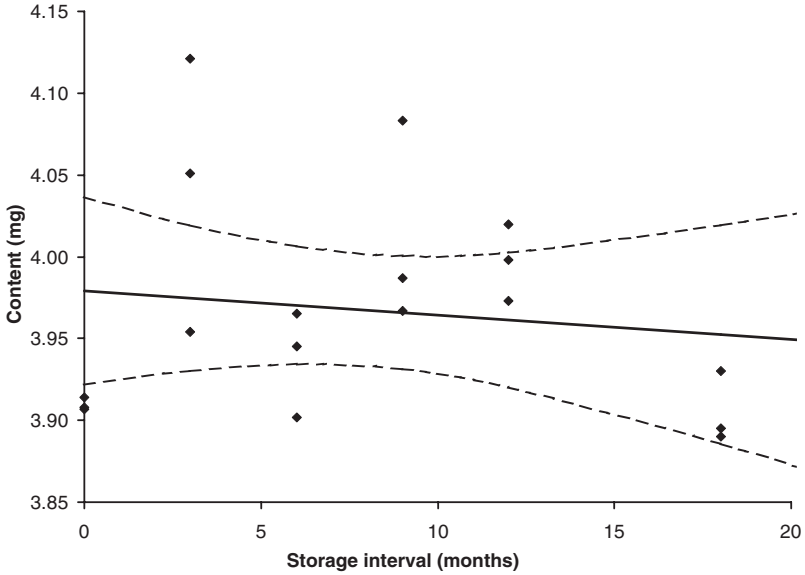


Figure 2.1-14: Stability study of a film tablet batch at 25 °C/60% relative humidity over 18 months. Besides the individual content determinations, the unweighted linear regression line with its 95% confidence intervals is shown.

2.1.5
Key Points

- Be aware of the large variability of experimental standard deviations!
- Do not calculate standard deviations with three values only, the true standard deviation can be up to 4.4 times the calculated result!
- Distinguish clearly between the precision levels, in order to assign the contributions of the various steps or factors of an analysis correctly.
- Repeatability, intermediate precision, and the ratio between the two precision levels are dependent on the type of sample (drug substance, drug product), mainly due to the different influence of the sample and its preparation.
- At low concentrations, the integration/detection variability becomes the dominating error source.

2.2 Specificity

Joachim Ermer

ICH

"Specificity is the ability to assess unequivocally the analyte in the presence of components which may be expected to be present. Typically these might include impurities, degradants, matrix, etc. Lack of specificity of an individual procedure may be compensated by other supporting analytical procedure(s)" [1a].

With respect to identification, discrimination between closely related compounds likely to be present should be demonstrated by positive and negative samples. In the case of chromatographic assay and impurity tests, available impurities/degradants can be spiked at appropriate levels to the corresponding matrix or else degraded samples can be used. For assay, it can be demonstrated that the result is unaffected by the spiked material. Impurities should be separated individually and/or from other matrix components. Specificity can also be demonstrated by verification of the result with an independent analytical procedure.

In the case of chromatographic separation, resolution factors should be obtained for critical separation. Tests for peak homogeneity, for example, by diode array detection (DAD) or mass spectrometry (MS) are recommended.

There has been some controversial discussion about the terminology for this validation characteristic. In contrast to the ICH, most other analytical organisations define this as *selectivity*, whereas specificity is regarded in an absolute sense, as the "*ultimate degree of selectivity*" (IUPAC) [68]. Despite this controversy, there is a broad agreement that specificity/selectivity is the critical basis of each analytical procedure. Without a sufficient selectivity, the other performance parameters are meaningless. In order to maintain a consistent terminology, in the following 'specificity' is used as the general term for the validation characteristic, whereas 'selective' and 'selectivity' describe its qualitative grade. The latter is important to realise, because there is no absolute measure of selectivity, there is only an *absence of evidence, no evidence of absence*.

In contrast to chemical analysis, where each analytical procedure is regarded (and evaluated) separately, in pharmaceutical analysis, a whole range of control tests is used to evaluate a batch. Therefore, the performance of these individual analytical procedures can complement each other in order to achieve the required overall level of selectivity. For example, an assay by means of a less selective titration that will include impurities with the same functional groups, can be confirmed (or corrected) by a selective impurity determination by LC [1b].

Specificity is to be considered from the beginning of the method development, taking into account the properties of both analyte and sample (matrix). The (sufficiently) selective determination of the analyte can be achieved by appropriate sample

preparation, separation, and/or detection. Usually, a combination of several approaches will be developed.

Selective detection

For a *selective detection*, unique properties of the analyte are used, for example, spectral properties (selective UV-wavelength, fluorescence), MS including fragmentation, selective reactions (sensors) or molecular recognition (antibodies, receptors).

An example of a highly selective detection for the determination of the enantiomeric purity of the constitutent amino acids of a synthetic tripeptide is shown in Figure 2.2-1. The hydrolysed tripeptide is derivatised with chiral Marfey's reagent [69], converting the amino acid enantiomers into pairs of diastereomers, which can be separated by RP-chromatography. However, as the upper trace shows, the UV chromatogram is rather complex – even for a tripeptide with only six enantiomers – due to additional peaks related to the reagent or side products. However, since the molecular masses of the derivatised amino acids are known, the respective mass chromatograms can easily be obtained (traces B–D), eliminating any interference from other compounds in the mixture.

Selectivity can also be achieved by means of the *sample preparation*, for example, by derivatisation, extraction, precipitation, adsorption, etc. However, a complex sample preparation will probably have a major influence on other validation characteristics, such as precision (see Section 2.1.3.2) and/or accuracy. Therefore, an overall acceptable balance needs to be found.

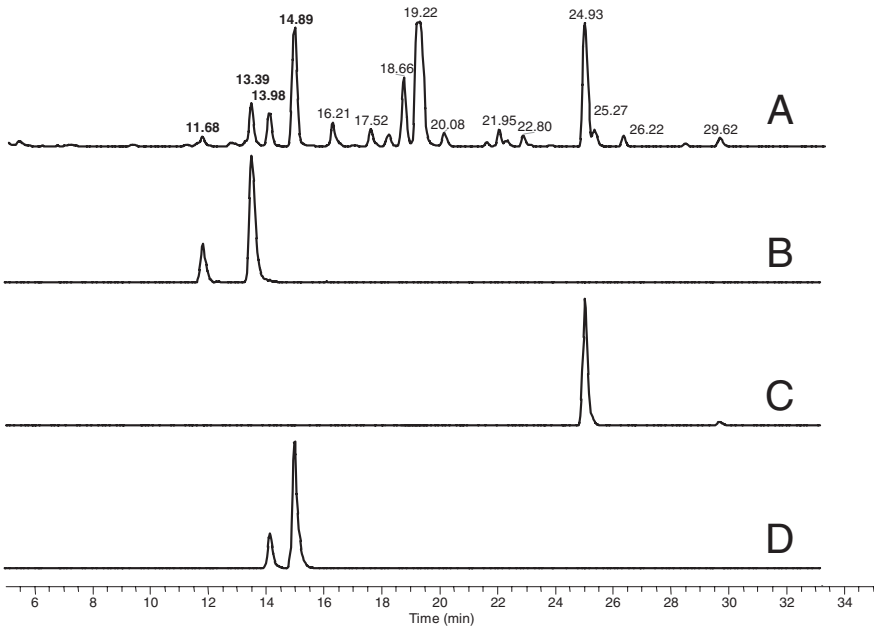


Figure 2.2-1: RP chromatography with UV (A) and mass spectrometric detection (B–D). The smaller peaks in traces B–D belong to the D-amino acids. (Reproduced with permission from [70].)

Stress samples

A moot point is the utilisation of *stress samples* for specificity investigations. As in the whole validation, the analyst should always have the routine application of the analytical procedure in mind, namely, what are the interferences that are likely to occur in practice? Therefore, with respect to degradants, only those that "*may be expected to be present*" [1b] are relevant and need to be taken into account. *"...it may not be necessary to examine specifically for certain degradation products if it has been demonstrated that they are not formed under accelerated or long term storage conditions."* [1g] However, in stress studies, artificial conditions are applied, often resulting to a large extent in degradants that will never be observed during routine storage or usage conditions. Of course, stress studies are essential as part of the stability program to elucidate the degradation pathway, but not (all of the samples) for validation. Some stress samples may be used, provided that the stress conditions are relevant for the prediction of 'routine' degradants, or to demonstrate general separation capability. *"As appropriate, this should include samples stored under relevant stress conditions."* [1b] In order to avoid atypical degradation, it's extent should be restricted to a maximum of 10 %. In addition, the purpose of the stress samples and particularly the evaluation of the results, should be clearly defined in the validation protocol. For example, peak purity investigations of stress samples which are not considered as relevant for routine storage conditions, should not be performed, because they do not provide added value with respect to the suitability of the (routine) procedure. However, such samples may be used to demonstrate the general capability of the method to separate a (larger) number of substances. For more detailed investigations, samples from accelerated storage conditions are preferable.

Clearing validation

For validation of *cleaning methods* (see also Section 2.3.4), it is most important to take interferences from the sampling procedure into account. This should include blank extractions of the swab material, as well as blank swabs from the respective surface. It must also be considered that the target substance may be altered during the cleaning process so that the analyte may be a degradant. Due to the small concentrations involved, peak purity investigations are difficult to perform, and are not normally essential. Therefore, specificity is usually demonstrated by sufficient chromatographic resolution, or lack of interference.

Approaches

Basically, specificity can be addressed directly or indirectly. The latter approach demonstrates acceptable accuracy for the results of the analytical procedure (see Section 2.2.1). The direct approaches demonstrate the lack of (or an acceptable) interference by other substances, for example, by obtaining the same result with and without such potentially interfering compounds (with respect to an acceptable difference see Section 2.3.5), sufficient chromatographic resolution (see Section 2.2.2), or peak purity (see Section 2.2.3).

2.2.1 Demonstration of Specificity by Accuracy

As an indirect approach, sufficient specificity can be concluded if an acceptable accuracy is demonstrated. If all components of the sample can be determined quantitatively, the overall accuracy can be verified by means of a mass balance, i.e., summing up all determined substances. With respect to the evaluation, i.e. an acceptable difference, the problem of error propagation needs to be considered (see Section 2.3.5).

The other possibility is to compare the results of the analytical procedure in question to another procedure (see Section 2.3.1).

2.2.2 Chromatographic Resolution

Chromatographic separation is usually quantified by resolution factors, according to EP (Eq. 2.2-1) or USP (Eq. 2.2-2) at half height or at the baseline, respectively. However, these equations only provide meaningful results for baseline-separated peaks. The USP resolution factor is less sensitive towards tailing, but is more complex to determine.

Resolution factors:

$$\text{EP:} \quad R_s = \frac{1.18(t_{Rb} - t_{Ra})}{w_{0.5a} + w_{0.5b}} \tag{2.2-1}$$

$$\text{USP:} \quad R_s = \frac{2(t_{Rb} - t_{Ra})}{w_a + w_b} \tag{2.2-2}$$

$t_{Ra,b}$: Retention time of peaks a and b with $t_{Rb} > t_{Ra}$
$w_{0.5a,b}$: Peak width a and b at half height
$w_{a,b}$: Peak width a and b at baseline.

Resolution factors are difficult to compare between methods, because they are defined for Gaussian peaks and are dependent on the tailing. A modified equation has been proposed in the event of tailing [71].

Peak-to-valley ratio

In the case of incomplete separation, the calculations according to Eqs. (2.2-1) and (2.2-2) are not possible or are biased due to the additivity of the peak curves, especially for peaks of different magnitude. Here, other separation parameters such as the peak-to-valley ratio (p/v) are more appropriate. This approach measures the height above the extrapolated baseline at the lowest point of the curve between the peaks (i.e. the 'valley') with respect to the height of the minor (impurity) peak (Fig. 2.2-2). Therefore, it is directly related to the peak integration and independent of tailing or 'smearing' effects in the elution range behind the main peak [15, 72]. However, care should be taken to define the accurate mode of integration, i.e., drop or rider integration (see Section 2.3.3).

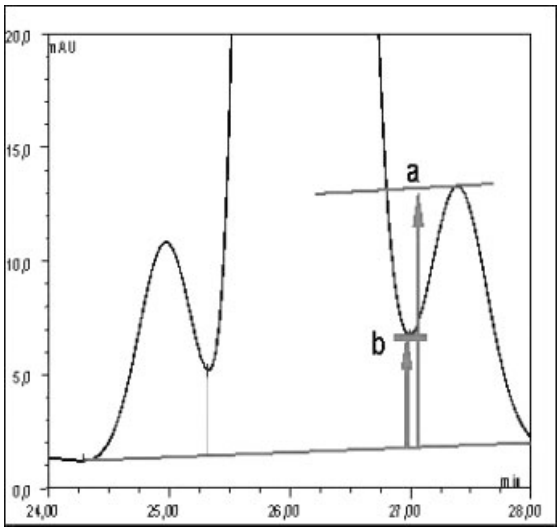


Figure 2.2-2: Peak separation indices for minor peaks [73]: Peak-to-valley ratio = 1-*b/a*
By subtracting the ratio *b/a* from 1, the parameter is easier to compare with the resolution factor *Rs*. A peak-to-valley ratio of 1 corresponds to baseline separation, a value of 0 to unresolved peaks. (Reproduced with permission from [74].)

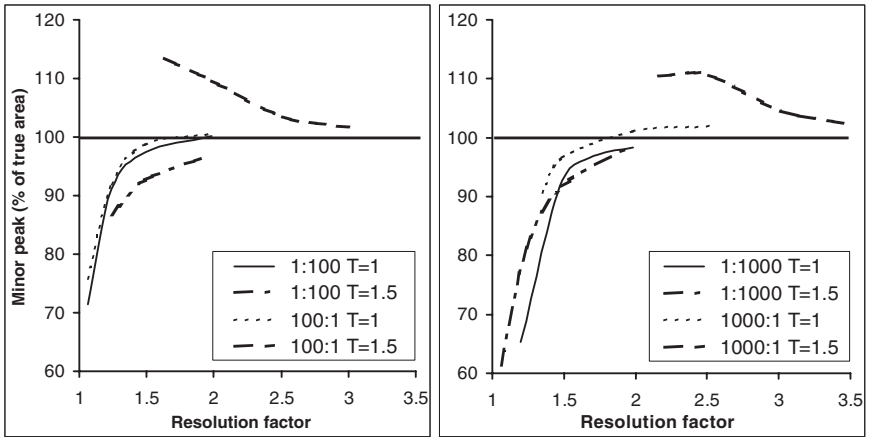


Figure 2.2-3: Resolution and (drop) integration accuracy with respect to the minor peak in dependence on peak size, elution order, and tailing (data from [75]). The numbers in the legend represent the ratio of the peak (area), their sequence corresponds to the elution order, and *T* indicates the tailing factor. For example, the solid line in the left diagram describes the accuracy of the integrated area of a 1 % impurity peak, eluting before the main peak with tailing factors of 1.

Resolution requirements

The *resolution requirements* are strongly dependent on the size difference, the elution order, and the tailing of the peaks involved [72, 75] (Fig. 2.2-3). Generally, larger resolution factors are required for elution of the minor peak after the main peak, and larger size differences and tailing to ensure satisfactory separation and quantification. If the factors are not sufficient for an accurate integration, then minor peaks eluting before the main peak and symmetric peaks, irrespective of the elution order, are underestimated, whereas tailed peaks eluting after the main peak will be overestimated. As a conclusion, if separation factors are determined, the *typical concentration levels* of the impurities or those at the specification limit (as the worst case) should be used.

2.2.3 Peak Purity (Co-elution)

There are many approaches to the investigation of co-elution, also called peak purity or peak homogeneity. However, only co-elution which results in interference in the detection mode of the analytical procedure should be taken into account. Approaches may include variations in the chromatographic conditions, peak shape analysis, re-chromatography of peak (fraction)s, DAD, MS, etc. The reader should be aware that only the absence of co-elution evidence is possible, but never the proof of peak homogeneity. However, applying several of the aforementioned approaches, preferably in combination, will greatly increase the confidence in the method. If other detection modes are applied, such as different wavelengths in DAD or MS, identified co-eluting substances must be further investigated to determine their relevance under routine conditions. For some of the approaches, such as a variation in the chromatographic conditions, the relation to the method development or robustness studies is obvious. For example, chromatograms obtained by varying the pH, modifier composition, temperature, etc. (see Section 2.7) can be inspected for new peaks (see Section 2.2.3.1) or else the peaks can be investigated by DAD (see Section 2.2.3.2) or MS (see Section 2.2.3.3).

2.2.3.1 Peak Shape Analysis

A very simple and straightforward, but nevertheless very efficient method, is the visual investigation of irregularities in the peak shape, i.e., shoulders and peak asymmetry. However, sometimes at low concentrations it is quite difficult to distinguish the former from 'smearing' effects at the foot of larger peaks, and the latter from tailing. These visual inspections can be assisted by mathematical evaluations: The *1st derivative* of the signal results in symmetrical curves for Gaussian peaks. Co-elution will decrease the height of the maximum or minimum, depending on whether the retention time of the co-eluting peak is smaller or larger than that of the major peak (Fig. 2.2-4a). However, the problem is that tailing peaks also produce asymmetric first-derivative curves, without any co-elution. In such cases, co-elution is indicated by irregularities or shoulders (Fig. 2.2-4b). If the first derivative cannot be provided by the chromatographic data system, the chromatogram can be exported as ASCII-format, imported into EXCEL and the ratio of differences $(y_{n+1}\text{-}y_n)/(x_{n+1}\text{-}x_n)$ plotted

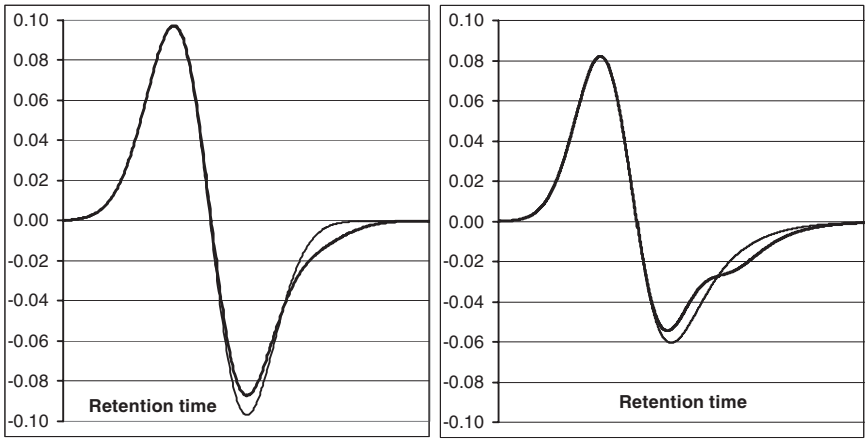


Figure 2.2-4: First derivative chromatograms of a symmetric (Gaussian, a) and an asymmetric peak (tailing factor = 1.33, b). The y-axes correspond to the first derivation of the signal. The thinner lines represent the derivation of the major peak alone, the thick lines those of the co-eluted peaks. The peak ratio of a major to minor peak is 10:1, the resolution factor is 0.5, and the minor peak is eluting after the main peak.

vs. x (where x= time and y= absorbance). The noise in the signal is considerably increased in derivative chromatograms, and therefore the sensitivity is rather poor (see Table 2.2-1).

For Gaussian peaks, the asymmetry (according to USP) is unity, independent of the peak height. In the case of co-eluting small impurities, the upper peak heights of the main peak are not affected, but only the lower ones. Consequently, the asymmetry only deviates from unity for lower peak heights. For tailing peaks, the asymmetry decreases continuously with the height, and co-elution will result in sigmoid asymmetry curves in the affected range of the peak height (Fig. 2.2-5b).

In Table 2.2-1, the various peak shape investigations are compared. It is obvious that the asymmetry approach has only marginal advantages over the visual inspection for smaller impurities (< 1 %). Therefore, it is only sensible to use it if the chromatographic data system provides an easily accessible height-dependent calculation of the asymmetry. For larger impurities (>10 %), the first derivative is a suitable option.

Table 2.2-1 Detection of co-elution for tailing peaks in dependence on their size difference. The tailing factor according to USP was 1.33.

Peak ratio	Minimum resolution factor for detection of co-elution		
	Visual	First derivation	Asymmetry curve
100 : 10	~ 0.7	~ 0.4	~ 0.5
100 : 1	~ 1.3	Not detectable (noise)	~ 1.1
100 : 0.5	~ 1.5	Not detectable (noise)	~ 1.3

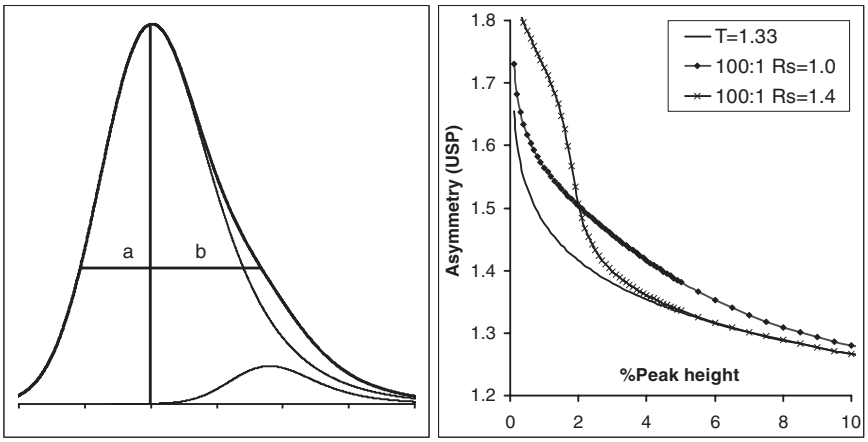


Figure 2.2-5: Peak asymmetry ($(a+b)/2a$, according to USP, A) and its dependence from the height (B) for various degrees of co-elution in the case of peak tailing (tailing factor 1.33) and a height ratio of 100:1. For smaller and larger size differences, the asymmetries at larger and smaller peak heights will be affected, respectively.

2.2.3.2 **Re-chromatography**

Re-chromatography of suspected peaks represents a simple, universally available and sensitive approach for small amounts of co-eluting impurities. Confidence in the results is the greater the more the two applied chromatographic methods differ. Various combinations can be taken into consideration, from variation of method conditions (different eluent, buffer, pH, column) to a change in the methodology such as size exclusion chromatography, ion chromatography, thin layer chromatography (TLC), capillary electrophoresis (CE) [76], gas chromatography, etc. The investigations can be performed off-line with isolated peaks or peak fractions (see Fig. 2.2-6), or as a direct orthogonal coupling of the two methods. Working with isolated peak fractions, care should be taken to avoid artefacts due to degradation.

2.2.3.3 **Diode Array Detection**

The spectral peak homogeneity can be investigated by means of diode array or scanning detectors [77], provided that there is a difference both in the spectra and in the retention time of the co-eluting substances. If this is fulfilled, detection of co-elution with commercially available software is easily achieved, provided that the concentration difference is not too large (Fig. 2.2-7 c). However, impurities below 1 % are usually difficult to detect (Fig. 2.2-7 d).

2.2.3.4 **LC-MS**

The most powerful technique for the investigation of the peak purity is mass spectrometry [70, 78]. Mass spectra are taken successively over the whole elution range of the suspected peak (Fig. 2.2-8, UV trace). If during this spectra 'scrolling' additional masses are detected such as can be seen in the insert a, the corresponding

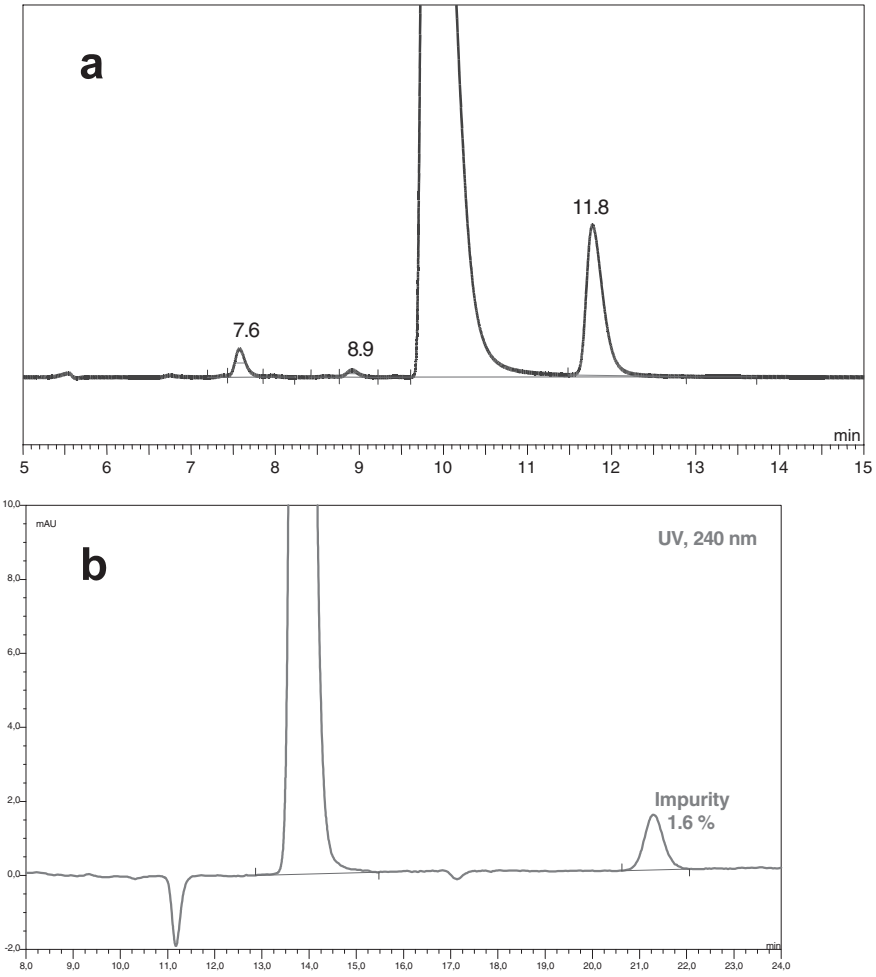


Figure 2.2-6: Investigation of chromatographic peak purity by means of re-chromatography. (a) Chromatogram of the method to be validated (acetonitrile/water/0.1 % trifluoroacetic acid, RP C8 column). The main peak was heart-cut from 9.6 to 11.0 minutes. (b) Re-chromatography of the main peak fraction using another method (acetonitrile / 0.2M sodium phosphate buffer pH 4.0, RP C8-column). (Reproduced with permission from [74].)

mass chromatogram is extracted (Fig. 2.2-8, lower trace). Differences in the retention time and/or elution behaviour with respect to the UV peak are proof of a co-eluting impurity. In the given example, 0.5 % of the impurity was present. Of course, the detection limit depends on the individual MS response of the impurities, and diastereomers cannot be detected. If LC procedures with non-volatile buffers are to be validated, the corresponding peak fractions can be isolated and re-chromatographed under MS-compatible conditions. Any co-eluting substances identified must be further investigated for their relevance under the control test conditions.

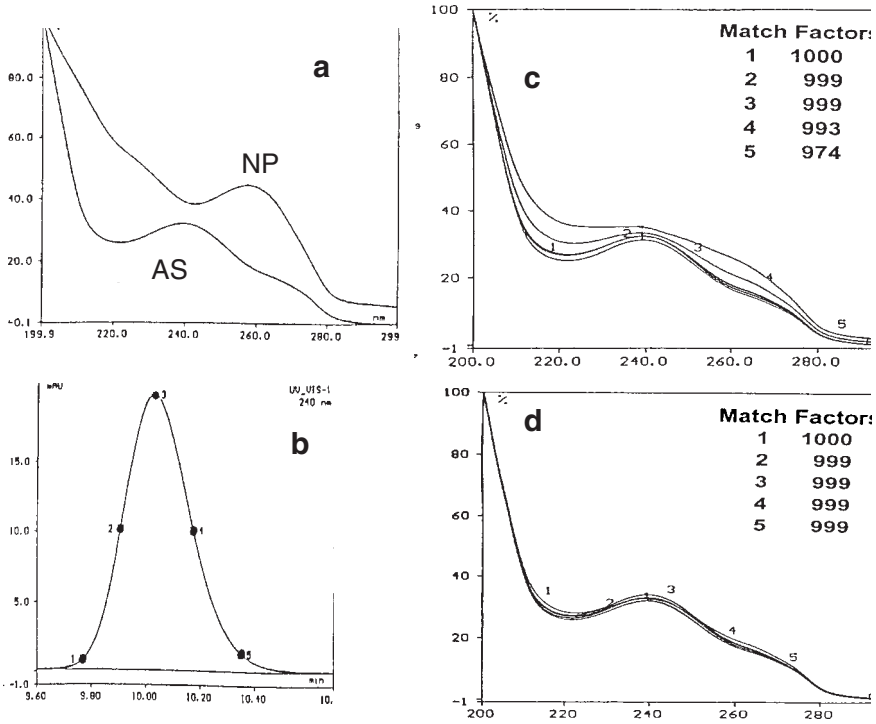


Figure 2.2-7: Investigation of chromatographic peak purity by means of diode array detection
(a) Spectra of drug substance (AS) and impurity (NP).
(b) Spectra were extracted in the peak maximum (3) and at approx. 5 % and 50 % peak height, each at the leading (1,2) and the tailing edge (4,5). The spectra were normalised with respect to spectrum 1 (match factor 1000).
(c) Co-elution of a mixture containing about 10 % impurity.
(d) Co-elution of a mixture containing about 0.5 % impurity.
(Reproduced with permission from [74].)

For example, an impurity producing a large MS peak may only be present in a very small and negligible mass concentration.

Although not often applied in routine (pharmaceutical) analysis, MS detection offers tremendous gains in efficiency and reliability of the procedures, due to the highly specific detection (largely) without interferences, for monitoring of impurity profiles and identification [70].

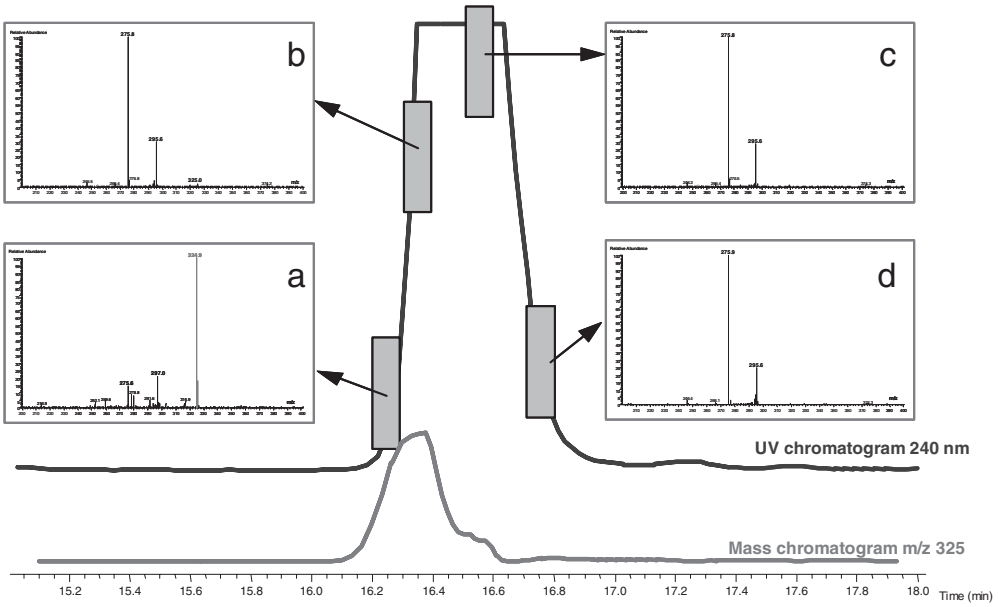


Figure 2.2-8: Investigation of chromatographic peak purity by means of LC-MS The upper and lower chromatogram display the UV (at 240 nm) and the extracted ion chromatogram (for m/z 325), respectively. The inserts (a) to (d) are representative mass spectra over the investigated (UV) peak. The amount of co-eluting impurity corresponds to 0.5 %. (Reproduced with permission from [74].)

2.2.4

Key Points

- Selectivity is the 'hallmark' of any analytical procedure.
- Apply scientific judgment to a selection of *relevant* substances and samples.
- Resolution requirements are concentration dependent, use relevant impurity levels.
- Peak purity investigations should be integrated into method development/robustness. Only the absence of evidence (for co-elution), no evidence of absence is possible! A combination of several approaches will considerably increase overall confidence. Co-elution identified under different detection conditions must be further investigated for relevance in routine applications.
  - Variation in chromatographic conditions and/or re-chromatography is a simple, sensitive approach.
  - Peak shape investigations and DAD are difficult for small co-eluting substances.
  - MS is very sensitive and highly selective, but is also dependent on substance properties.

2.3 Accuracy

Joachim Ermer

ICH

"*The accuracy of an analytical procedure expresses the closeness of agreement between the value which is accepted either as a conventional true value or an accepted reference value and the value found*". [1a]

Accuracy can be demonstrated by the following approaches:

- Inferred from precision, linearity and specificity
- Comparison of the results with those of a well characterised, independent procedure
- Application to a reference material (for drug substance)
- Recovery of drug substance spiked to placebo or drug product (for drug product)
- Recovery of the impurity spiked to drug substance or drug product (for impurities)

For the quantitative approaches, at least nine determinations across the specified range should be obtained, for example, three replicates at three concentration levels each. The percentage recovery or the difference between the mean and the accepted true value together with the confidence intervals are recommended.

It is important to use the *same quantitation method (calibration model)* in the accuracy studies as used in the control test procedure. Sometimes in the literature, the data from linearity studies are simply used to calculate the content of spiked samples. However, the validation linearity study is usually not identical to the calibration applied in routine analysis. Again, validation has to demonstrate the suitability of the *routine* analytical procedure. Deviations from the theoretical recovery values, while performing a calibration with a drug substance alone, may indicate interferences between the analyte and placebo components, incomplete extraction, etc. In such a case, the calibration should be done with a synthetic mixture of placebo and drug substance standard. Such interferences will also be detected by comparing the linearities of diluted drug substance and of spiked placebo, but the evaluation is more complex (for example the influence of extrapolation on the intercept, see Section 2.4.1.4). In contrast, recovery studies usually concentrate directly on the working range and are simpler (but not always easy) to evaluate.

2.3.1 Drug Substance

It can be rather difficult to demonstrate accuracy for a drug substance appropriately, especially if no (independently) characterised reference standard is available. Other, independent analytical procedures are often not readily found. Nevertheless, every effort should be made to identify a suitable method for comparison, because this is the only way to verify accuracy objectively. Instead of quantitative comparison, the results could also be supported by another method, for example, the verification of a very high purity of a drug substance by differential scanning calorimetry (DSC) [79]. Inferring accuracy from the other validation characteristics should be the 'last resort', because it does not provide absolute measures. The only exception is if the analytical procedure to be validated is based on an absolute method itself (see below), or permits a universal calibration, i.e., a calibration with another, well characterised substance, such as LC with refractive index [80] or nitrogen detection [80, 81].

Sometimes in validation literature, recovery is reported for a drug substance. However, 'recovery' from simple solutions does not provide meaningful information (at least if all determinations are traced back to a reference standard characterised with the same analytical procedure) and is therefore not appropriate to demonstrate accuracy.

What can be considered as a well-characterised, independent procedure?

Preferably, it should provide an *absolute* measure of the analyte, such as titration, nitrogen (or other constituent elements) determination (whole sample, or coupled to LC [81]), NMR [55, 56]), or indirectly by specific reactions of the analyte (e.g., enzymatic assays). For such absolute methods, according to the well-known fundamentals (such as defined stoichiometry, composition, reaction mechanism), their accuracy can be assumed. If no absolute method is available, sufficient agreement between the results of two independent analytical procedures may be used to conclude accuracy (although with less confidence compared to the absolute approaches).

Mean value or two-sample *t*-test: significant difference if

$$\frac{|\bar{x}_1 - \bar{x}_2|}{s_{av}} \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \quad > \quad t\,(P, n_1 + n_2 - 2) \text{ with}$$

$$s_{av} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} \tag{2.3-1}$$

Prerequisite of equal variances: $\frac{s_1^2}{s_2^2} < F(\,P, n_1 - 1,\ n_2 - 1)$ with $s_1^2 > s_2^2$ (*F*-test)

$F(P,df_1,df_2)$ = Fisher's *F*-value for the statistical confidence *P* and the degrees of freedom *df* corresponding to s_1 and s_2. Excel®: $F = FINV(\alpha, df_1, df_2)$; $\alpha = 1\text{-}P$>

Nominal value *t*-test: $\frac{|\bar{x}_1 - x_{reference}|}{s_1} \sqrt{n_1} \rangle\ t(P, n_1 - 1)$ (2.3-2)

Paired sample *t*-test: $\frac{\bar{d}}{s_d} \sqrt{n_d} \rangle\ t(P, n_d - 1)$ (2.3-3)

It is obvious from the (far from exhaustive) list that we have to acknowledge compromises with respect to specificity and precision. For example, in titration we can expect quite a high precision, but impurities with the same functional groups will also respond. Usually, these absolute methods are less selective compared to chromatographic separations. However, the objective is to verify accuracy of the procedure to be validated, not to demonstrate that the two methods are identical. Therefore, statistical significance tests should be used with caution (if at all, as acceptance criterion, see Section 2.3.5). The difference in specificity will most likely lead to a systematic influence on the results. If the effect can be quantified, the results should be corrected before performing the statistical comparison. If a correction is not possible, the presumptions of the significance test are violated and the *t*-test should consequently not be performed.

t-tests

These t-tests investigate whether a difference between two means (mean *t*-test, two-sample *t*-tests, Eq. 2.3-1), between a mean and a reference or target value (nominal value *t*-test, Eq. 2.3-2), or between replicated determinations of samples by both methods (paired *t*-test, Eq. 2.3-3) becomes significant. However, whether a significant difference is of practical importance is not included in the test (Section 1.4.1, Table 2.3-1). The t statistics calculated according to Eqs (2.3-1 to 2.3-3) is then compared to the critical Student-*t*-values, which are dependent on the statistical level of confidence P (or the error probability α=1-*P*, i.e., a 95 % confidence level corresponds to an error probability of 0.05), and the number of determinations (degrees of freedom, df). These values can be found tabulated in statistical textbooks, or are available in Excel®: *t*=TINV(α, df). If the calculated *t* is larger than the critical one, the difference is significant. Another way to present the test result is the calculation of the *p*-value, i.e., the *probability of error* in accepting the observed result as valid. A value of $p < 0.05$ means a significant difference at 95 % statistical confidence (Excel®: *p*=TDIST(*t*, df, 2), for a two-tailed distribution). As a more visual description of the *t*-test, it is checked, whether the confidence intervals of the means overlap each other, or the nominal value (see Fig. 1-4), i.e., if the true difference is zero.

Equivalence tests

A statistical alternative consists of the so-called *equivalence tests* (Eqs. 2.3-4 to 2.3-6) [28]. Here, the user can define an acceptable difference, i.e., a measure of the practical relevance (see Section 1.4.2). Equivalence can be assumed, if the lower and upper limit of the equivalence interval ($C_{L,U}$) are within the defined acceptance interval ($-\delta \leq C_L \wedge C_U \leq \delta$). Technically, this corresponds to performing two one-sided t-tests.

Equivalence tests [28]: Equivalence can be assumed if $-\delta \leq C_L \wedge C_U \leq \delta$

For mean results: $C_L = 100\left\{\left(\frac{\bar{x}_1}{\bar{x}_2}\right)\exp[-t(P, n_1 + n_2 - 2)\, s] - 1\right\}$

$$C_U = 100\left\{\left(\frac{\bar{x}_1}{\bar{x}_2}\right)\exp[t(P, n_1 + n_2 - 2)\, s] - 1\right\} \qquad (2.3\text{-}4)$$

with $s = s_{av}\sqrt{\left(\frac{1}{n_1 * \overline{x_1}^2} + \frac{1}{n_2 * \overline{x_2}^2}\right)}$

$\delta\%$ = acceptable difference (percentage)
For nominal or target values:

$$C_L = T - \bar{x}_1 - t(P, n_1 - 1)\frac{s_1}{\sqrt{n_1}},\ C_U = T - \bar{x}_1 + t(P, n_1 - 1)\frac{s_1}{\sqrt{n_1}} \qquad (2.3\text{-}5)$$

T = nominal value
δ = acceptable difference (absolute value)

For paired samples:

$$C_L = -\bar{d} - t(P, n - 1)\frac{s_d}{\sqrt{n}},\ C_U = \bar{d} + t(P, n - 1)\frac{s_d}{\sqrt{n}} \qquad (2.3\text{-}6)$$

δ = acceptable difference (absolute value)
$\bar{x}_{1,2}, s_{1,2}, n_{1,2}$ = Mean, standard deviation, and number of determinations of series 1 and 2.
$\bar{d}, s_d, n_d$ = Mean and standard deviation of the differences between, and the number of the pairs of samples.
$t(P,df)$ = Student t-value for the statistical confidence P and the degrees of freedom. Excel®: $t = TINV(\alpha, df)$; $\alpha = 1\text{-}P$ (Note that in case of equivalence tests, α must be multiplied by two, in order to correspond to a one-tailed distribution, i. e., for 95% confidence, α is 0.10)

Another alternative is the simple evaluation of whether the absolute magnitude of the difference is below an acceptable value (for example, below 2.0 %). To define the acceptance criteria, the performance characteristics of the reference procedure, particularly its precision, should also be taken into consideration.

In Table 2.3-1, example A, the nominal value t-test results in a highly significant difference, although it amounts to less than 0.1 %. The reason is that the high number of determinations cause very small confidence intervals. With 23 determinations, the p-value would be larger than 0.05 and the difference would not become significant (at a confidence level of 95 %, assuming the same standard deviation). This (practically absurd) problem of high reliability is converted into the opposite in case of the equivalence test. The equivalence interval becomes very tight with a range from 0.03 to 0.15 % and would be compatible with very narrow acceptance limits. This situation is also illustrated in Figure 1-4, scenario 1.

Comparing two analytical procedures in example B, results in a difference of 0.57 %, which is not significant. However, the equivalence interval is rather wide being between 1.25 and 2.42 %. This is caused by the larger variability of the CE-method, which needs to be considered when establishing the acceptance limits (see Section 2.3.5).

Table 2.3-1 Comparison between two independent procedures or to a reference.

Analyte	A Benzoic acid [56] vs. NIST certificate (99.99%)	B Amoxicillin [51]		C Insulin	
Procedures	^{1}H-NMR	CE	LC	LC	N[b]
Number of determinations	48	12	6	10	9
Mean	99.9%	74.39%	73.97%	92.21%	93.03%
95% Confidence limit:					
lower	99.84%	73.46%	73.20%	91.53%	92.67%
upper	99.96%	75.32%	74.74%	92.88%	93.39%
Relative standard deviation	0.21%	1.96%	0.99%	1.02%	0.50%
Difference (relative)	0.09%	0.57%		0.89%	
t-test[a]	$p = 0.005$	$p = 0.52$		$p = 0.03$	
Equivalence interval (95% confidence)	0.04 to 0.16%	–2.07 to –0.94%		0.23 to 1.54%	

a) The test result is given as the *p*-value
b) Nitrogen determination according to Dumas

Example C shows both small variabilities and a difference, but a significant one. However, from a practical point of view, a difference of less than 1 % is certainly acceptable with respect to the very different methods used. Note that in the case of nitrogen determination, the results were already corrected for the amount of impurities, obtained by LC, assuming the same nitrogen content as for the active substance.

Usually, the concentration range of an available drug substance is very limited and (dedicated) investigations over a larger range are not possible (at least at the upper end, the maximum true content is 100 %). A variation in the concentration of the test solution is also not meaningful for comparison to another procedure, because the conditions for the two procedures (sample preparation, absolute concentrations, etc.) are likely to be different. Therefore, in this case, at least six determinations of an authentic sample should be performed.

2.3.2 Drug Product

Usually, the accuracy is validated by analysing a synthetic mixture of the drug product components, which contain known amounts of a drug substance, also termed spiking or reconstituted drug product. The experimentally obtained amount of active substance is then compared to the true, added amount (recovery). It can be calculated either at each level separately as a percentage recovery, or as a linear regression of the found analyte versus the added one (recovery function). Sometimes, the term 'recovery' is misused when reporting the content of active in drug product batches. This is misleading, because in such cases, the true amount of active is influenced by the manufacturing variability and is not exactly known. Preferably, the result should be termed '% label claim'.

The analyst should be aware of two important aspects with respect to recovery. First, it is based on the (validated) accuracy of the drug substance procedure, otherwise the added amount will already be wrong. Secondly, in preparing the reconstituted drug product, the analyst deviates (more or less) from the routine analytical procedure. Of course, there is no other possibility of adding exactly known amounts, but consideration should be given to the *possible implications*. If, for example, solutions of the placebo are spiked with a stock solution of the active substance, the influence of the missed sample preparation steps, such as grinding, extracting, etc. on the analysis should be considered. Here, information obtained during method development is helpful (for example, homogeneity or extraction investigations). If such steps are of importance, any problems related to them will not influence the experimental recovery, and therefore are not identified. Spiking is also not appropriate, when the properties of the authentic sample are important for analytical measurement, such as in quantitative NIR [82].

2.3.2.1 **Percentage Recovery**

The author recommends applying the percentage recovery calculation, because it gives *easily interpretable* results, at least for narrow working ranges (see Chapter 3, Tables 3-7 and 3-8). The mean recovery can be tested statistically versus the theoretical value of 100 %, i.e., if the 95 % confidence intervals include the theoretical value, with the known problems of statistical significance tests (see Section 1.4.2 and

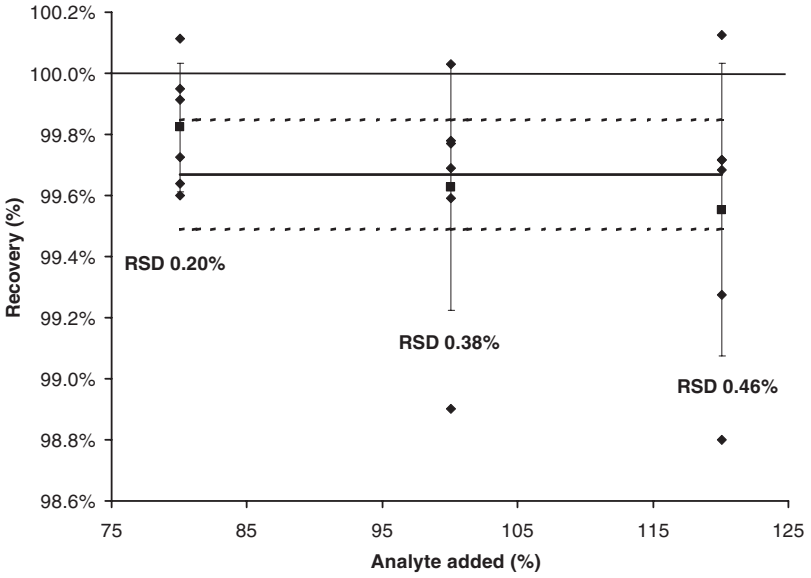


Figure 2.3-1: Recovery investigations for a lyophilised drug product. The diamonds represent the six individual spikings at 80, 100, and 120 % each (of the nominal concentration level). The relative standard deviations of the recoveries for each level are given. The mean and the 95 % confidence intervals for each level are shown as squares and bars, respectively. The overall mean, with 95 % confidence interval, is indicated by horizontal, solid and dotted lines, respectively.

2.3.1). They can be expected especially if the variability of the spiked preparation is the same or even lower than that of the standard preparation and the number of determinations is high. This is illustrated in Figure 2.3-1. Although the 95 % confidence intervals at each level include the theoretical value, the overall interval does not. The small deviation of 0.33 % and the overall relative standard deviation of 0.36 % are certainly acceptable, from a practical point of view. Because the same standards were used for all concentration levels, the small bias can be explained by the variability of the standard preparation.

Alternatively, an equivalence test can be applied. For the purpose of recovery investigations, the nominal value T in Eq. (2.3-5) is 100. The equivalence interval for the mean recovery in the example ranges from 0.18 to 0.48 %, in the case study (Tables 3-7 and 3-8) it ranges from –1.02 to –0.38 % and –3.18 to 0.03 % for the main component and the degradation product, respectively.

The analyst can also establish absolute acceptable limits for the deviation from the theoretical recovery. If these limits apply to the mean recovery, they should be smaller than those for the equivalence test, because the variability of the individual recoveries is reduced for the mean (see Section 2.1.1.3 and Fig. 1-4). Therefore, the scattering of individual recoveries or their standard deviation should additionally be limited. Recovery values should always be plotted, in order to detect trends or concentration dependency (see Fig. 3-6 and 3-7).

2.3.2.2 **Recovery Function**

The recovery function of an unbiased analysis has a slope and an intercept of one and zero, respectively. The experimental results can be tested statistically versus the theoretical values by their 95 % confidence intervals (Eqs. 2.4-5 and 2.4-9). Here we may face the same problem of statistical significance vs. practical relevance as discussed before, although by the process of spiking and sample preparation and the dominating effect of the larger concentrations, enough variability is often present.

Alternatively, equivalence tests can be applied to test the slope and intercept for an acceptable deviation from the theoretical values (Eqs. 2.3-7 and 2.4-14). The limits of the equivalence interval are compared to a previously defined acceptable deviation (see 1.4.2). It is obvious from the equations (see also Fig. 1-4) that the variability of the experimental results (here as residual standard deviation of the regression line s_y) is included in the test and must be taken into consideration during the establishment of acceptance limits.

Equivalence test for slope of one [28] (for explanation of variables, see 2.4.1.1):

$$C_L = b - 1 - t(P, n-2)\frac{s_y}{\sqrt{Q_{xx}}},\ C_U = b - 1 + t(P, n-2)\frac{s_y}{\sqrt{Q_{xx}}} \qquad (2.3\text{-}7)$$

The slope and intercept can also be compared to absolute acceptance limits, as proposed for volumetric titrations (see Section 8.4.1). The variability of the experimental results should be limited in this case by a separate acceptance test.

The evaluation of the intercept may pose a more serious problem. For assay, recovery is usually investigated within the working range from 80 to 120 % [1b].

This results in a large extrapolation and consequently in a higher uncertainty of the intercept (see Section 2.4.1.4 and Fig. 2.4-8).

Due to the different weighting effects, percentage recovery and recovery function may lead to different statistical results. This will be more pronounced, the larger the concentration range.

2.3.2.3 Standard Addition

If no adequate placebo can be prepared, a known amount of drug substance can also be added to an authentic batch of drug product (standard addition). Of course, in this case, only the range above the nominal content is accessible. In order to provide practically relevant information, the upper limit of the investigated spiking range should not be too high, i.e., not greater than 150 %. Because the precision is concentration dependent, the percentage recovery calculation should be based on the *overall amount* of active present, i.e., the theoretical amount is the sum of the original content in the batch, and the spiked amount. This is illustrated in the following example: Assuming a constant deviation of 1 % at each spiking level, this would result in the same percentage recovery of 101 % for all levels of spiked placebo, and also for the proposed calculation mode in case of a standard addition. However, if only the additions that are made to the batch are to be considered, the recoveries of a 10 %, 5 %, and 1 % standard addition would be 111 %, 121 %, and even 201 %.

2.3.2.4 Accuracy of Drug Product by Comparison

If accuracy is investigated by comparison with another analytical procedure with samples over a concentration range (as is often done for quantitative NIR [61, 62]), a linear (least square) regression may not be suitable. For the application of this regression, it is assumed that no error is present in the independent variable (*x*-values) (see Section 2.4). If this cannot be ensured, then the error in the *x*-values should be much less than those expected for the *y*-values, otherwise other statistical regressions [83] must be applied. As another option, the ratio between the validation method and the reference method can be calculated, by analogy with percentage recovery. The ratio (or *percentage accuracy*) should be plotted with respect to the content (from the reference method) and evaluated for systematic deviations and concentration dependencies (beyond an (absolute) acceptable limit). As a quantitative measure of accuracy for NIR, the standard error of prediction is recommended (Eq. 2.3-8) [82] and it should be no larger than 1.4 times the intermediate precision/reproducibility of the reference method.

$$\text{Standard error of prediction:} \quad SEP = \sqrt{\frac{\sum (y_i - Y_i)^2}{n}} \tag{2.3-8}$$

n = number of batches
y = reference method value
Y = NIR value

2.3.3 Impurities/Degradants and Water

2.3.3.1 Recovery of Spiked Impurities or Water

For impurities and degradants, an individual validation is only required if they are *specified* in the analytical procedure, by acceptance (specification) limits. If accuracy is verified by spiking, the same calculations as those described in Section 2.3.2 can be performed. Of course, for the evaluation, the larger variability at low concentration range must be taken into account. Therefore, larger differences may also be acceptable (such as 10–20 %). Often at very low concentrations matrix effects can occur, so that statistical tests should be applied with great caution. In order to reflect the conditions of the routine test appropriately, impurities and water should always be *spiked to a drug substance or drug product*, i.e., the final sample must consist of all components at (approximately) the nominal level, but with varying concentrations of the impurity to be validated. In order to avoid handling artefacts in the case of water determination, according to Karl Fischer, it may be more appropriate to spike with water after the sample has been added to the titration vessel (for example, as standard additions, see Section 8.3.9).

If several impurities are validated simultaneously, co-elution of a given impurity with peaks from other impurities may increase the peak area. In this case, these 'contributions' should be obtained from individual chromatograms and added to the spiked amount. This overall value is then the theoretical concentration. The matrix should, as far as possible, be free from the respective impurity. A matrix containing less than 10 – 20 % of the lowest spiked impurity concentration will be acceptable, because this amount does not influence the result markedly (see Table 2.3-2, lines 3 and 4). If the matrix contains more impurity, the recovered amount cannot only be related to the spiked amount, because the variability is a function of the overall concentration, not only of the spiked one. Thus, the error would be overestimated (see Table 2.3-3, lines 1 and 2). Instead, the spiked amount and the amount already present in the matrix should be combined to calculate the overall amount of impurity in the matrix (see Table 2.3-2, columns 'overall').

Table 2.3-2: Recovery from an impurity-containing matrix.

No.	Concentration of impurity					Recovery	
	in matrix	spiked	overall	found with respect to		with respect to	
				overall	spiked	overall	spiked
1	0.02	0.01	0.03	0.039	0.019	130.0 %	190.0 %
2	0.02	0.03	0.05	0.060	0.040	120.0 %	133.3 %
3	0.02	0.08	0.10	0.090	0.070	90.0 %	87.5 %
4	0.02	0.23	0.25	0.280	0.260	112.0 %	113.0 %
5	0.02	0.48	0.50	0.475	0.455	95.0 %	94.8 %
6	0.02	0.73	0.75	0.813	0.793	108.4 %	108.6 %
7	0.02	0.98	1.00	1.010	0.990	101.0 %	101.0 %

2.3.3.2 Accuracy of the Integration Mode

Non-linear behaviour and/or systematic deviation in the recoveries of small concentrations in the case of *partly resolved impurity peaks* might be caused by the use of an inappropriate integration mode (drop or rider). This is also of importance for unknown (or unspecified) impurities, as the peak area can vary substantially according to the integration mode (Fig. 2.3-2). The correct (or acceptable) mode can be verified by comparing the results of the method to be validated with those obtained by a more selective method (e.g., by extended chromatography or column switching). Another possibility is to investigate the elution behaviour of the active substance without the respective impurity, either using batches with lower impurity content or by re-chromatography of a heart-cut peak fraction without the respective impurity (Fig. 2.3-2, dotted line). In the example of a semi-synthetic peptide shown in Figure 2.3-2, integration as a rider peak would underestimate the amount of impurity substantially.

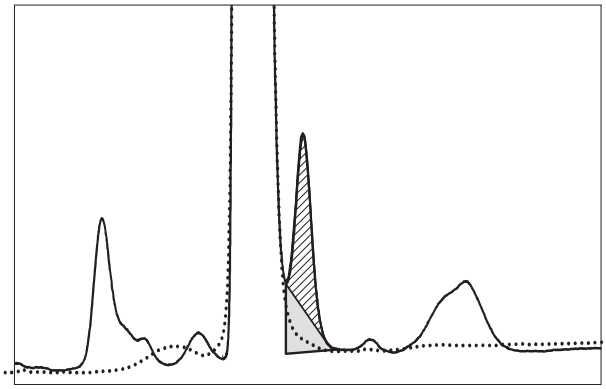


Figure 2.3-2: The influence of the mode of peak integration on the result. The solid and dotted lines represent the chromatogram of a peptide sample and an overlay of a heart-cut, re-chromatographed fraction from the main peak (in order to obtain an impurity-free sample), respectively. In the case of an integration as rider (hatched area), the impurity peak is only 62 % of the area obtained by a drop integration (grey area).

2.3.3.3 Response Factors

If the analytical response for identical concentrations of active and impurity is different, and the latter is to be quantified by area normalisation (100 % standard) or by using an external calibration with the active itself, a correction or response factor must be determined. In contrast to recovery, the calculation of response factors can be performed with the impurities alone, because they are an absolute property of the substances involved. In order to minimise the experimental variabilities, both impurity and active should be analysed in the same concentration range, sufficiently above the quantitation limit. The response factor can be calculated from the slopes of the two regression lines. If a linear relationship and a negligible intercept is dem-

onstrated (see Section 2.4.1), the response factor can also be calculated from a single concentration, with an appropriate number of determinations, i.e., at least six. An appropriate rounding should be applied, taking into account the variability of the determination, and also the uncertainty in the assigned content of the impurity reference material. One decimal figure is usually sufficient. Taking all uncertainties into account, the response factor can also defined to be unity, if it is within an acceptable range, for example, 0.8–1.2.

Unknown or unavailable impurities

In the case of unknown or unavailable impurities, response factors of unity are usually assumed [1c]. However, it is recommended to check whether the response factors deviate substantially from unity. This can be done initially by comparing the spectra of all impurities to those of the active, or by comparing the normal chromatogram with one obtained at a different, preferably low, wavelength. Large differences (factor of 5–10), as observed in Figure 2.3-3 for the peak pair at about 27 minutes, may indicate different extinction coefficients, i.e., response factors. However, it cannot provide information on whether the response factor is larger or smaller than unity.

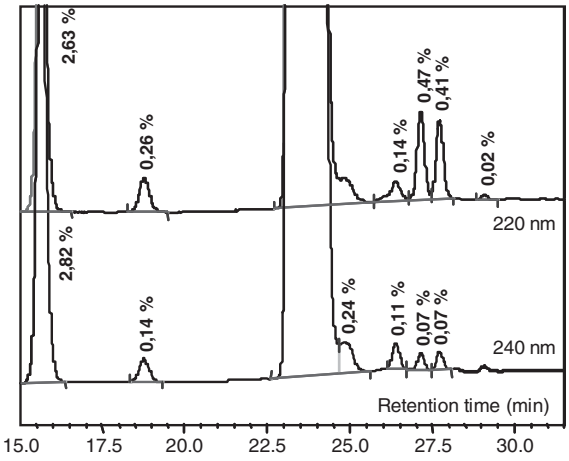


Figure 2.3-3: Chromatogram overlay at the nominal wavelength (240nm) and a check-wavelength (220nm). The chromatograms were normalised with respect to the main peak and the impurity peaks are quantified as percentage area.

Further investigations into such suspect impurities may include studies to identify their structure, their synthesis, and the experimental determination of their response factors. Alternatively, the absolute content, or ratio of impurity and drug substance could be determined analytically, for example, by refractive index (RI) detection. This detection mode results in a mass specific response [80] (at least in the same class of compounds) but is very sensitive to variations in the method condition and has therefore a poor sensitivity towards the analyte, as is obvious from

Figure 2.3-4A. The upper UV-chromatogram was obtained by injecting 20 µg of the analyte, whereas for the lower RI-chromatogram an amount of 300 µg was required, with a quantitation limit of about 0.5 %. However, this was still sufficient to estimate a large response factor for the impurities indicated by the two arrows, of about nine to ten, directly from the ratio of the RI and UV impurity peak area. The main problem in this chromatography was to achieve good separation with short retention times for a sufficient peak height, under the isocratic conditions required. For a more accurate determination, the suspected impurity peaks were collected and re-chromatographed (Fig. 2.3-4B). Here, separation is not of primary concern; therefore, the chromatographic conditions can be optimised for RI detection. The same is done for the drug substance itself. The response factor can then be calculated from the ratio of UV/RI area for the drug substance and impurity.

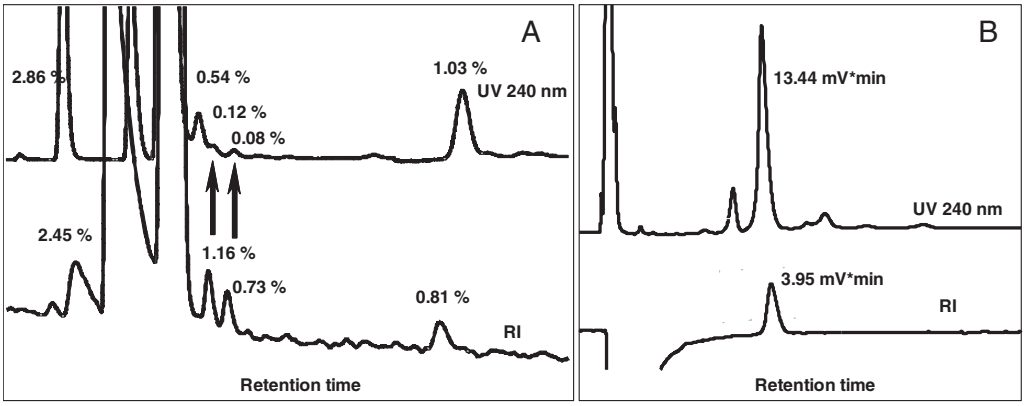


Figure 2.3-4: The estimation (A) and determination (B) of the response factor of an impurity by means of UV (upper chromatograms) and refractive index detection (lower chromatograms). The left chromatograms (A) show a whole sample, whereas in the right chromatograms (B) one of the relevant impurity peaks (labelled with arrows) was collected and re-chromatographed under conditions optimised for the RI detection.

2.3.4 Cleaning Validation Methods

2.3.4.1 Requirements

In order to prevent cross-contamination of drugs in pharmaceutical production, the cleaning of the manufacturing equipment is an important GMP aspect [84-86]. The process of demonstrating the efficiency of the cleaning procedure is known as cleaning validation. As one part of the whole process, the analytical procedures applied (in this section termed 'cleaning methods' to distinguish them from the (equipment) cleaning procedure) must of course be validated. Often, the efficiency of the cleaning procedure is investigated by swabbing defined areas of the cleaned equipment surfaces with an appropriate material. The residual substance(s) sampled from the cleaned surface are then extracted and their amount analysed. With respect to chemical substances, it includes primarily active ingredients and cleaning agents,

but degradants, raw materials, intermediates, or excipients may also be of concern. The maximum acceptable amount of residue is dependent on the pharmacological or toxicological activity of the respective substance [1f, 84], on the batch sizes and doses of the previous and next product, and on the equipment surface. A maximum limit of 10 ppm in the next batch is often established. This 'residual cleaning limit' is then normalised with respect to the sampled equipment area, as the '*specific residual cleaning limit*' (SRCL). Reported SRCLs are between 4 ng/cm^2 and 3 µg/cm^2 [13, 87, 88].

2.3.4.2 Integration of development and validation

Of course, validation of cleaning methods should follow the same rules [1a,b], but some aspects need special consideration, as regulated by the intended application. Therefore, sensitivity [86] and recovery [1f] are of particular importance. It is crucial to realise that the sampling procedure is an integral – and often the dominating – part of the cleaning method! Therefore, validating the analytical technique alone, with standard solutions, is not appropriate. Development and optimisation of the analytical procedure and its validation is an iterative process in which the influence of the cleaning solvent, the swab material, the swabbing solvent, the sampling technique, and the extraction of the analyte, on the recovery is investigated. In this *explorative stage*, the recovery at one single concentration, preferably at the defined limit (i.e., 100 % target concentration), is sufficient.

2.3.4.3 Recovery investigations

After the conditions of sampling and sample preparation have been optimised, the accuracy, precision, and quantitation limit can be validated simultaneously in a range of at least 50 – 250 % of the cleaning limit, using at least nine spikings. The higher upper range is required, because the SRCL is usually defined for the average of several (e.g., three) individual sampling sites of an equipment part, whereas the individual residues can be up to twice the limit. In cleaning (validation), no 'authentic', homogeneous samples are available. Therefore, precision of the analytical procedure must be estimated using spiked samples. Recovery is performed from the spiked surface(s) of identical equipment material, often also from spiked swabs. The latter may be omitted if the former recovery is acceptable. If interference is suspected from excipients, the spiking of the active should be performed in their presence. The robustness of the recovery, which will also include the swabbing, should be investigated by repeating the recovery with another operator. The contribution of other factors, such as analytical instrument, reagents, etc., may be investigated as well, but can be expected to be small compared to the sampling. Due to dependence on surface properties and material, it is often not possible to recover the analyte completely. Values larger than 80 % and 50 % are regarded as good and reasonable, respectively, whilst less than 50 % is questionable [85]. When relevant (with respect to the precision), the – appropriately rounded – recovery factor should be used to correct the results of the cleaning method. In order to allow a straightforward evaluation, it is preferable that the recoveries are presented graphically as percentage recovery with respect to the spiked concentration. This plot should be inspected for

a concentration dependent behaviour. An acceptable range corresponds, approximately, to three times an acceptable precision (see Section 2.3.5). When there is no (practically relevant) concentration effect, the average and the relative standard deviation of all recoveries can be calculated. The overall average from the intermediate recovery study is then used as the recovery factor. In the case of a concentration dependency, either a concentration dependent recovery factor is used, or it is calculated at the cleaning limit as the relevant concentration. If a sufficient number of determinations have been performed (at least five), the average and relative standard deviation can be calculated for each concentration level.

In Figure 2.3-5, an example is given for the recoveries of a drug substance from swabs and stainless steel plates [13]. Meclizine, i.e., 1-(*p*-chloro-(-phenyl-benzyl)-4-*m*-methylbenzyl) piperazine dihydrochloride, is practically insoluble in water and slightly soluble in diluted acids. The SRCL was established to 50 µg/100 cm^2. The system precision of the LC assay of 0.21 % and 0.41 % was obtained from five repeated injections at a concentration corresponding to 200 % SRCL. The recoveries from five swabs and five plates at each of the three concentration levels were obtained. No relevant difference was observed between the recoveries from swabs and plates, therefore, it can be concluded that meclizine can be recovered almost completely from the stainless steel surface, and the main loss is due to adsorption on the swab material. However, the recoveries are well within the limits of acceptability, with only slight concentration dependency. The much larger standard deviations of the recoveries compared to the system precisions, show that the latter precision levels contribute only marginally to the analytical variability.

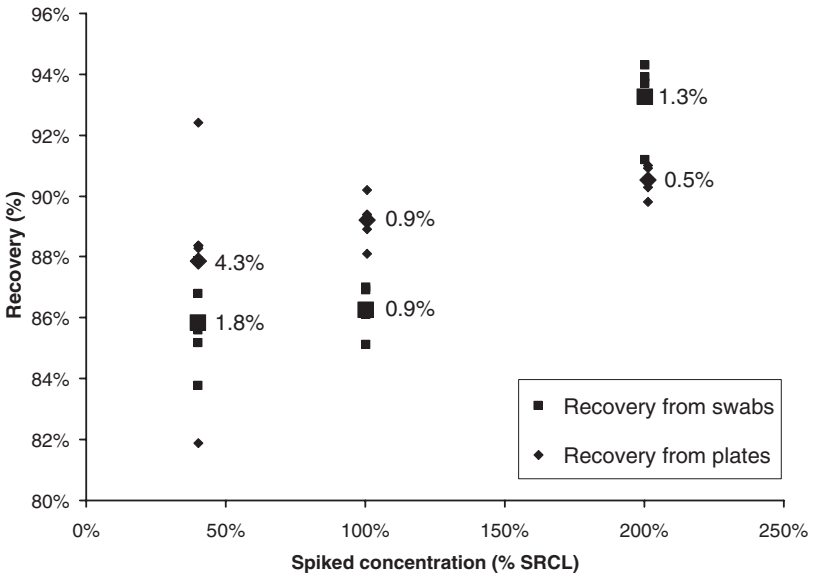


Figure 2.3-5: Recovery of meclizine from stainless steel plates and from swabs (data from [13]). Five individual spikings per concentration level were performed (smaller symbols). To the right of each mean (larger symbols), the relative standard deviation is indicated.

2.3.4.4 Stability investigations

A very important aspect to be considered in cleaning validation is the aging of the sample at the various stages of the process, in order to define the conditions appropriately. In the given example, the aging effect was investigated with respect to the analyte on dry stainless steel plates, on moistened swabs, and in the extraction solution from the swabs (Fig. 2.3-6). Large effects can be observed for the aging on the plates and in the extraction solution, both with respect to the recovery and the variability. Therefore, the time between cleaning and sampling from the steel surface and the shelf life of the extraction solution needs to be limited in the cleaning validation process. No changes were observed for the storage of the moistened swabs. Due to the intermediate recovery conditions in the stability study, the average of these twelve samples of 89.9 % would be very suitable for defining the recovery factor.

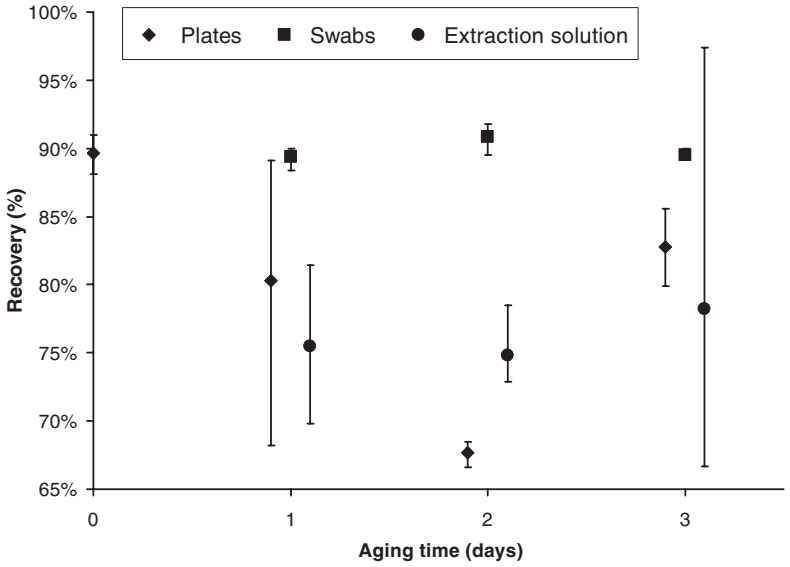


Figure 2.3-6: Influence of aging on the recovery of meclizine (data from [13]). The investigations were performed at 200 % SCRL with three spikings per storage interval. The means are symbolised and the bars indicate the minimum and maximum recovery of each series. Aging of the samples on the swabs did not affect the result. The overall average and the relative standard deviation for these 12 samples were 89.9 % and 1.15 %, respectively.

2.3.5 Acceptance Criteria

A statistical significance test, such as the *t*-test or 95 % confidence intervals for recovery should be used cautiously as acceptance criteria, because they do not take into consideration the *practical relevance* and are sensitive to small variabilities and a large number of determinations (see Sections 1.4.2 and 2.3.1). A maximum permitted absolute difference of the mean recovery to the theoretical value of 100 % or

between the means and the reference, in the case of comparison, can be defined from experience, taking the requirements of the analytical procedure into account, for example, ±2 % for LC assay [66] or 0.3 % for the proportional systematic error of volumetric titrations (see Section 8.4.1).

The maximum acceptable difference may also be derived from *statistical considerations*. The *t*-test can be regarded as the description of the relationship between a difference (between two means or to a reference) and a standard deviation. Rearranging the corresponding equations (2.3-1) and (2.3-2), the maximum permitted difference is given as a function of the (maximum permitted) standard deviation (Eq. 2.3-9). The factor *F* depends only on the number of determinations and whether the comparison is to a nominal value or another experimental mean result. Under the usually applied conditions, the factors are near unity (Table 2.3-3). Therefore, an acceptable precision ought to be used as an indication of a suitable difference acceptance limit with respect to means.

If individual recoveries are evaluated, larger ranges must be taken into account. For nine determinations, a maximum range of approximately 4.5 times the standard deviation can be expected, corresponding to 6 times the target standard deviation (see Section 2.1.3).

Relation between precision and difference:

$$\text{to a nominal value:} \quad D \leq \frac{t\,(P,df)}{\sqrt{n}}\,s = Fs \quad df = n\text{-}1$$

$$\text{between (two) means} \quad D \leq t\,(P,df)\sqrt{\frac{2}{n}}\,s = Fs \quad df = 2n\text{-}2 \qquad (2.3\text{-}9)$$

Table 2.3-3 Factors to obtain the maximum permitted difference from the standard deviation (Eq. 2.3-9).

n	Factor *F* for comparison with	
	a nominal value	another mean
6	1.05	1.29
9	0.77	1.00

Can this theoretically obtained relationship be supported by experimental results?

The means and the relative standard deviations of 36 recovery series for LC assays of 18 drug products are shown in Figure 2.3-7. The usual spiking range of the active into the placebo was 80–120 or 70–130 %, the number of determinations ranged from five to nine. If sufficient data were available, the concentration levels are shown separately. In order to limit the influence of possible extreme results, only 90 % of all results were taken into account. The mean recoveries range from 99.5 to 101.4 %. Due to the relatively small number of data, further classification according to the type of drug product is not possible. It seems, nonetheless, that the deviations from the theoretical value are slightly larger for tablets. The average bias was calculated to be 0.5 %, and the individual deviations range from 0.1 to 1.4 % (90 % distribution). From the relative standard deviations of the individual recoveries observed there seems to be no relation to the type of drug product (Fig. 2.3-7b). The lower and

upper limit of the 90 % distribution were determined to 0.20 % and 1.22 %, respectively. The average RSD was calculated to be 0.6 %. These results are very similar to the repeatabilities obtained for drug substances, lyophilisates, and solutions (see Section 2.1.3.2, Table 2.1-5). The variability contributions in the recovery experiments are different from those related to authentic samples. However, the results demonstrate that the variability of the recovery studies is not so different from the repeatability observed with drug substances and drug products requiring less complex sample preparation. Therefore, the theoretical ratio of approximately unity between analytical variability and recovery deviation could indeed be confirmed experimentally.

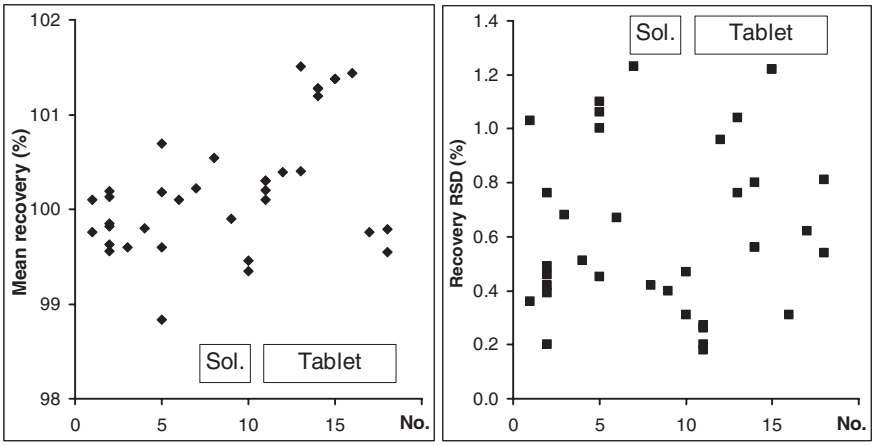


Figure 2.3-7: Mean (a) and relative standard deviation (b) of recoveries for LC-assays of 18 drug products. The results are sorted according to the type of drug product, with lyophilisates (No. 1-2), others (No. 3-6), solutions (No. 7-10), and tablets (No. 11-18).

2.3.6
Key Points

- The same calibration should be used as is intended for the routine application.
- The accuracy of drug substance assay should be validated by comparison with another (preferably) absolute procedure.
- For drug product, the evaluation of percentage recovery is recommended, due to simpler interpretation.
- Impurities (if specified and available) should be spiked into the drug substance or drug product.
- Absolute acceptance criteria (for deviation between mean results or to a target) or statistical equivalence tests should be preferred, because here a measure of the practical relevance can be included.
- An acceptable absolute difference between means corresponds approximately to an acceptable precision.

2.4 Linearity

Joachim Ermer

ICH

"*The linearity of an analytical procedure is its ability (within a given range) to obtain test results which are directly proportional to the concentration (amount) of analyte in the sample*". [1a]

It may be demonstrated directly on the analyte, or on spiked samples using at least five concentrations over the whole working range.

Besides a visual evaluation of the analyte signal as a function of the concentration, appropriate statistical calculations are recommended, such as a linear regression. The parameters slope and intercept, residual sum of squares and the coefficient of correlation should be reported. A graphical presentation of the data and the residuals is recommended.

The terminology for this validation characteristic is somewhat misleading, because linearity in the inner sense, i.e., a linear relationship between analyte concentration and test results is certainly preferable, but not essential. A better term would have been 'analytical response'. Some analytical procedures have intrinsic non-linear response functions, such as quantitative TLC, fluorescence detection, etc., but they can of course be validated. The primary objective is to validate or verify the *calibration model*. Consequently, the requirements and the relevant parameters depend on the intended mode of calibration (see Table 2.4-1).

The response function of a given analytical procedure is an intrinsic property of the respective analyte. That means, with respect to validation, that the answer is of a qualitative kind: Can the intended calibration be applied, yes or no? Therefore, solutions of the analyte itself are sufficient and there is no need to repeat linearity. Potential influences by the matrix, i.e., the linearity of the analytical procedure would be better addressed in accuracy (see Section 2.3.2).

Often, the fundamental response function is known for a given type of analytical procedure, such as a linear function for LC with UV detection, according to the Lambert–Beer law. In such cases, validation of linearity can be regarded more as a *verification* of the assumed response function, i.e., the absence of (unacceptable) deviations. Primarily, this should be performed by means of *graphical evaluation* of the deviations of the experimental data from the assumed response model (residual analysis), known as residual plots. The evaluation of numerical parameters is only sensible after verification of the response function, i.e., if only random errors exist.

Table 2.4-1 Requirements for different calibration models.

Calibration model	Requirements
Single-point calibration	Linear response function
(single external standard concentration)	Negligible constant systematic error (ordinate intercept)
	Homogeneity of variances[a]
Multiple-point calibration	
Linear, unweighted	Linear response function
	Homogeneity of variances[a]
Linear, weighted	Linear response function
Non-linear	Continuous response function
100%-method	For main peak
(area normalisation for impurities):	Linear response function
	Negligible constant systematic error (ordinate intercept)
	Homogeneity of variances[a]
	For impurities:
	Linear response function
	Negligible constant systematic error (ordinate intercept)

a) Homoscedasticity, constant variance: may be assumed within a limited concentration range (factor ~10)

2.4.1 Unweighted Linear Regression

Prerequisites

The most simple and popular calibration is a linear model, which is usually validated by means of an unweighted linear regression. In order to highlight some practical requirements that need to be fulfilled, but are sometimes neglected in validation literature, the fundamentals are briefly illustrated in Figure 2.4-1. In this regression, the straight line, which produces the best fit to the experimental data, is constructed. This best fit, or smallest possible difference is obtained by minimising the distances between the experimental points and the regression line, the so-called residuals. Since positive and negative deviations would cancel each other out, summarising the residuals, the squares of the residuals are summarised and minimised. Therefore, this regression is also called a least-squares regression. It is an intrinsic property that each regression line passes the data pair of the averaged experimental x and y-values (for details, see statistical text books). It is important to be aware that the x-values (or independent variables) are assumed to be *error-free* (because the vertical residuals are minimised). Only the y-values (or dependent variables) are assumed to be randomly distributed. These prerequisites are often fulfilled, because the x-values in a calibration are usually obtained from preparing well-characterised materials and the preparation error is much less and is also negligible compared to

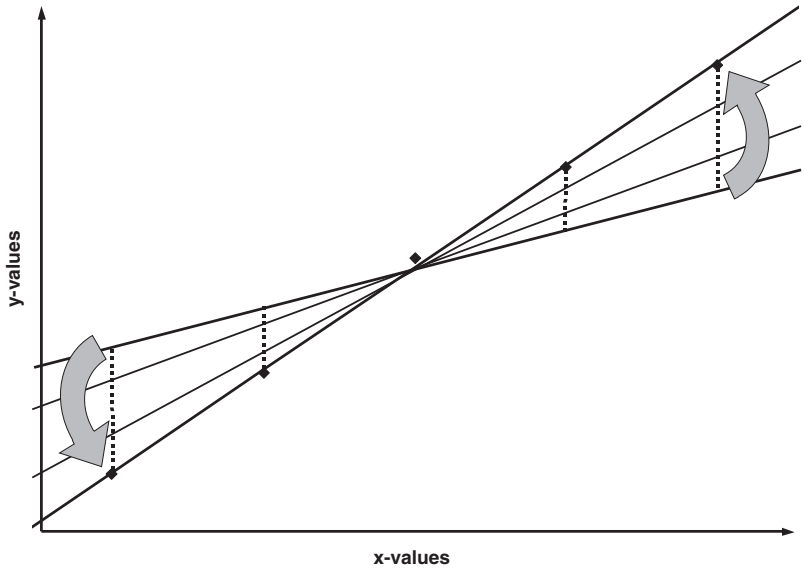


Figure 2.4-1: The principle of least-squares regression. The vertical distances between the experimental data and the regression line (i.e., the residuals, dotted lines) are squared and the line is varied until the sum of the squared residuals is at the minimum.

the measurement variability of the y-values. However, the analyst needs to take this into consideration for some applications, such as linearity from (complex) spiked samples, calibration against the results of another analytical procedure, etc.

Larger concentrations with larger response values will have a greater influence on this type of regression, because reducing larger residuals has more impact in the minimisation of the sum of squares. Consequently, very small concentrations are more or less neglected, as is obvious from their large (relative) deviation from the regression line (see Figs. 2.4-9 and 2.4-10). Therefore, an essential prerequisite for the unweighted linear regression is to use only concentration ranges in which the response data have a comparable variability, also termed homogeneity of variances or homoscedasticity (see also Section 2.4.2). This prerequisite can be assumed to be fulfilled, if the standard deviations of the data do not vary by more than a factor of 1.5–3 [89]. With respect to UV-detection, this corresponds approximately to a ten-fold concentration range. A too-large range for linearity data can sometimes also be observed in the validation literature. In six out of 46 validation papers reviewed, published between 1997 and 2003, inappropriate ranges for unweighted linear regression were used, with a ratio between the minimum and maximum concentration of up to 2000! Such mistakes usually do not impair the performance of the analytical procedure with respect to linearity (in contrast to the quantitation limit, see Section 2.6.4), but only due to the fact that they have intrinsically a linear response function. However, this should be no excuse for inappropriate experimental design. If large concentration ranges are (really) required, i.e., if analyte concentrations can be expected (anywhere) within a larger range, a *weighted regression* must be performed.

Here, the influence of the small concentrations increases (see Section 2.4.2). The equations for the parameters of both weighted and unweighted linear regression follow. For unweighted regression, the weighing factor w_i to calculate the means and the sum of squares is set to unity.

Residual sum of squares: $RSS = Q_{yy} - \frac{Q_{xy}^2}{Q_{xx}} = \sum (y_i - (a + bx_i))^2$ (2.4-1)

Residual standard deviation: $s_y = \sqrt{\frac{RSS}{n-2}}$ (2.4-2)

Slope: $b = \frac{Q_{xy}}{Q_{xx}}$ (2.4-3)

Standard deviation of the slope: $s_b = \sqrt{\frac{s_y^2}{Q_{xx}}}$ (2.4-4)

Relative confidence interval of the slope: $CI_b = \frac{100 \;\; t(P,n-2) \;\; \sqrt{s_b^2}}{b}$ [%] (2.4-5)

Intercept: $a = \bar{y} - b\,\bar{x}$ (2.4-6)

Standard deviation of the intercept: $s_a = s_y \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{Q_{xx}}}$ (2.4-7)

Confidence interval of the intercept: $CI_a = \; t(P, n-2)\; s_a$ (2.4-8)

Relative residual standard deviation: $V_{x0} = \; 100 \; \frac{s_y}{b\bar{x}}$ [%] (2.4-9)

Coefficient of correlation: $r = \; \frac{Q_{xy}}{\sqrt{Q_{xx} \; Q_{yy}}}$ (2.4-10)

Means: $\overline{x} = \frac{\sum (x_i k_w w_i)}{n} \qquad \overline{y} = \frac{\sum (y_i k_w w_i)}{n}$ (2.4-11)

Sum of squares: $Q_{xx} = \sum (k_w w_i (x_i - \bar{x})^2)$

$Q_{yy} = \sum (k_w w_i * (y_i - \bar{y})^2) \qquad Q_{xy} = \sum (k_w w_i (x_i - \bar{x}) \;\; (y_i - \bar{y}))$ (2.4-12)

$$\text{Normalisation factor: } k_w = \frac{n}{\sum (w_i)} \tag{2.4-13}$$

$$\text{Confidence interval at } x_i\text{: } y_i \pm t(P\%, n-2)\ s_y \sqrt{\frac{1}{n} + \frac{(x_i - \bar{x})^2}{Q_{xx}}} \tag{2.4-14}$$

$$\text{Prediction interval at } x_i\text{: } y_i \pm t(P\%, n-2)\ s_y \sqrt{\frac{1}{m} + \frac{1}{n} + \frac{(x_i - \bar{x})^2}{Q_{xx}}} \tag{2.4-15}$$

$$\text{Uncertainty of } x_i\text{: } x_i \pm \frac{t(P\%, n-2)\ s_y}{b} \sqrt{\frac{1}{m} + \frac{1}{n} + \frac{(x_i - \bar{x})^2}{Q_{xx}}} \tag{2.4-16}$$

n = number of data for validation

m = number of determinations in future application for which the prediction interval / uncertainty is intended, i.e., m=1 f or a single determination (also called single-use), $m > 1$ for means (multiple-use)

P = statistical confidence, 100 – error probability. Usually, an error probability α of 5 % is used, i.e., $P = 95\,\%$

$t(P,n\text{-}2)$ = Student t-factor for the given statistical confidence P and the degrees of freedom (n-2), i.e., for two estimated parameters in a linear regression.

Confidence and prediction intervals

Important parameters used to evaluate the variability of regression lines and data are the confidence and prediction intervals (Fig. 2.4-2). The former refer to the variability of the regression line, i.e., the true line can be expected within this interval. The term under the square-root in Eq. (2.4-14) is also called *leverage*. It increases with the distance of the data from their mean. Therefore, a distant data point which is biased or exhibits an extreme variability will have a high impact on the regression line, i.e., a high leverage effect. It is proposed that leverage values larger than 0.5 should be avoided [23].

The *prediction interval* aims at future data, i.e., within this interval, around the regression line, a further determination (or the mean of several determinations, if m>1) can be expected, with the defined statistical confidence P. The prediction interval can also be used to investigate suspect data, by means of the outlier-test according to Huber [90]. The regression is repeated without the suspected data pair and if it is now outside the new prediction interval, they may be regarded as outliers (indicted by square in Fig. 2.4-2). However, the cause of a possible outlying result should always be identified before data are removed from the statistical analysis (see discussion in Chapter 10). There are other regression techniques, which are less sensitive to outlying results (robust statistics) (see Section 2.4.3). The prediction interval can also provide information about the variability (or uncertainty) of the analyte determination itself. In routine analysis, the inverse of the defined calibration function is used to calculate the corresponding analyte concentration from the ana-

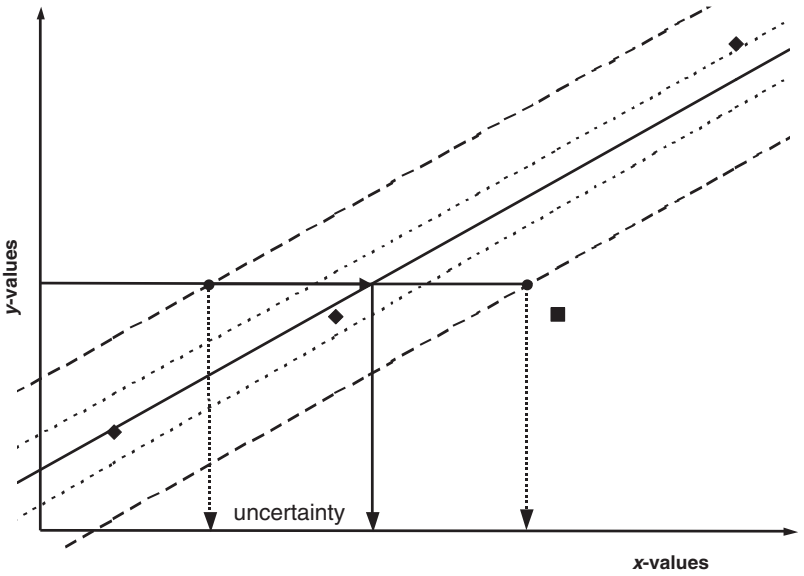


Figure 2.4-2: Linear regression line (solid) with the limits of the confidence (dotted line) and prediction intervals (broken line). The resultant uncertainty Δx and a possible outlier (square) are indicated. For details, see text.

lytical response. This is illustrated by the horizontal line in Figure 2.4-2. The intersection of this line with the regression line and its projection onto the x-axis provides the estimated analyte concentration (solid vertical arrow). The corresponding intersections with the upper and lower limits of the prediction interval provide the *uncertainty range* for this concentration, i.e., the confidence interval of the predicted analyte concentration. However, the reader must be very aware that this uncertainty is only valid for the given regression. Because the ultimate target is the performance of the routine application, the described approach must be applied to the routinely intended calibration model to achieve relevant variability estimates. For a single point calibration, the prediction interval is calculated from the response of the repeated standard determinations using Eq. (6-2) and divided by the slope b, in order to convert the response into a concentration, i.e., to obtain the prediction interval of x_i. As these prediction intervals include the variability of the standard determination, they can be regarded as the minimum estimate of the intermediate precision, of course without the contribution of the other variables such as the operator, the equipment or the time, etc.

2.4.1.1 Graphical Evaluation of Linearity

Residuals

The simplest approach in order to identify deviations from the assumed linear model is the investigation of the residuals, i.e., the difference between the experimental response, and that calculated from the regression line. If the model is correct, the residuals are randomly distributed, i.e., they are normally distributed, with

a (true) mean of zero. Therefore, the actual distribution of the residuals can be examined for deviations, such as non-normality, non-linearity (or in general, lack of fit), and heteroscedasticity (see Section 2.4.2) by means of graphical presentation or statistical tests. The residuals can also be normalised by their standard deviation, also called scaled or Studentised residuals (Eq. 2.4-17), in order to avoid scaling distortions.

$$\text{Studentised residuals: } r_{s,i} = \frac{r_i}{s_r \sqrt{1 - \frac{1}{n} + \frac{(x_i - \bar{x})^2}{Q_{xx}}}} \quad r_i = y_i - (a + bx_i) \tag{2.4-17}$$

with s_r = standard deviation of the residuals r_i

A plot of the residuals should always be performed, usually versus the x-values. They can also be plotted versus other parameters such as the response calculated from the regression line or in serial order, to reveal instabilities or progressive shifts in the analytical conditions (provided that the various concentrations were analysed in a random order) [91]. A *visual evaluation* of the pattern of the residuals is a very simple and straightforward, but nevertheless powerful, tool to detect deviations from the regression model [1b, 8]. If the linear, unweighted regression model is correct, the residual plot must show random behaviour in a constant range, without systematic pattern or regularities (Fig. 2.4-3B). Non-linear behaviour will result in systematic or curved pattern of the residuals, heteroscedasticity in a wedge-shaped distribution, with increasing residuals (Fig. 2.4-3A). In order to investigate if the pattern is significant, replicate measurements are required, to provide information about the inherent variability of the response for each concentration (corresponding to the 'pure error' in the statistical lack of fit test). This measurement variability is then compared to the (systematic) deviations of the residuals from zero. If the latter is much larger than the former, the linear model may be inappropriate. Of course, to use this approach, as well as the corresponding statistical tests (see below), requires a sufficient number of data. It is recommended to use eight or more concentrations with duplicate determinations [92], or a three-point design with six to eight replications [91]. Another (or additional) option is to define an acceptable dispersion range of the residuals is increasing with the number of data and corresponds to about four to five times the (true) standard deviation. For this purpose, it is recommended that the residuals are normalised with respect to the calculated response. By defining an acceptable range, rejection of the linear model due to slight systematic deviations, which are of no practical relevance, can be avoided (Fig. 2.4-6).

Sensitivities

Another, very powerful approach used to detect deviations from linearity is the graphical presentation of the *sensitivities*, i.e., the ratio of the analytical signal and the corresponding concentration (also called response factor) as a function of the concentration. In the case of a linear response function with zero intercept, the sensitiv-

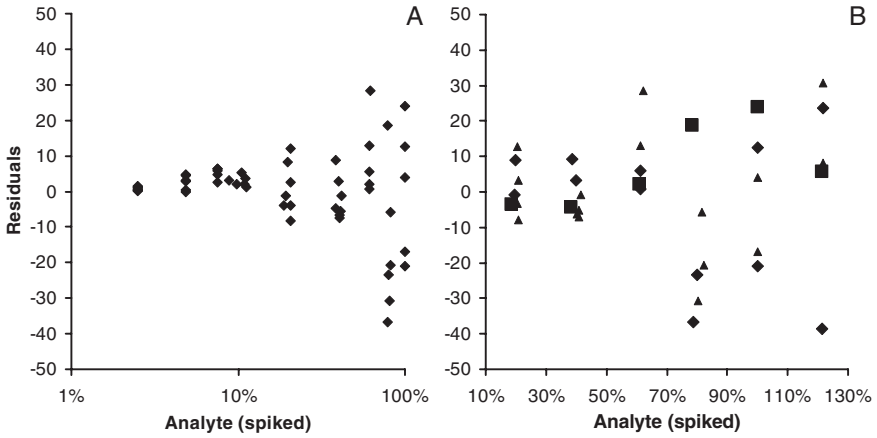


Figure 2.4-3: Residual plot for an unweighted linear regression of an LC assay (data from [65]). A: Non-appropriate concentration range 0.025 – 120 %. B: Suitable concentration range 20 – 120 %. Usually, the number of determinations per concentration will be smaller, but the example was chosen to illustrate the non-constant variability in A (heteroscedasticity). In order to visualise the effect of a smaller number of repetitions in B, the first sample of each concentration is symbolised by a square, the second and third by diamonds, and the remaining by triangles. The scale of the residuals corresponds to ±1.5 % with respect to the peak area at the working concentration.

ities are constant within a certain distribution range (Fig. 2.4-4). The ASTM recommends an interval of 5 % around the sensitivity average for the linear range of a detector [93]. However, this interval should be adjusted to the concentration range and application in question. Again, the dispersion range with four to five times the (true) standard deviation (Eq. 2.1-4) can be used, for example 2–3 % for an LC assay. For larger concentration ranges, the expected precision at the lower end should be taken for orientation, because these concentrations have a greater influence on the sensitivities. Even for constant variability, the dispersion of the sensitivities is increased for smaller concentrations, because they appear in the denominator of the ratio (Fig. 2.4-4). The advantage of the sensitivity plot is that deviations are easily identified even in a small number of data points, where the randomness of the residuals is difficult to evaluate. However, a constant systematic error, represented by a significant intercept, will also cause a particular trend in the sensitivities.

The lack of a measure of practical relevance is the main disadvantage of statistical linearity tests (see also Section 1.4.2). Therefore, it is recommended that such tests be applied (see Section 2.4.1.3) only if deviations from linearity must be assumed or are indicated by the graphical evaluation. For verification purposes, the evaluation of the plots is usually sufficient.

2.4.1.2 Numerical Linearity Parameters

Numerical parameters of the regression are only meaningful for evaluating the performance of the analytical procedure after verification of a linear response function.

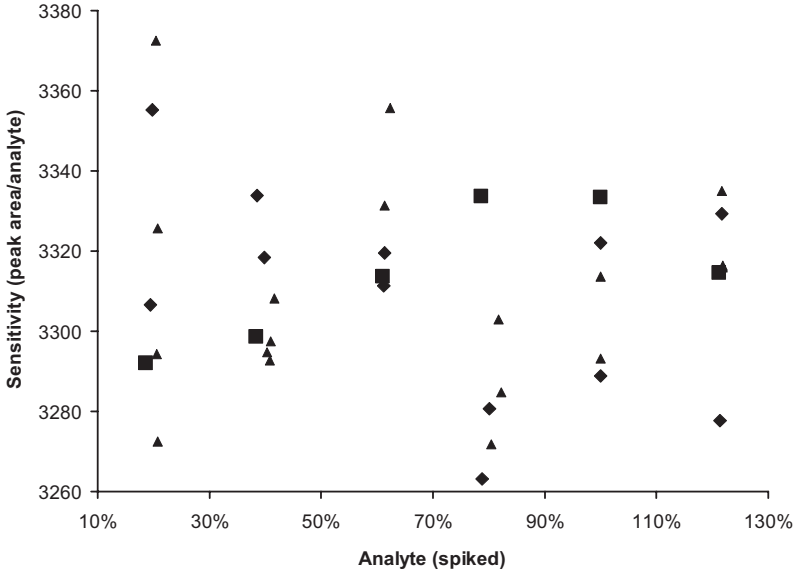


Figure 2.4-4: Sensitivity Plot for linearity of an LC assay (data from [65]). Usually, the number of determinations per concentration will be smaller, but the example was chosen to illustrate the different influence of the concentration on the data dispersion in comparison with the residual plot (Fig. 2.4-3B). In order to visualise the effect of a smaller number of repetitions, the first sample of each concentration is symbolised by a square, the second and third by diamonds, and the remaining by triangles. The scale of the sensitivities corresponds to 3.6 % with respect to their average. The relative standard deviation is 0.75 %.

Coefficient of correlation

The coefficient of correlation is almost uniformly (mis)used, since it is neither a proof of linearity, nor a suitable general quantitative measure [92, 94, 95]. In contrast, it requires linearity as a prerequisite; therefore it cannot be used in its proof. In other words, the correlation coefficient requires random scatter around the linear regression line to have a quantitative meaning et all, but even then the numerical values cannot be properly compared, because they depend on the slope [91], as well as on the number of determinations and the regression concentration range (Fig. 2.4-5). Therefore, this parameter is not suitable as a general acceptance criterion for the performance of an analytical procedure, i.e., as a measure of the calibration variability. Whether there is a significant correlation between two variables or not is primarily dependent on the number of determinations (see Table 2.4-2). The values indicating a significant linear correlation, such as 0.878 for five, or even 0.632 for 10 determinations, will usually not be accepted for the calibration for a (chemical) assay.

Residual standard deviation

The residual standard deviation (Eq. 2.4-9) measures the deviation of the experimental values from the regression line and thus represents a good performance parameter with respect to the precision of the regression. Expressed as a percentage

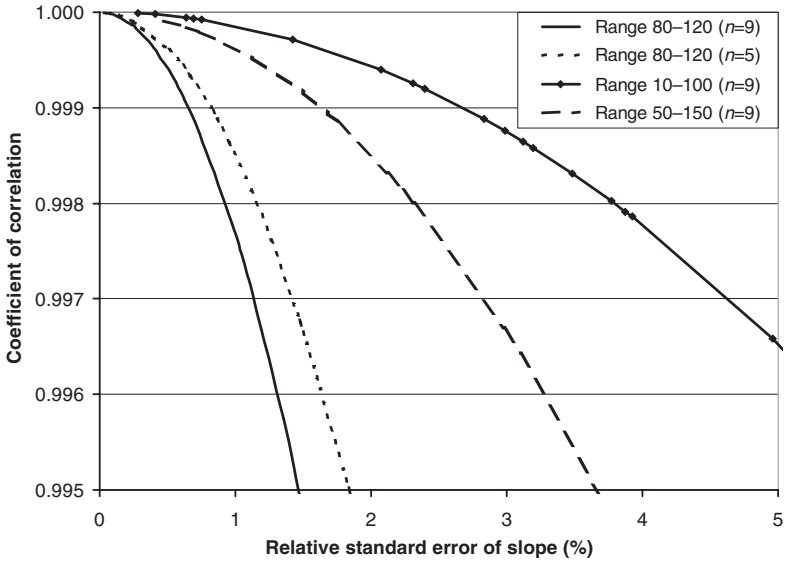


Figure 2.4-5: Relationship between the coefficient of correlation and the experimental variability. Data sets were simulated using the response function $y = x$ and normally distributed errors with a constant standard deviation for each data set. The dependence of r is shown for several concentration ranges and numbers of data.

Table 2.4-2 Statistical significance of the correlation coefficient dependent on the number of determinations (Pearson's correlation coefficient test). If the experimental value of the coefficient of correlation (r) is larger than the tabulated one for the given number of determinations (n) and a statistical confidence of 95%, a linear relationship is statistically confirmed.

N	Significant *r**	*n*	Significant *r**
5	0.878	12	0.576
6	0.811	15	0.514
7	0.754	20	0.423
8	0.707	25	0.396
9	0.666	50	0.279
10	0.632	100	0.197

*: The test is based on the following t-statistics: $t(P, n-2) = \frac{|r|\sqrt{n-2}}{\sqrt{1-r^2}}$

(relative residual error), it is comparable to the relative standard deviation obtained in precision studies in the given concentration range. Therefore, this parameter is better suited to evaluation purposes than the residual sum of squares and the residual standard deviation, which are also dispersion parameters, but they depend on the absolute magnitude of the signal values and are difficult to compare with results from other equipment or other procedures. The normalisation is performed by the mean of all x-values and the slope of the regression line. Therefore, the relative stan-

dard error is (slightly) dependent on the distribution of the x-values within the range. As an alternative, the x-mean can be replaced by the target concentration (100 % test concentration or specification limit, as proposed in Section 5.2.4, Note 2) of the whole denominator, by the target y-value (Eq. 2.1-12, Section 2.1.4.1).

2.4.1.3 **Statistical Linearity Tests**

Statistical linearity investigations, also called 'lack-of-fit' tests, are only recommended if deviations from linearity are suspected or if the intrinsic response function is unknown. The practical relevance of a statistically significant deviation must always be considered, as well as the opposite. This is illustrated in Figure 2.4-6 and Table 2.4-3. Using the range from 50 to 150 %, the quadratic fit is significantly better and a systematic behaviour of the residuals can be recognised for the linear regression. However, a spread of residuals of less than 0.6 % is irrelevant for practical purposes. This is also supported by the loss of the significance of the quadratic coefficient and the wider range for the residuals, adding another data point. In the case of intrinsic non-linearity, an extension of the regression range is likely to confirm this. In order to distinguish more reliably between intrinsic systematic behaviour of the residuals and a grouping of experimental data by chance, a larger number of concentrations (at least ten [8]) and/or replicate determinations on each concentration level (at least two) should be used. Then, a possible lack of fit can be better evaluated with respect to the variability of the data itself, as a kind of 'visual variance analysis' [91] (see below).

Alternative model

One statistical approach is to check the significance of the quadratic coefficient in a second order polynomial. This can be done by calculating the confidence intervals of the quadratic coefficient (see statistical textbooks or software, e.g., [28]). If zero is included, the coefficient is not significant, and the quadratic function is reduced to a linear one. An equivalent approach is to compare the residual standard deviations of

Table 2.4-3 Linearity of ascorbate in a drug product.

Ascorbate (%)	Peak area	Parameter	Regression range 50 – 150%	Regression range 25 – 150%
25	476	Unweighted linear regression ($y = a + b*x$)		
50	973	Slope	20.16	20.15
80	1586	Intercept	– 29.78	– 28.60
100	1993	Confidence interval (95%)	–56.46 to –3.10	– 42.44 to –14.75
120	2391	Significant difference to 0?	Yes	Yes
150	2990	As percentage signal at 100%	– 1.50%	– 1.44%
		Relative residual error	0.30%	0.30%
		Coefficient of correlation	0.99998	0.99999
		Statistical linearity tests (significance of the quadratic coefficient: $y = a + b*x + c*x^2$)		
		95% Confidence interval of c	– 0.0070 to – 0.0013	– 0.0056 to 0.0028
		Significance of c	Yes	No

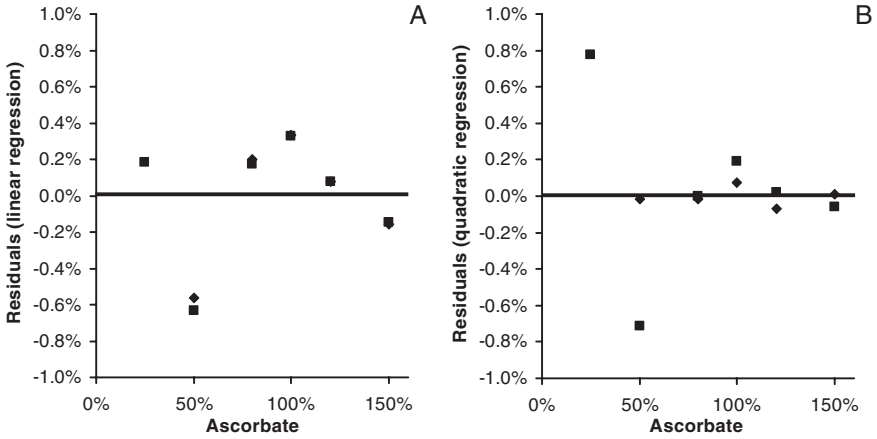


Figure 2.4-6: Residual plot for unweighted regression of an LC assay of ascorbate. A linear (A) and a quadratic (B) regression was performed in a concentration range of 25–150 % (squares) and of 50–150 % (diamonds). In order to facilitate the evaluation, the residuals are presented as percentages with respect to the fitted peak area.

a linear and a quadratic regression in order to investigate whether the latter results in a significantly better fit. This is also known as the Mandel test [96]. An essentially the same, but more complicated, test calculation has been described in [97]. These tests are also sensitive to heteroscedasticity, i.e., a statistically significant better quadratic fit could be the result of a regression range being too large.

Model independent

The disadvantage of these tests is the need for an alternative model, which also may not be the intrinsic one. This is avoided in the so-called *ANOVA lack-of fit test* [91, 98]. This test is based on an analysis of variances and requires replicated measurements for each concentration. The variability of the measurement is then compared with the deviation from the calibration model (Fig. 2.4-7). Mathematically, the sum of the squared deviations of the replicates from their respective mean at each concentration is calculated and summed for all concentrations. This is an estimator of the variability of the measurement ('pure error' SS_E). Then, the residual sum of squares of the regression (RSS, Eq. 2.4-1) is calculated from all data. This parameter includes both the pure error SS_E, and the sum of the squares due to the deviation from the regression line or due to the 'lack-of-fit error' (SS_{lof}). In case of no deviation, the latter is zero, and the RSS is identical to the measurement variability, i.e., the pure error. If not, the lack-of-fit error can be calculated from RSS and SS_E. Now, the significance of the SS_{lof} can be tested by comparing it to the SS_E. Both parameters are divided by their respective degrees of freedom and the ratio of these mean squares is used in an F-test.

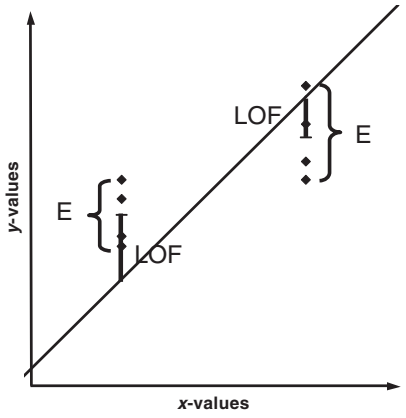


Figure 2.4-7: Illustration of the ANOVA lack-of-fit test. The measurement variability *E* ('pure error') from replicate determinations is obtained for all concentration levels and compared with the deviation of the means from the regression model LOF (lack of fit error).

$$\text{ANOVA lack-of-fit: } \frac{SS_{lof}/(k-2)}{SS_E/(n-k)} \leq F(P, k-2, n-k) \tag{2.4-18}$$

n = overall number of determinations
k = number of concentrations (with repetitions).

In order to have a sufficient number of data, it is recommended to use eight or more concentrations with duplicate determinations [92], or a three-point design with six to eight replications [91]. It is important to ensure that the variability contributions are the same for both the replicates and the preparation of the concentrations; otherwise a lack-of-fit may be identified because of additional variability in the latter. This may occur if the repetitions are obtained by repeated injections. To cope with such a situation, a modified version of the test is described in [98].

2.4.1.4 Evaluation of the Intercept (Absence of Systematic Errors)

The absence of constant systematic errors is a prerequisite for a single-point calibration and for the 100 %-method for the determination of impurities. The so-called single-point calibration represents, in fact, a two-point calibration line where one point equals zero and the other the standard concentration. This negligible intercept has to be demonstrated experimentally, a regression forced through zero is only justified afterwards.

Statistical evaluations

A negligible intercept can be demonstrated *statistically* by means of the confidence interval of the intercept, usually at 95 % level of significance (Eq. 2.4-8). If it includes zero, the true intercept can also be assumed to take zero, i.e., the intercept is statistically insignificant. Performing a *t*-test with the ratio from the intercept and its standard deviation is an identical approach to testing its statistical significance. However, a small variability may result in a significant intercept, but without any practical relevance (see Table 2.4-3). In contrast, a large variability can obscure a substantial deviation of the intercept from zero.

An alternative statistical approach, i.e., the equivalence test for the intercept, includes a measure of its practical relevance, see Section 1.4.2. A check is carried out as to whether the equivalence interval of the intercept (Eq. 2.4-19) is included in the acceptance interval around zero, as defined by the analyst.

Equivalence interval intercept (lower and upper limits)

$$C_L = a - t(P, n-2)s_y\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{Q_{xx}}},\ C_u = a + t(P, n-2)s_y\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{Q_{xx}}} \qquad (2.4\text{-}19)$$

Absolute evaluation

For an absolute evaluation, the intercept can be expressed as a percentage of the analytical signal at the target or a reference concentration, such as 100 % working concentration, in the case of assays. In fact, this approach can be regarded as an extrapolation of the variability at the working concentration, to the origin. Therefore, an acceptable precision value can be used as the acceptance limit.

Risks of extrapolation

An often-encountered problem is the impact of too large an *extrapolation*. For example, the minimum range required for an assay is 80–120 % (see Table 2.5-1). Using linearity data only in this range, the intercept is affected by a large extrapolation. This results in a high uncertainty and less reliable estimates for the intercept. The true value of the intercept for the simulations shown in Figure 2.4-8 is zero, but 'experimentally', results up to more than 5 % are obtained if the intercept is extrapolated from 80 % as the minimum concentration used for the regression. The variability of the calculated intercepts is much reduced, if lower concentrations are included. In the case of 20 % as the minimum concentration, the most extreme intercept is about 1 %, which corresponds well to the error introduced during the simulation of the data sets. Extrapolation also makes the statistical evaluation of the intercept meaningless, because the confidence intervals become very wide. Therefore, the linearity to justify a single-point calibration should be validated starting with lower concentrations, 20–40 %, but the range should not exceed about one order of magnitude to avoid heteroscedasticity (see Section 2.4.2).

Acceptable deviation

The *absence of a systematic error* can also be investigated by comparing a single-point calibration versus a multiple-point calibration (as a better estimate if an intrinsic constant error would exist), within the range required. If the difference between the two calibrations is acceptable within the working range, a (practically important) systematic error can be ruled out and the single-point calibration is justified. In the range 80–120 % (Fig. 2.4-8), the largest deviation between the regression line and a single-point calibration using the mean of all specific signals (y/x) is 1.2 %. This is still a rather large deviation, but much less compared to the extrapolated intercept of 5.5 % for the same data set. Ignoring this data set as an extreme example, the second

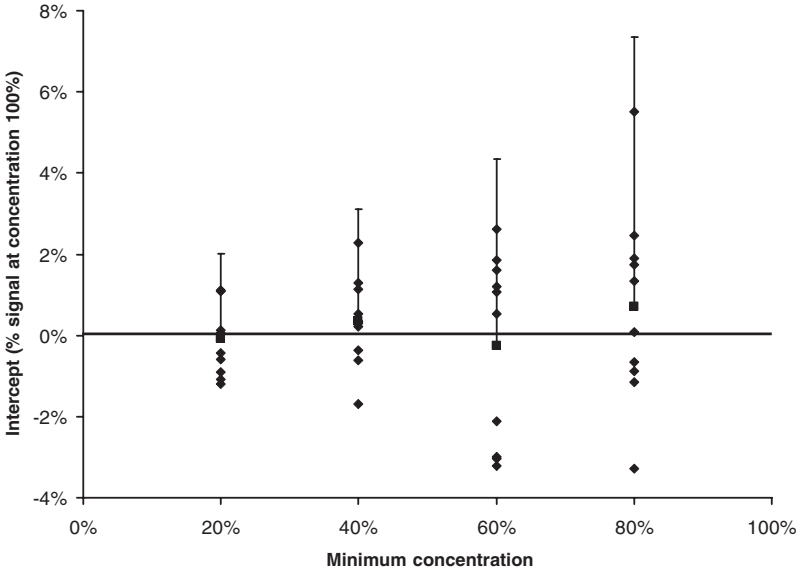


Figure 2.4-8: Evaluation of intercept and extrapolation. Data sets of seven concentrations each were used, equally distributed within the range between the minimum concentration indicated on the x-axis and 120 %. Ten data sets were simulated using the response function $y = x$ and normally distributed errors with a standard deviation of 1. The calculated intercepts are presented with respect to the theoretical signal concentration at 100 %, as diamonds. The average of the ten intercepts is indicated by squares, together with the upper average 95 % confidence interval shown as error bars.

largest deviation is 0.7 %, and the second largest extrapolated intercept 3.3 %. However, even if this approach is sufficient to justify a single-point calibration, the author recommends investigating the intercept, if an intrinsic value of zero can be expected for the given calibration. Then, possible systematic errors can be detected.

Such an approach can also be applied to justify a calibration model within a defined working range, even if it is not the intrinsic response function.

2.4.2 Weighted Linear Regression

One prerequisite for an unweighted linear regression is a constant variability of the y-values over the whole concentration range. In LC-UV and CE-UV procedures, this can be expected for one order of magnitude. A concentration range of more than two orders of magnitude will most probably violate this assumption [91]. However, it should be considered if the quantitation is really required over the whole concentration range (see discussion on area normalisation, Section 2.5). Non-constant variability (or hetereoscedasticity, or inhomogeneity of variances) can be identified by graphical evaluation [95], or a statistical test, such as the F-test at the upper and lower limit of the range (Eq. 2.3-1), or over the whole range, such as Cochran's or

Bartlett's test (see statistical textbooks or software), or according to Cook and Weisberg [91, 99]. Non-constant variability is easily recognised by a wedge-shaped distribution of residuals, best seen in the case of repeated measurements per concentration level (Fig. 2.4-3A), but also as a concentration dependency of the residuals. In such a case, values with larger variability, i.e., usually the larger concentrations, dominate the unweighted linear regression, because minimising their residuals has much more impact in the overall minimisation than those of smaller concentrations that also have smaller (absolute) residuals. Therefore, these concentrations and data points are more or less ignored, resulting in large deviations from the regression line (Fig. 2.4-9, broken line), especially obvious if relative residuals are plotted (Fig. 2.4-10, squares).

In order to achieve the same representation for all data, the 'weight' of the smaller concentrations must be increased in the regression. This is achieved by using weighting factors in the least-squares regression (Eqs. 2.4-11 to 13). Either the reciprocals of the actual variability (variance or standard deviation), or generalised estimates of the error function, are used [91, 100]. There can either be an individual model of the specific error function (obtained from repeated determinations over the required concentration range) or a suitable approximation may be used taking the respective concentration into account, for example $1/x$ or $1/x^2$. The best weighting scheme can also be experimentally determined by means of minimisation of the

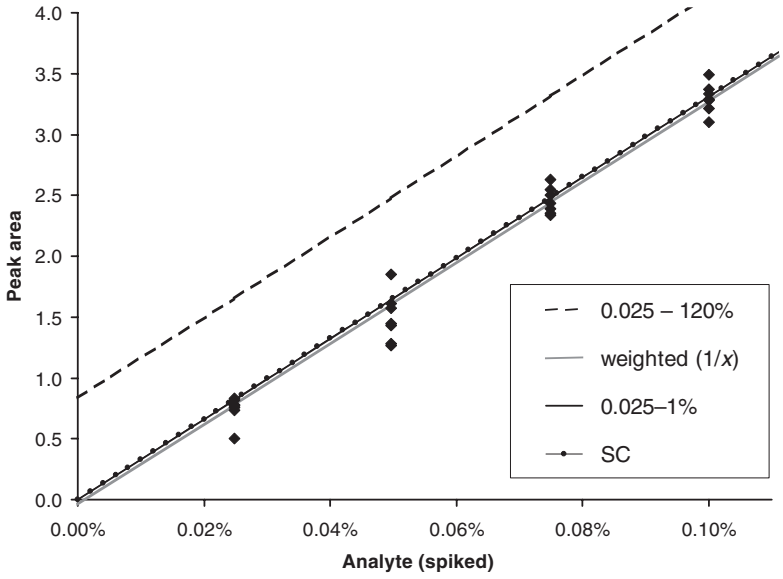


Figure 2.4-9: Linearity data (diamonds) and linear regression lines from 0.025 to 120 % (data from [65]). Only the lower concentration range is shown. The regression lines for unweighted and weighted (weighting factor $1/x$) regression are shown (broken and shaded line, respectively). Additionally, the line corresponding to the single-point calibration (SC) using the data at 100 %, and the line for the regression from 0.025 to 1 % (pearl and solid line, respectively) are shown.

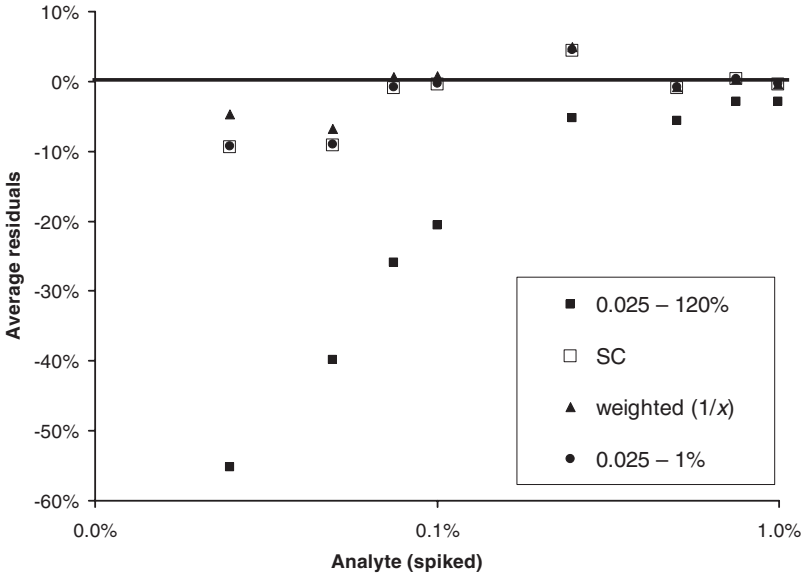


Figure 2.4-10: Residual plots for the regressions shown in Figure 2.4-9. The residuals are calculated with respect to the fitted signal (relative residuals), in order to illustrate the deviation for small concentrations, and only the average per concentration is reported. The concentrations on the x-axis are given in a logarithmic scale to facilitate the inspection of a larger concentration range.

sum of the relative errors [95]. In the example given in Figures 2.4-9 and 2.4-10, it is shown that, applying a weighting factor of 1/x, the lower concentrations are much better fitted to the regression line and therefore provide a better estimation of the intrinsic response function. The same effect can be observed by restricting the unweighted regression range to use only the small concentrations. It is interesting to note that a single-point calibration also provides an equally good fit with respect to the small concentration data. In the concentration range shown in Figure 2.4-9, the weighted regression over the whole concentration range and the single-point calibration, as well as the unweighted regression in the small concentration range alone, are almost identical. Of course, for a single- point calibration, the absence of a constant systematic error, i.e., a negligible intercept must be demonstrated beforehand (see Section 2.4.1.4). For bio-analytical applications, this prerequisite is not likely to be fulfilled, due to varying matrices. However, for large-range applications in pharmaceutical analysis, for example, dissolution testing or impurity determinations, a single-point calibration can be an appropriate choice even in the case of heteroscedasticity.

2.4.3 Non-linear and Other Regression Techniques

If an unacceptable deviation from a linear response function is demonstrated, or can be assumed from the fundamentals of the respective analytical procedure, non-linear regression models must be applied (see statistical textbooks or software). The best fit of a polynomial regression to the experimental data may be tested [28], or – preferably – the intrinsic response model can be fitted, if known. Alternatively, suitable data transformations to achieve a linear function is also possible [1b, 101]. However, transformation will probably lead to rather complex error functions, which must be investigated [91].

If there are indications that the prerequisites for an ordinary least-squares analysis are not fulfilled, such as normal distribution, no outliers, or error-free x-values, other techniques, such as *non-parametric or robust regression* techniques can be applied.

One very straightforward approach makes use of the medians of all possible slopes and intercepts [91, 102]. Here, the median of all possible slopes between one data pair and all the remaining data pairs is calculated. This is repeated for all other data pairs and the median of all medians is the robust estimate of the slope, and therefore is called the 'repeated median' estimator. However, further details and other approaches are beyond the scope of this book and the reader is referred to specialised literature. The same applies to multivariate calibration, where a multitude of response variables are processed simultaneously.

Robust slope: $$b = median\left\{ median\ (j \neq i)\ \left(\frac{y_j - y_i}{x_j - x_i} \right) \right\} \tag{2.4-20}$$

Robust intercept: $$a = median\left\{ median\ (j \neq i)\ \left(\frac{x_j y_i - x_i y_j}{x_j - x_i} \right) \right\} \tag{2.4-21}$$

2.4.4
Key Points

- In the validation characteristic of linearity, the intended calibration model must be justified.
- When there is previous knowledge about the intrinsic (linear) response function (e.g., LC-UV), a verification is sufficient.
- Deviation from the assumed calibration model should be assessed primarily by graphical inspection:
  - Residual plot: random scatter of the residuals vs. concentration (or fitted signal) within an acceptable range around zero.
  - Sensitivity plot (ratio of signal vs. concentration): in the case of a linear function without intercept (y= x), the sensitivities must scatter in an acceptable horizontal range without systematic trends.
- Numerical parameters are only meaningful after verification/proof of the model.
  - The relative residual error represents the dispersion of the data around the regression line. Due to the normalisation, this percentage parameter can easily be compared with an acceptable precision.
  - The coefficient of correlation is neither a proof of linearity, nor a suitable quantitative linearity parameter.
- The prerequisites for an unweighted linear regression, such as error-free x-values, normally distributed y-values, or constant variability over the whole regression range, must be fulfilled (either by reasonable assumption/experience, or by experimental investigation). For UV-detection, constant variability can only be assumed within a ten-fold concentration range.
- If quantitation is required over larger concentration ranges (more than two orders of magnitude), the non-constant variability of the response data must be compensated for using a weighted regression. In the case of a constant matrix and a proven zero intercept, a single-point calibration is also appropriate.
- Statistical linearity tests and non-linear regression models are only recommended to be applied in the case of indication or assumption of deviations from a linear response function.

2.5 Range

Joachim Ermer

ICH

"The range of an analytical procedure is the interval between the upper and lower concentration (amounts) of analyte in the sample (including these concentrations) for which it has been demonstrated that the analytical procedure has a suitable level of precision, accuracy and linearity." [1a]

The required range depends on the application intended for the analytical procedure, see Table 2.5-1.

The working range of an analytical procedure is usually derived from the results of the other validation characteristics. It must include at least the expected or required range of analytical results, the latter being directly linked to the acceptance limits of the specification, or the target test concentration (Table 2.5-1). In the case of other applications, the range can be derived by the same considerations. For example, a water determination with an upper and lower specification limit would require a range of 20 % below and above the limits, as would also be the case in dissolution testing. When there is only an upper limit, the same requirements as for impurities are appropriate, i.e., from the reporting threshhold up to 120 %.

Table 2.5-1 Minimum ranges for different types of analytical procedures [1b].

Analytical procedure	Recommended minimum range
Assay	80–120 % test concentration
Content uniformity	70–130 % test concentration
Dissolution	± 20 % upper/lower specification limit
Impurities	Reporting threshold to 120 % specification limit
in drug substance	Reporting threshold: 0.05 % / 0.03 % (daily intake < 2 g / > 2 g)
in drug product	Reporting threshold: 0.1 % / 0.05 % (daily intake < 1 g / > 1 g)
100 % Standard (area normalisation)	Reporting threshold impurity to 120 % specification limit active

The ICH statement for the 100 % standard method (also called the area normalisation or 100 % method) needs some interpretation. As discussed in Section 2.4.1, an unweighted linear regression is inappropriate to perform over the whole range of more than four orders of magnitude, due to the heteroscedasticity of the data. However, it is not necessary to address the whole range simultaneously. Quantitation is performed in the concentration range of the impurity, i.e., the required range is

from the reporting threshold up to at least 120 % of the impurity specification limit. Because the impurity peak area is (mainly) related to the main peak area, the latter is extrapolated to proportionally smaller concentrations. Therefore, a linear response function and a negligible intercept needs to be demonstrated for the active substance. However, these requirements, corresponding to those for a single-point calibration must be verified under appropriate conditions, as described in the Sections 2.4.1.1 and 2.4.1.4.

In order to avoid extrapolation artefacts in the evaluation of the intercept, it is strongly recommended to extend the investigation of linearity below the minimum required range (see Section 2.4.1.4).

2.6
Detection and Quantitation Limit

Joachim Ermer and Christopher Burgess

Regulatory authorities require impurity profiling of drug substances and drug products as part of the marketing authorization process. The safety requirements are linked to toxicological studies for the active substance itself as well as the impurities of synthesis and degradation. Hence there is a need to demonstrate that impurity profiles are within the ranges examined within the toxicological studies and to limit any degradation products. The purpose of this section is to examine the methods available for determining when an analyte is present (Detection Limit, DL) and for the smallest amount of analyte that can be reliably measured (Quantitation Limit, QL).

ICH

"The detection limit of an individual analytical procedure is the lowest amount of analyte in a sample which can be detected but not necessarily quantitated as an exact value. The quantitation limit of an individual analytical procedure is the lowest concentration of analyte in a sample which can be quantitatively determined with suitable precision and accuracy." [1a]

Various approaches can be applied:

- Visual definition
- Calculation from the signal-to-noise ratio (DL and QL correspond to 3 or 2 and 10 times the noise level, respectively)
- Calculation from the standard deviation of the blank (Eq. 2.6-1)
- Calculation from the calibration line at low concentrations (Eq. 2.6-1)

$DL;\ QL = \frac{F \cdot SD}{b}$ (2.6-1)

F: factor of 3.3 and 10 for DL and QL, respectively

SD: standard deviation of the blank, standard deviation of the ordinate intercept, or residual standard deviation of the linear regression

b: slope of the regression line

The estimated limits should be verified by analysing a suitable number of samples containing the analyte at the corresponding concentrations. The DL or QL and the procedure used for determination, as well as relevant chromatograms, should be reported.

2.6.1 Analytical Detector Responses

The most common type of analysis undertaken in impurity analysis is chromatographic separation.

In most instances, HPLC analytical detectors give a continuous voltage output. In order to be able to compute the peak areas, etc., it is necessary to convert this voltage into a time sequenced discrete (digital) signal that is able to be processed by the Chromatography Data System. In order to perform this conversion an A/D (Analogue to Digital) converter is used (Fig. 2.6-1).

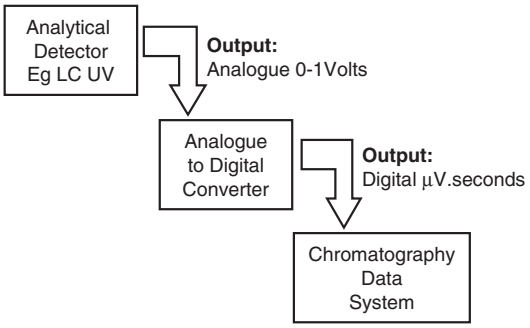


Figure 2.6-1: Typical A/D conversion of chromatography signals.

The resolution of the A/D converter determines the accuracy to which the voltage is represented. For most chromatographic applications the number of bits the A/D converter has is normally in excess of 16 and in fact modern systems use 24 bits. The converter has to be linear over the application range. For more details, see reference [111] and the references contained therein. Even if we are able to achieve perfect A/D conversion, the overall system introduces noise and drift which distort the measurement signal and hence our ability to detect and integrate peaks. The data analysis associated with chromatography is a complex matter and the reader is referred to Dyson [112] and Felinger [113] for an in-depth discussion. We will restrict ourselves here to a brief overview and some practical implications.

2.6.1.1 Noise and Drift

Noise and drift are the bane of the chromatographer's existence. The lower the level of the analyte to be detected or quantified, the worse the problem becomes. The task in essence is simple; find a peak which is not noise. The presence of other effects, such as drift or spiking, make the problem worse. Typical examples encountered in chromatography are shown in Figure 2.6-2.

The most common method of measuring noise is the peak-to-peak method. This is illustrated in Figure 2.6-3. A set of parallel lines are drawn over the time required and the maximum distance measured.

Sometimes the situation is complicated by baseline shifts. This is illustrated in Figure 2.6-4 where it would not be correct to estimate the noise from the highest to

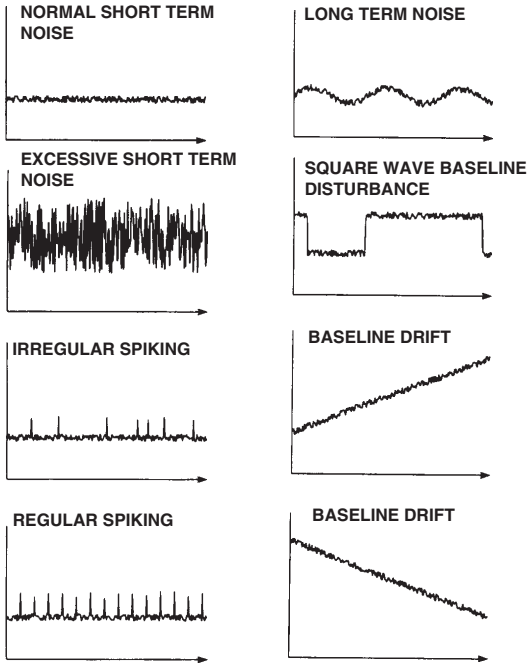


Figure 2.6-2: Noise and drift types (adapted from Table 6.1 [113]).

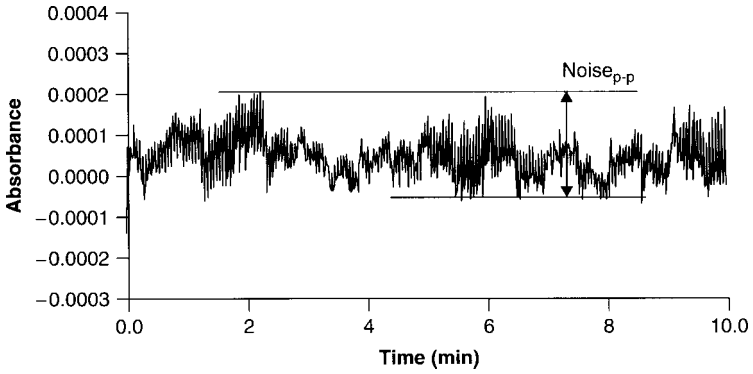


Figure 2.6-3: Peak-to-peak noise (Figure 10.8 [114]).

the lowest as this clearly includes a baseline shift. The better way is to estimate the peak-to-peak noise for each region as illustrated by the dashed lines.

The American Society for Testing and Materials (ASTM) has developed an approach for the measurement of noise and drift for photometric detectors used in HPLC. Short-term noise is defined as that which occurs over a span of half to one minute over a period of 15 minutes, long-term noise over a span ten minutes within a 20 minute period and drift over the course of 60 minutes. The peak-to-peak noise measurements are illustrated in Figure 2.6-5.

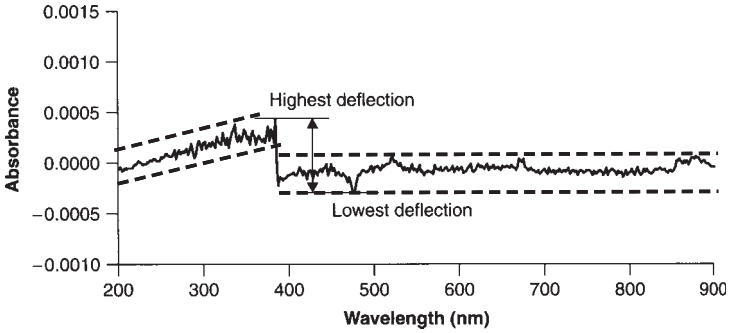


Figure 2.6-4: Baseline shifts and peak-to-peak noise (adapted from Figure 10.9 [114]).

2.6.2 Requirements for DL/QL in Pharmaceutical Impurity Determination

2.6.2.1 Variability of the Actual QL

Figure 2.6-6 shows the results of a repeated experimental QL determination using five different LC systems. Several calculation modes described in this chapter were applied and the investigations were repeated, both on the same LC system and on others over a time interval of about nine months. The experimentally obtained QLs vary within the range of a factor between 2 and 5. As the experimental conditions were rather simple (isocratic elution, dilution of the analyte from a stock solution into the mobile phase), the results of the investigations mainly reflect the instrumental influences. (Therefore, an acceptance limit of 10 % relative standard deviation was defined to estimate QL from precision.) Under authentic conditions, i.e., the analyte (impurity) in a complex matrix of the active, other impurities, and placebo (in the case of drug products), additional variability can be assumed. With respect to the calculation modes, there are (minor) differences according to their fundamentals, which will be discussed in the next sections. Therefore, before reporting the DL/QL, the calculation mode always needs to be specified and referred to in sufficient detail. From 30 validation papers reviewed, dealing with DL/QL and published between 1995 and 2003, five were deficient in this respect.

However, even applying the same calculation, a high variability in the actual QL result must be considered. This is crucial, because in pharmaceutical analysis, *fixed acceptance limits* for impurities [1c–e] are required, and the analytical procedure needs to be able to quantify reliably in all future applications. This is especially important for long-term applications such as stability studies, or in the case when different equipment is used or methods are transferred to other laboratories. As a consequence, the QL of the analytical procedure has the character of a general parameter.

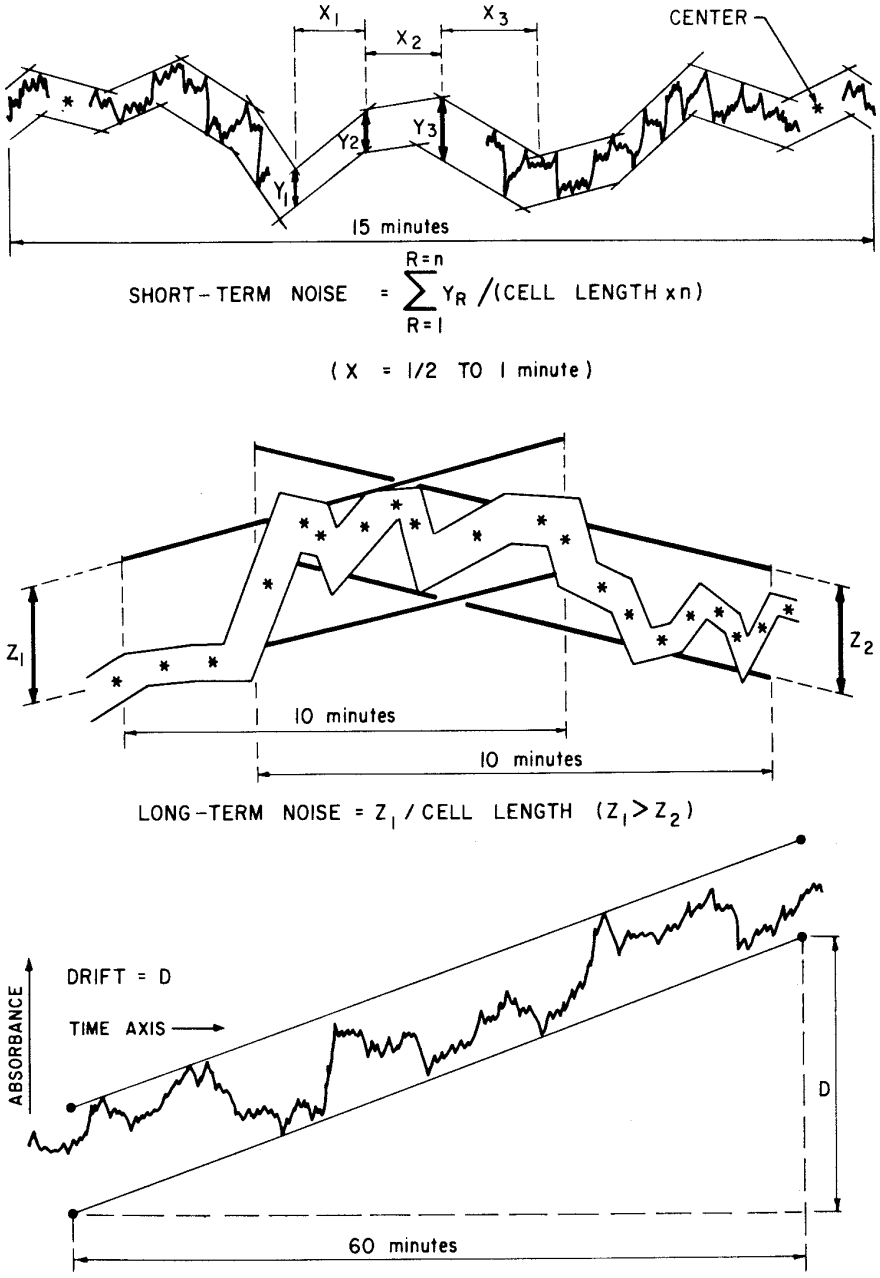


Figure 2.6-5: ASTM noise and drift measurements for LC UV detectors ([115]).

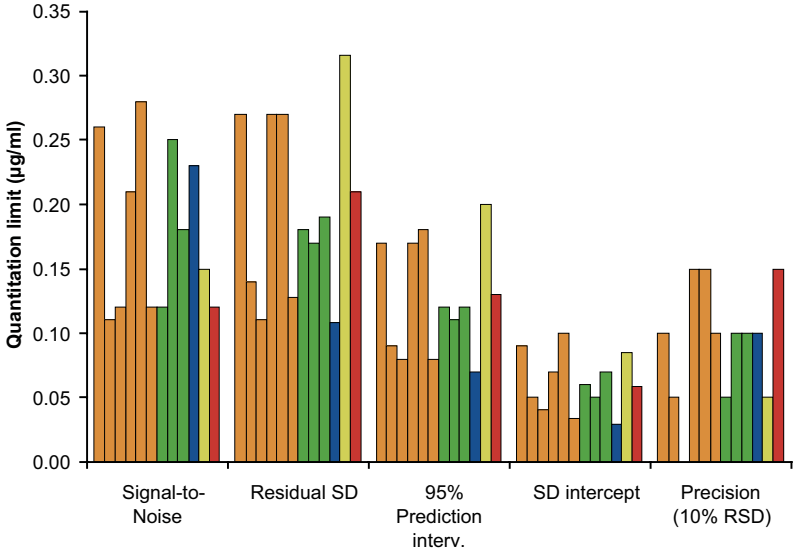


Figure 2.6-6: Intermediate QL study using five LC systems with one to six repetitions per system over nine months. The columns of the same colour illustrate the same LC-system used, and their sequence the repetition, i.e., the first orange column for each calculation mode corresponds to the first QL-study on LC-system 1, the second column to the second series on system 1, etc.

2.6.2.2 General Quantitation Limit

Statistically, the *general QL* can be regarded as the upper limit of the distribution of all individual QLs. Thus, one method of obtaining a reliable result is to perform an 'intermediate QL' study. In such an investigation, as for precision, all factors likely to vary in the future routine application should be included, for example, different reagents, equipment, analysts, etc. Depending on the number of repeated determinations, the upper limit may either be defined as the largest experimental result, or (based on at least six QL determinations, in order to ensure sufficient reliability) calculated from the mean result and the standard deviation (Eq. 2.6-2). For the study shown in Figure 2.6-6, the result for $QL_{general}$ from the residual standard deviation of the regression line is 0.43 µg/ml, the largest individual QL was 0.32 µg/ml.

$$QL_{general} = \bar{Q}\bar{L} + 3.3s_{QL} \tag{2.6-2}$$

$\bar{Q}\bar{L}$ mean of all individual QL

s_{QL} standard deviation of all (at least six) individual QL.

The approach just described is rather extensive and should only be followed in cases where the aim is quantitation of impurities as low as (reliably) possible and justified. For all other cases, it is recommended to begin with the requirements. The ICH guidelines define reporting thresholds for unknown related substances [1c,d] (Table 2.6-1). These reporting thresholds can be regarded as the minimum require-

ments for quantitation and therefore can be directly used as a '*general QL*'. From Table 2.6-1, it is also obvious that the reporting thresholds correspond to 50 % of the respective specification acceptance limit. This relationship can also be applied – as a minimum requirement – for specified impurities and degradants, such as for residual solvents limited according to [1e] or for cleaning validation methods (specific residual cleaning limit, see Section 2.3.4).

However, if technically feasible, the thresholds for unknown related substances should also be used as the general QL of specified related impurities, both from the perspective of the 'analytical state of the art' as well as for consistent reporting in batch release and stability and for reasons of practicability. Of course, the 50 % requirement is also valid if specification acceptance limits need to established at lower levels than usual, for example, for safety reasons.

Table 2.6-1 Thresholds for unknown impurities according to ICH [1c,d]

	Maximum daily dose	Reporting threshold (%)[a]
Drug substance	≤ 2 g	0.05
	> 2 g	0.03
Drug product	≤ 1 g	0.10
	> 1 g	0.05

a) response with respect to active, e.g., area percentage

If it is necessary to go to the (performance) limits of the analytical procedure, the QL can be specifically calculated using the actual precision of the analytical procedure at this concentration. The calculation is based on the compatibility between analytical variability and specification acceptance limits, as described in Section 6.3. QL can be regarded as the maximum true impurity content of the manufactured batch (Fig. 2.6-7), i.e,. as the basic limit in Eq. (6-12). Rearranging, leads to Eq. (2.6-3).

$$QL_{general} = AL - \frac{(st_{df,95\%})_{validation}}{\sqrt{n_{assay}}} \qquad (2.6\text{-}3)$$

AL: Acceptance limit of the specification for the impurity.

s Precision standard deviation at QL, preferably under intermediate or reproducibility conditions. AL and s must have the same unit (e.g. percentage with respect to active, µg, µg/ml, etc.)

n_{assay}: Number of repeated, independent determinations in routine analyses, as far as the mean is the reportable result (see Chapter 10), i.e., is compared to the acceptance limits. If each individual determination is defined as the reportable result, n=1 has to be used.

t_{df}: Student t-factor for the degrees of freedom during determination of the precision, usually at 95 % level of statistical confidence.

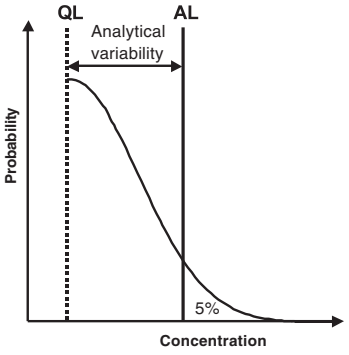


Figure 2.6-7: The acceptance limits of the specification (AL) must include both analytical and manufacturing variability. The (true) quantitation limit (QL) represents the maximum manufacturing variability with respect to impurity content. The analytical variability can be described by a 95 % prediction interval (for details, see Section 6.3).

It is also possible to apply a combination approach. For example, if the analyst is confident that the analytical procedure is capable of reliably quantifying very small amounts of the respective impurity (say less than 0.01 %), the general QL can be defined as 0.02 % with only limited experimental confirmation. Any value of the QL found to be below the requirements is therefore scientifically justified.

Once the general QL is established as part of the validation effort, it is then only necessary to verify that the actual QL is below the defined limit [16], regardless of how far below.

2.6.3 Approaches Based on the Blank

There are two different approaches which have been used to derive practical estimations of DL and QL from the blank. The first is based on a simple measurement of the *signal-to-noise ratio* of a peak using the peak-to-peak approach. A test sample with the analyte at the level at which detection is required or determined is chromatographed over a period of time equivalent to 20 times the peak width at half-height. The signal-to-noise ratio is calculated from Eq. (2.6-4).

$$S/N = \frac{2H}{h} \tag{2.6-4}$$

H is the height of the peak, corresponding to the component concerned, in the chromatogram obtained with the prescribed reference solution, and measured from the maximum of the peak to the extrapolated baseline of the signal observed over a distance equal to 20 times the width at half-height

h is the peak-to-peak background noise in a chromatogram obtained after injection or application of a blank, observed over a distance equal to 20 times the width at half-height of the peak in the chromatogram obtained.

This approach is specified in the European Pharmacopoeia [15]. It is important that the system is free from significant baseline drift and/or shifts during this determination.

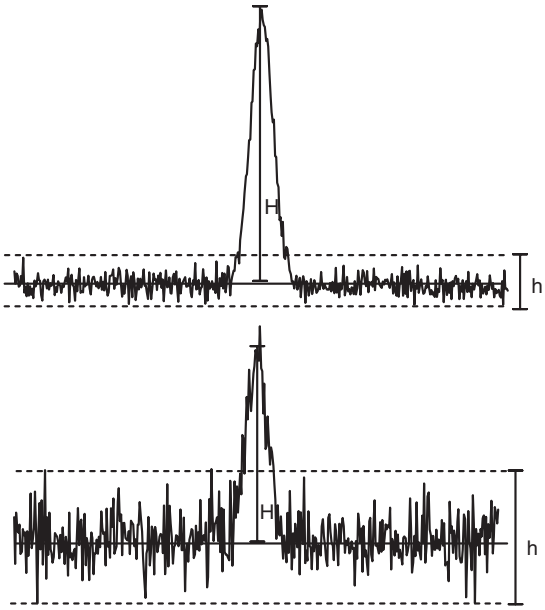


Figure 2.6-8: Signal-to-noise examples of 10:1 (top) and 3:1 (bottom), using the method of the EP [15].

Figure 2.6-8 shows examples of *S/N* ratios of 10: 1 and 3:1 which approximate the requirements for the QL and DL, respectively. This approach works only for peak height measurements.

For *peak area measurements*, the standard deviation of the blank must be considered. The statistical basis on which the DL is defined is shown graphically in Figure 2.6-9. The dashed curve represents the distribution of the blank values and the solid line that of the analyte to be detected. It is assumed that they both have the same variance and are normally distributed. As the curves overlap there is a probability that we could conclude that we have detected the analyte when this is in fact due to the blank signal (false positive, α error or type 1 error). Alternatively, we can conclude that the analyte is not detected when it is in fact present (false negative, β error or type 2 error). When addressing the issue about when an analyte has been detected it is always a matter of risk. In some analytical techniques, particularly atomic spec-

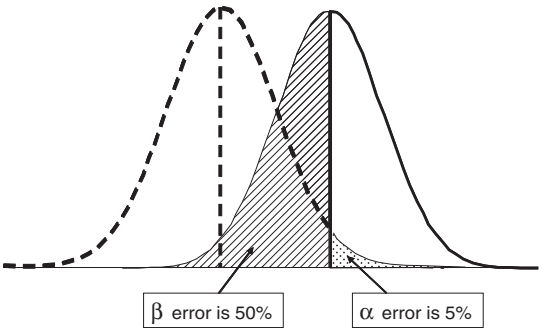


Figure 2.6-9: Statistical basis for the detection limit.

troscopy, this is defined as when there is an even chance of a false negative, i.e., a 50 % β error. This is illustrated in Figure 2.6-9. Note, however, that there is also a false positive risk in this situation of 5 % (α error).

In ICH, the detection limit and quantitation limits are described in similar terms but with a different risk basis. They define the DL and QL as multiples of the standard deviation of the blank noise (Eq. 2.6-1). These multiples are 3.3 for the DL and 10 for the QL. This is illustrated in Figure 2.6-10.

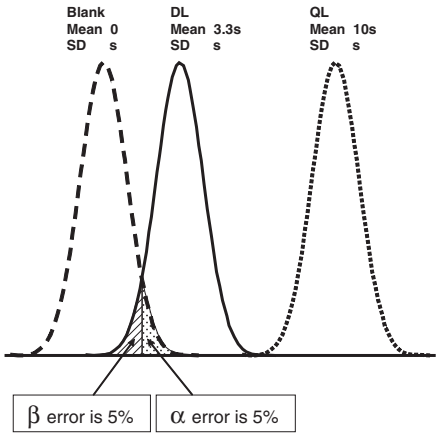


Figure 2.6-10: Statistical basis for the ICH detection limit and quantitation limits (DL and QL).

Here we can see that although the false positive error is still 5 % it is balanced by the same false negative error for the confidence of the DL. The choice of a factor of 10 for the QL is arbitrary but it demonstrates that the possibilities of either α and β errors are very small indeed.

2.6.4
Determination of DL/QL from Linearity

These approaches are based on parameters of an unweighted linear regression using low analyte concentrations. Therefore, all requirements for an unweighted linear regression must be fulfilled, i.e., the homogeneity of variances and a linear response function (see also Section 2.4). This is imperative for DL/QL calculations, because here regression parameters are used that describe the scattering (dispersion) of the analytical results. As described in Section 2.4.2, (too) high concentrations with large responses would dominate these parameters and lead to incorrectly large DL/QL (Fig. 2.6-11). Obviously, for DL/QL, the data variability at very low concentrations is relevant. As a rule of thumb, for LC–UV, the concentration range used for the calibration line should not exceed the 10–20 fold of DL [103]. In this range, the increase of the variances can usually be assumed to have minor influence on the dispersion parameters of an unweighted linear regression (see Fig. 2.6-11). Otherwise, it needs to be verified experimentally, for example, by means of the F-test (at

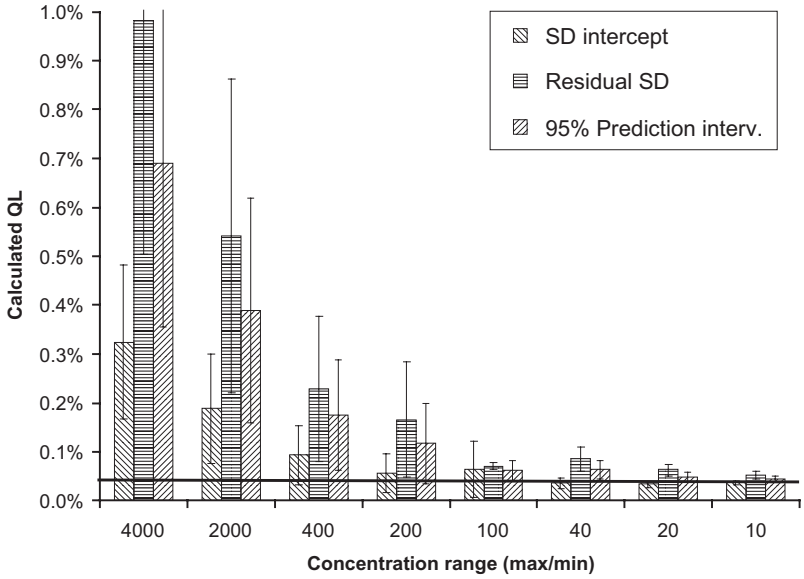


Figure 2.6-11: Influence of the concentration range used for linear regression on the QL for various calculation modes. The concentration range is represented as the ratio between the largest concentration and the smallest concentration of 0.025 %. In order to have a reliable reference (true QL), as well as taking the result variability into account, simulated data based on the experimental standard deviation curve given in Figure 2.1-6 were used. Between eight and twelve data were generated for each concentration range. The columns represent the average QL of six simulations, the bars the standard deviations. The (true) standard deviation of the blank was estimated from the pooled results of the four smallest concentrations (0.025 – 0.1 %) to 0.14. The true QL was calculated as ten times the standard deviation of the blank (ICH definition), corresponding to 0.042 % and is indicated as a horizontal line.

the upper and lower limit of the range, Eq. 2.3-1), or the Cochran test over the whole range, see statistical textbooks, for example [116]).

Selecting too large a range for linearity data to calculate the DL/QL is a frequent mistake in validation literature. From 30 validation papers reviewed dealing with DL/QL and published between 1995 and 2003, eight of them obtained DL/QL from linearity measurements. Six of these studies, i.e,. 75 % (!) used an inappropriate concentration range, with the ratio between the minimum and maximum concentration of up to 2000!

There are some proposals which avoid the problem of inhomogeneous variances by using weighted linear regression [104–107]. However, this cannot really solve the problem, because due to the increased weight of smaller concentrations (see Section 2.4.2), the larger ones are more or less neglected in the calculated dispersion parameters. Therefore, the QL calculation result obtained is not very different from the one using the small concentrations only, provided that the number of determinations is still large enough (Fig. 2.6-12, A(w) vs. A(uw)). However, as soon as the lowest concentration is not in the vicinity of the QL, the calculated values from a

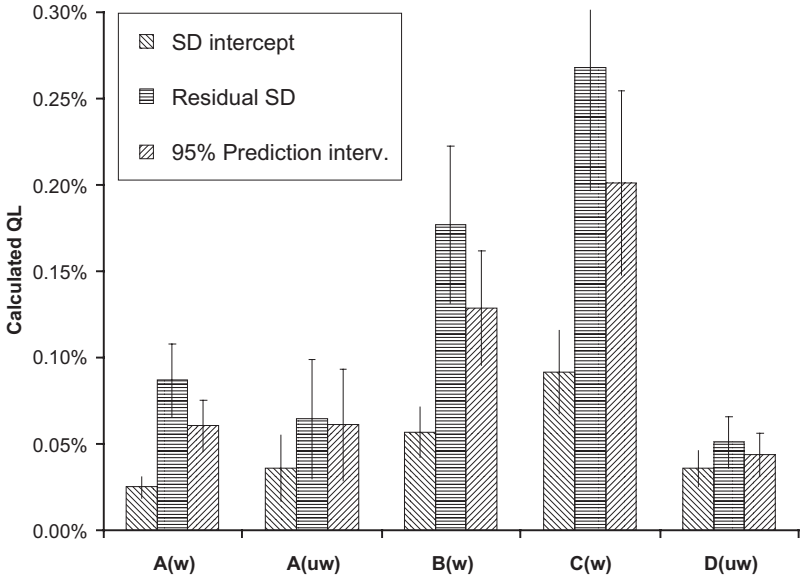


Figure 2.6-12: Comparison between weighted and unweighted regression for calculation of QL. The data were simulated as described in Fig. 2.6-11. The columns represent the mean result of six simulations, the bars the standard deviations. A(w): the concentration range 0.025–100 %, $n = 12$, weighting factor $1/x$ A(uw): the (lower) concentration range 0.025–1 %, $n = 6$, the unweighted regression B(w): concentration range 0.1–100 %, $n = 10$, weighting factor $1/x$ C(w): concentration range 0.25–100 %, $n = 9$, weighting factor $1/x$ D(uw): concentration range 0.025–0.25 %, $n = 8$, unweighted regression

weighted regression are also biased (Fig. 2.6-12, B(w) and C(w)). Therefore, extrapolation must be strictly avoided!

The dependence of the calculated QL on the number of data used is shown in Figure 2.6-13. As is to be expected, a larger number of data increases the reliability of the dispersion parameter, and consequently of the calculated QL, to a different extent dependent on the calculation mode (see next sections). Generally, a minimum of about eight concentrations is recommended.

2.6.4.1 **Standard Deviation of the Response**

One option according to ICH [1b] (Eq. 2.6-1) is to use the *residual standard deviation* of the regression. This parameter describes the scattering of the experimental data around the regression line and can thus be regarded as a measure of the variability. Dividing the standard deviation by the slope converts the response (signal) into the corresponding concentration. The factors of 3.3 and 10 for DL and QL, respectively, are again used to discriminate between the distributions of blank and analyte. This calculation results in a slight overestimation of QL (Fig. 2.6-13), probably due to the impact of the higher concentrations on the residual standard deviation (even in the present example of an only ten-fold range). The approach is less sensitive to small data numbers (Fig. 2.6-13), but largely influenced by too large a concentration range (Fig. 2.6-11).

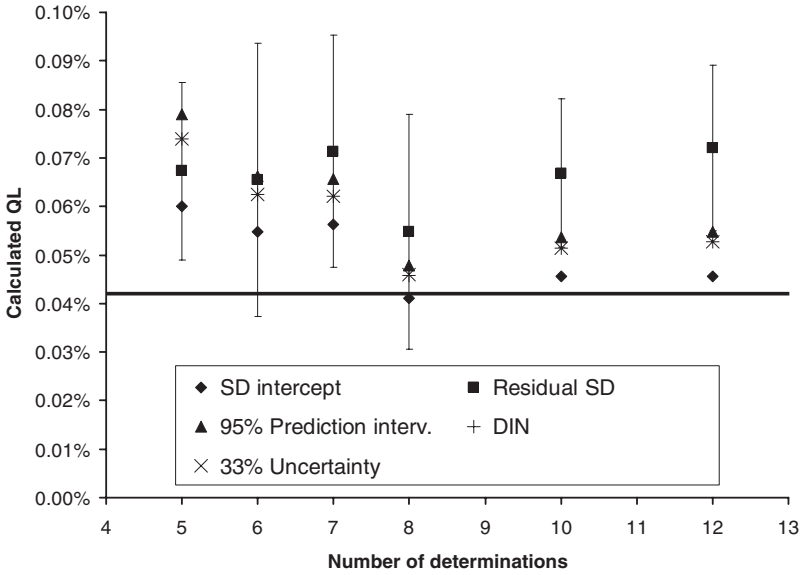


Figure 2.6-13: Dependence of QL on the number of data and the calculation mode. Six sets of data with the respective numbers each were simulated as described in Fig. 2.6-11, for a concentration range 0.025–0.25 %. The average QL for the various calculation modes is symbolised. The standard deviations of the six QLs, for the calculation from the residual standard deviation, are indicated by error bars; for the other approaches, similar variabilities were obtained. The true ICH- based QL value is given by the horizontal line.

The *standard deviation of the intercept* can be regarded as an extrapolated variability of the blank determination. The QL calculated in such a way are substantially lower than those obtained from the residual standard deviation of the regression. This behaviour can be explained from the respective equations (Eq. 2.4-7 and 2.4-2). The two parameters are directly correlated, with the following ratio:

$$\frac{s_{int\ ercept}}{s_{res}} = \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{Q_{xx}}} \tag{2.6-5}$$

Therefore, the ratio is only dependent on the number of values used for the regression and the concentration range (x-values). Under conditions usually applied, the QL calculated from the standard deviation of the intercept will be lower by a factor of between 0.8 and 0.5, compared with the calculation from the residual standard deviation.

The standard deviation of the intercept seems to be a good approximation of the true blank variability, as is obvious from Figure 2.6-13. In the simulated examples, the QLs calculated in this way are nearest to the true value, provided that a sufficient number of determinations are available. This calculation mode is also less sensitive towards non-optimal concentration ranges (Fig. 2.6-11), because the increased dispersion parameter is partly compensated for by the decreased ratio according to Eq. (2.6-5).

2.6.4.2 95 % Prediction Interval of the Regression Line

The prediction interval of the regression line is a measure of the variability of the experimental determination. This interval can be interpreted as the probability distribution of future determinations that can be experimentally expected (see Section 2.4.1, Eq. 2.4-15). As illustrated in Figure 2.6-9, DL and QL can be defined by different degrees of overlapping of their probability distribution with that of the blank. The upper limit of the analyte concentration, whose probability distribution has a 50 % overlap with the distribution of the blank (i.e. a ß-error of 50 %) is defined as the detection limit (Fig. 2.6-14, DL). With respect to the quantitation limit the overlapping is reduced to 5 %, guaranteeing a reliable quantification (Fig. 2.6-14, QL) [103]. The difference in the calculation approach using the standard deviation of the blank (Fig. 2.6-10) is that experimental results from small analyte concentrations are used for the regression, not from the blank alone, and that the prediction interval describes the probability of future determinations. Therefore, a larger uncertainty is included, resulting in different risk assumptions.

Figure 2.6-14 illustrates the graphical derivation of DL and QL from the 95 % prediction intervals, their numerical calculation is given in Eqs (2.6-6) and (2.6-7).

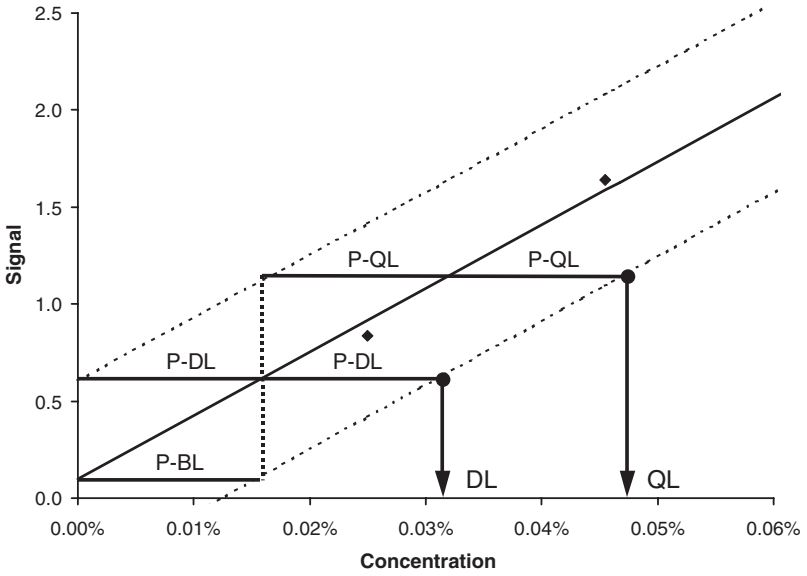


Figure 2.6-14: Utilisation of the 95 % prediction intervals (dotted lines) of an unweighted linear regression (solid line) to obtain DL and QL (according to [103]). The horizontal lines indicate the 95 % width of the probability distributions of the (upper half of the) blank (P-BL), the detection limit (P-DL), and the quantitation limit (P-QL).

Calculation of DL from 95 % prediction interval

$$y_c = t(P, n-2)\, s_y \sqrt{\frac{1}{n} + \frac{1}{m} + \frac{\bar{x}^2}{Q_{xx}}}$$

$$DL = \frac{2\ t(P,n-2)\ s_y}{b} \sqrt{\frac{1}{n} + \frac{1}{m} + \frac{(a+y_c-\bar{y})^2}{b^2\ Q_{xx}}} \tag{2.6-6}$$

Calculation of QL from 95 % prediction interval:

$$y_h = a + 2\ t(P, n-2)\ s_y \sqrt{\frac{1}{n} + \frac{1}{m} + \frac{\left(\frac{y_c}{b} - \bar{x}\right)^2}{Q_{xx}}}$$

$$QL = \frac{y_h - a}{b} + \frac{t(P,n-2)\ s_y}{b} \sqrt{\frac{1}{n} + \frac{1}{m} + \frac{(y_h-\bar{y})^2}{b^2\ Q_{xx}}} \tag{2.6-7}$$

This calculation mode is moderately sensitive to large regression ranges (Fig. 2.6-11), but requires a sufficient number of determinations (about eight, Fig. 2.6-13). It results in a slight overestimation of QL, compared to the true value, probably also due to the influence of the higher concentrations.

2.6.4.3 Using the German Standard DIN 32645

This approach [108] is also based on the variability of the concentration dependent experimental determination (Eqs. 2.6-8 to 2.6-10). Above the detection limit, a statistical decision is possible that the analyte content in the sample is higher than in the blank (i.e., the presence of the analyte can be proved qualitatively). A *recording limit* is (additionally) defined as the lowest content of analyte which can be detected with a certain degree of probability. Assuming the same probability of errors types α and β, the recording limit corresponds to twice the detection limit.

The quantitation limit according to DIN is calculated from the recording limit using a level of uncertainty (see Section 2.6.4.4), which can be individually defined according to the requirements of the analytical procedure. The factor k_f used in Eq. (2.6-10) corresponds to the reciprocal of the relative uncertainty, i.e., a factor of 3 (that is usually applied) corresponds to an uncertainty of 33.3 %. The factor must be chosen to obtain a quantitation limit that is larger than the recording limit.

The results obtained are very similar to those from the 95 % prediction interval, especially for a higher number of determinations (Fig. 2.6-13).

Detection limit (DIN 32645):

$$x_{NG} = \frac{t(P,n-2)_{\ one\ sided}\ s_y}{b} \sqrt{\frac{1}{m} + \frac{1}{n} + \frac{(\bar{x})^2}{Q_{xx}}} \tag{2.6-8}$$

Recording limit (DIN 32645): $x_{EG} = 2\ x_{NG}$ (2.6-9)

Quantitation limit (DIN 32645):

$$x_{BG} = \frac{k_f t(P,n-2)_{two\ sided}\ s_y}{b} \sqrt{\frac{1}{m} + \frac{1}{n} + \frac{(x_{BG} - \bar{x})^2}{Q_{xx}}} \tag{2.6-10}$$

2.6.4.4 From the Relative Uncertainty

The uncertainty of a given concentration corresponds to half the horizontal prediction interval around the regression line at this concentration (Δx_i, Fig. 2.4-2, Eq. 2.4-16). A repeated analysis is expected (at the defined level of statistical significance) to give a value anywhere in this range. The relative uncertainty is the ratio between Δx_i and the corresponding concentration x_i. Apart from extrapolations, the prediction interval is only slightly curved, and consequently Δx_i increases only slightly towards the upper and lower limits of the regression range. (It is narrowest in the centroid, i.e., at the average of all concentrations used for the regression.) When the concentration x_i decreases towards zero, the relative uncertainty grows exponentially (Fig. 2.6-15). As this value represents a direct measure of the reliability of a determination at the corresponding concentration, DL and QL can be directly calculated by defining an *acceptable relative uncertainty* each, for example, 50 % and 33 %, respectively (Eq. 2.6-11). Applying the same relative uncertainty, this approach and that of DIN (see Section 2.6.4.3) result in identical QL (up to the third significant figure, see Fig. 2.6-13). However, using the relative uncertainty directly has the advantage of a straightforward and easily comprehensible approach.

$$A = \frac{1}{n} + \frac{1}{m} + \frac{\bar{x}^2}{Q_{xx}} \qquad B = \frac{\bar{x}}{Q_{xx}} \qquad C = \frac{1}{Q_{xx}} - \left(\frac{\Delta \cdot b}{100 t(P,n-2) s_y}\right)^2$$

$$DL/QL = \frac{B}{C} \pm \sqrt{\frac{B^2}{C^2} - \frac{A}{C}} \tag{2.6-11}$$

Δ = the acceptable relative uncertainty for DL or QL (in %)
The smallest positive solution of the equation corresponds to DL/QL

What is the relationship between the relative uncertainty of a given concentration and the relative standard deviation of its signal(s)? The latter describes the variability in the past, corresponding to an interval around the mean which includes 68 % of (the normally distributed) signal values (vertically). The relative uncertainty is based on the prediction interval, which indicates the variability of experimental values expected in the future (at the given level of statistical confidence, e.g., 95 %), calculated with respect to concentrations (horizontal prediction intervals). Therefore, the number of data and the distance from the mean concentration (leverage, see Section 2.4.1) have an effect, but the main contribution comes from the student t-value. Therefore, the relative uncertainty can be estimated to be larger for a factor of about three compared with the signal precision at the given concentration.

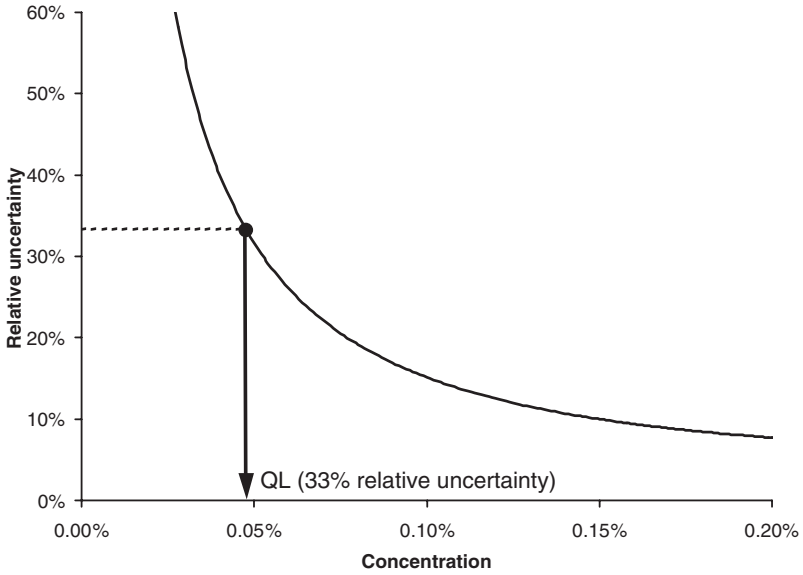


Figure 2.6-15 Relative uncertainty as a function of the concentration. In dependence on the regression range, the uncertainty function may display a minimum, i.e., not every uncertainty value has a corresponding concentration.

2.6.5
Precision-based Approaches

The quantitation limit can also be obtained from precision studies. For this approach, decreasing analyte concentrations are analysed repeatedly and the relative standard deviation is plotted against the corresponding concentration (precision function). If a pre-defined limit is exceeded (such as 10 or 20 %), the corresponding concentration is established as the quantitation limit [7, 109]. In the literature describing this approach, often a nice continuous increase in the variability with decreasing concentration is envisaged (such as the broken line in Fig. 2.6-16). However, in practice, due to the high variability of standard deviations (see 2.1.1.1), the true precision function is much more difficult to draw (see individual series in Fig. 2.6-16), unless a large number of concentrations is included. It should also be noted, that the average precision curve represents the true variability for a given concentration, whereas the individually obtained results scatter in a much larger range, for example, at 0.05 µg/ml from 5 to 25 %, with an average of about 15 %.

However, as discussed in Section 2.6.2, it is often not necessary to establish the intrinsic QL of the analytical procedure. Defining a general QL from the requirements and an acceptable precision for the quantitation, only the precision at a concentration corresponding to QL needs to be performed. Any result below the acceptance limit will suffice to demonstrate the suitability of the procedure. However,

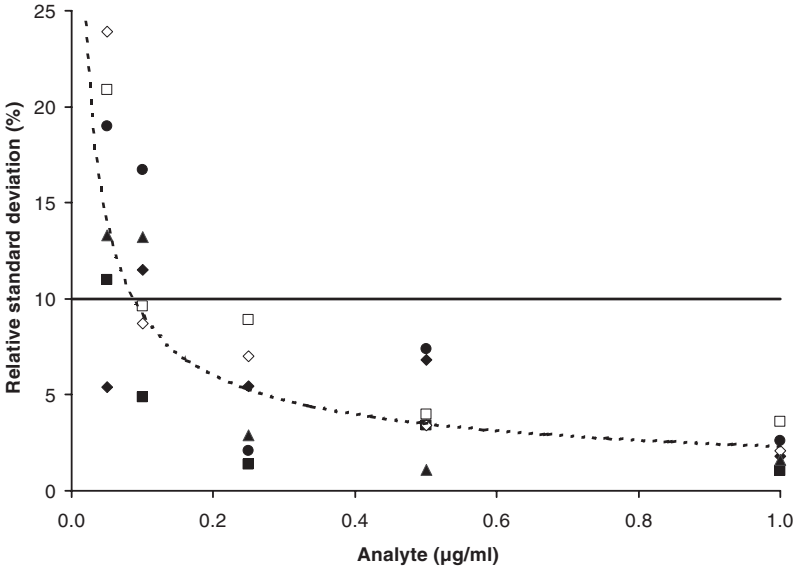


Figure 2.6-16: Repeated series of precision for decreasing concentrations. Six determinations were performed for each concentration. The estimation of the average precision function is illustrated by the dotted line, the defined precision acceptance limit by a solid line. Note that, for an acceptable individual precision (in the case of a defined general QL), a larger limit would be required, e.g., at 20 % RSD.

these acceptance limits must correspond to the upper distribution limit of individual precisions at the given concentration (Fig. 2.6-16).

2.6.6 Comparison of the Various Approaches

In order to obtain practically relevant DL and QL, available impurities and degradants should be spiked to the drug substance or drug product and the quantitation procedure described in the control test should be applied. The concentration of the active and/or matrix components (placebo, cleaning solutions, swab interferences, see Section 2.3.4) should be maintained at the nominal level of the test. The QL of unknown substances can be obtained using representative peaks or inferred from the QL of known impurities/degradants. If the required range is not too large (see Section 2.6.4, Fig. 2.6-11), the spiked samples can be used to validate accuracy, linearity, and DL/QL together.

If properly applied, all the described approaches lead to comparable results, taking into account the variability range to be expected for the low concentration range. The calculations from the 95 % prediction interval, according to the German Standard DIN 3265, and from the relative uncertainty, lead to almost identical results (Fig. 2.6-13), because they are all based on the prediction interval of the regression line. They are slightly higher than QL values calculated from the standard deviation

of the intercept. The latter seems to agree best with the theoretical QL (Fig. 2.6-13). However, this may also be a consequence of the definition of the true QL as tenfold the blank standard deviation. From a practical perspective, the differences are not large and the other calculation modes lead always to an overestimation of QL. The calculation from the residual standard deviation of the regression is less sensitive to a small number of data (Fig. 2.6-13). All approaches based on linearity are more exposed to possible mistakes, if the analyst is not aware of an appropriately limited concentration range (Fig. 2.6-11). The signal-to-noise ratio may be prone to subjectivity [110], but this can be limited by strictly defined conditions. The intrinsically most robust approach is probably the precision-based one, which is required anyway to verify the calculated or defined QL [1b].

Therefore, the choice can be made from the perspective of the most pragmatic approach. For example, if the analytical method is the same for assay and impurities and the batch used for precision investigations contain impurities at the defined (required) QL, the same experimental runs can be used for precision and QL. Of course, in some cases, practical restrictions will be faced. If no spiking is possible or if no sufficiently impurity-free matrix is available, QL can only be obtained from the signal-to-noise ratio, as the other approaches cannot be applied.

As a consequence of the high variability of the experimental QL, their validity should be routinely confirmed within the system suitability test [15], for example, from the signal-to-noise ratio or the system precision of a representative impurity peak.

2.6.7
Key Points

- The calculation mode always needs to be specified and referred to in sufficient detail.
- The high variability in the actual QL determination must be considered.
- Fixed acceptance limits for impurities are required (in pharmaceutical analysis), therefore a 'general QL' should be established, from the requirements or sufficiently reliable experimental determinations.
- For a practically relevant QL determination, impurities should be spiked to the drug substance or drug product, i. e., the respective matrix.
- If properly applied, all QL approaches lead to comparable and correct results, therefore, the most pragmatic approach can be chosen.
- If obtained from linearity investigations, avoid too large concentration ranges (> ten to 20 fold) and extrapolations, and use a sufficient number of determinations (at least eight).
- The validity of the established QL should be routinely confirmed within the system suitability test.

2.7 Robustness

Gerd Kleinschmidt

Although robustness of analytical procedures is generally noticed least of all, it is one of the most important validation parameters. Fortunately, in pharmaceutical analysis more and more attention is paid to it. Basically, robustness testing means to evaluate the ability of a method to perform effectively in a typical laboratory environment and with acceptable variations. Robustness definitions have been widely harmonised among international drug authorities, which is mainly the merit of the International Conference on Harmonisation (ICH).

2.7.1 Terminology and Definitions

Definitions provided by regulatory bodies, which play a significant role in the pharmaceutical world are itemised below.

2.7.1.1 International Conference on Harmonisation (ICH)

According to ICH Q2A [1a] *"the robustness of an analytical procedure is a measure of its capacity to remain unaffected by small, but deliberate variations in method parameters and provides an indication of its reliability during normal usage"*.

Furthermore, it is stated in ICH Q2B [1b], *"The evaluation of robustness should be considered during the development phase and depends on the type of procedure under study. It should show the reliability of an analysis with respect to deliberate variations in method parameters. If measurements are susceptible to variations in analytical conditions, the analytical conditions should be suitably controlled or a precautionary statement should be included in the procedure. One consequence of the evaluation of robustness should be that a series of system suitability parameters (e.g., resolution test) is established to ensure that the validity of the analytical procedure is maintained whenever used"*.

Additionally, the ICH guideline Q2B lists examples of typical variations such as extraction time or in case of liquid chromatography the mobile phase pH, the mobile phase composition and flow rate etc.

Even though these explanations are not very detailed, they guide an analyst on when and how to evaluate robustness. To decide what is small, but deliberate depends on the method and is the responsibility of the analyst.

2.7.1.2 Food and Drug Administration (FDA)

The FDA utilises the ICH definition for robustness and remarks that *"data obtained from studies for robustness, though not usually submitted, are recommended to be included as part of method validation"*. This is stated in the Reviewer Guidance "Validation of Chromatographic Methods" [3].

Corresponding to ICH, robustness testing *"should be performed during development of the analytical procedure and the data discussed and / or submitted. In cases where an*

effect is observed, a representative instrument output (e.g., chromatograms) should be submitted", which is explained in the Guidance for Industry document "Analytical Procedures and Methods Validation" [4].

2.7.1.3 **European Pharmacopoeia (EP)**

The European Pharmacopoeia [117] does not comprise a general chapter on validation of analytical procedures, but in chapter 2.6.21 there is reference to ICH guideline Q2B and it is recommended to evaluate the robustness of nucleic acid amplification analytical procedures.

2.7.1.4 **Japanese Pharmacopoeia (JP)**

The Japanese Pharmacopoeia [118] provides a chapter on validation of analytical procedures within which the ICH terms and definitions for the validation parameters are used. In the case of robustness it is set out that *"the stability of observed values may be studied by changing various analytical conditions within suitable ranges including pH values of solutions, reaction temperature, reaction time or amount of reagent added. When observed values are unstable, the analytical procedure should be improved. Results on studying robustness may be reflected in the developed analytical procedure as precautions or significant digits describing analytical conditions"*.

A point of interest is the clear statement that an analytical procedure should be improved when observed values are unstable. Such a statement cannot be found in any of the documents and monographs mentioned in this chapter, although the improvement of an analytical method should always be paramount and should be performed before precautions or significant digits describing analytical conditions form part of the procedure.

2.7.1.5 **United States Pharmacopoeia (USP)**

Definition of robustness in the United States Pharmacopoeia [5] corresponds to that given in the ICH guidelines. But apart from robustness a further parameter is defined, which is called ruggedness. *"The ruggedness of an analytical method is the degree of reproducibility of test results obtained by the analysis of the same samples under a variety of conditions, such as different laboratories, different analysts, different instruments, different days, etc. Ruggedness is normally expressed as the lack of influence on test results of operational and environmental variables of the analytical method. Ruggedness is a measure of reproducibility of test results under the variation in conditions normally expected from laboratory to laboratory and from analyst to analyst"*.

According to USP, ruggedness is determined by analysis of aliquots from homogeneous batches in different laboratories, by different analysts, using operational and environmental conditions prescribed for the assay. The degree of reproducibility is then evaluated by comparison of the results obtained under varied conditions with those under standard conditions.

2.7.2 Fundamentals of Robustness Testing

From the aforementioned definitions and explanations it follows that, due to the successful ICH process, the main regulatory bodies have the same understanding of robustness and robustness testing. Furthermore, a comparison of the various documents reveals that, in addition to the term robustness, the USP defines and explains the term ruggedness. Although, there is a clear difference between robustness and ruggedness with regard to the parameters usually changed to evaluate them, the term ruggedness is sometimes used as a synonym [119–122].

For evaluation of robustness, method parameters are varied, which are directly linked to the analytical equipment used, such as instrument settings. Therefore, these parameters are referred to as internal parameters. Ruggedness evaluation involves varying parameters such as laboratories etc. Hence, these parameters are described as external parameters.

The terms internal and external parameters do not cover all variables necessary to completely assess an analytical method's robustness and ruggedness. Further parameters such as the stability of standard and test solutions under the conditions needed (also mentioned in the ICH guideline Q2B) as well as the age and condition of consumable material (e.g., analytical columns) are no less important. Here, such parameters are considered as basic parameters.

In summary, an analyst must have a critical look at three different types of parameters when robustness and ruggedness are investigated:

- Internal parameters (e.g., temperature, pH, etc., in the case of HPLC).
- External parameters (e.g., different analysts, instruments, laboratories, etc.).
- Basic parameters (e.g., stability of test solutions, etc.).

The order of the three-parameter classes is not a ranking. Each of the parameters is of the same importance. A certain sequence can be derived merely with respect to an analytical method's life cycle. This means that at the beginning the basic parameters should be evaluated, then the influence of the internal parameters on the method's performance and finally the effect of the external parameters. It is useful to evaluate basic and internal parameters together, since both are implicated in analytical method development. Therefore, they are the first parameters determined in an analytical method's life cycle. The external parameters are estimated at a later point in time to a greater or lesser extent. The scope of this estimation depends on whether it is an intra-laboratory (precision, intermediate precision) or an inter-laboratory study (reproducibility, ruggedness).

2.7.2.1 Basic and Internal Parameters

Before beginning to develop a new analytical method, one decisive pre-requisite needs to be fulfilled: The analytical equipment must be qualified and the consumable material (e.g. columns) must be in good condition, or ideally, new or unused with defined performance characteristics, in order to generate meaningful data.

Furthermore, it has to be ensured that freshly prepared solutions are employed for method development.

After the method development process, the basic parameters are usually evaluated. The most important among these, and relevant for all analytical techniques, is the stability of test solutions used. For analytical separation techniques, such as chromatography (LC, GC) and capillary electrophoresis (CE), the evaluation of the following characteristics is also helpful [123]:

- Relative retention time / migration time.
- Column efficiency / capillary efficiency.
- Peak symmetry / peak shape.

The values obtained for these characteristics will serve as references for the robustness experiments and as a basis for establishing the system suitability test.

Generally, stability studies with test solutions are performed over 12, 24, 48 or even 72- hour periods of time. At each interval at least six replicated analytical runs (assays of one sample solution) are carried out. For the assessment of the results certain statistical tools are available:

- Trend test according to Neumann (statistical test versus tabulated values; significance level at $P = 95\,\%$ probability; [124–126]):

$$Q = \frac{1}{(n-1)s^2} \sum_{i=1}^{n-1} \left(x_i - x_{i+1}\right)^2 \qquad (2.7\text{-}1)$$

  n = number of measured values
  x_i, $x_{i\pm1}$ = measured values in chronological order
  s = standard deviation
  → Assessment criterion: If $Q >$ the respective tabulated value, then no trend exists.
- Trend test by linear regression
  → Assessment criterion: If the confidence interval (CI) of the slope includes zero and the CI of the y-intercept includes the assay found at t_0 (calculation by, e.g., MVA [28] or SQS [127]), then no trend exists.
- Coefficient of variation (CV)
  → Assessment criterion: If the CV of all values obtained at different time intervals does not exceed more than 20 % of the corresponding value at t_0, then no trend exists [128]. However, this depends on the respective method, the test item (assay, related impurities, etc.), the time interval and the measuring concentration. For assay and even for related impurities determined by HPLC it is recommended that the acceptance limits are tightened to 5 % and 10 %, respectively.
- Comparison of assay results at each time interval with the assay at the starting point (t_0)
  → Assessment criterion: If the assay at a certain time interval is within a predefined tolerance (that can be derived from the intermediate precision for instance), then no trend exists.

Normally, the information on the method characteristics is considered in the system suitability test and the results of the stability study are included in the control test as the defined shelf-life of the test solutions. A control test is the document describing the conduct of an analytical procedure.

The items that have been discussed here regarding the evaluation of basic parameters are analytical tasks that must be carried out between the development phase and the start of the basic validation of a new analytical procedure. The knowledge gained about basic parameters is a necessary pre-requisite before performing further studies on internal parameters, which can be considered as the "real robustness parameters". In order to have an idea of which internal parameters (robustness parameters) may be varied for typical analytical techniques predominantly used in pharmaceutical analysis, some examples are given below. This list does not claim to be complete:

- Gas Chromatography (GC)
  - Gas flow
  - Heating rate
  - Split ratio
  - Column type (manufacturer, batch of the stationary phase)
  - Sample preparation (pH of solutions, reagent concentration, etc.)
  - Injection temperature
  - Column temperature
  - Detection temperature.
- Capillary Electrophoresis (CE)
  - Voltage
  - Injection
  - Buffer concentration
  - Buffer pH
  - Buffer stability
  - Cooling (heat removal)
  - Sample preparation (pH of solutions, reagent concentrations, etc.)
  - Temperature
  - Detection wavelength.
- High Performance Liquid Chromatography (HPLC)
  - Column type (manufacturer, batch of stationary phase)
  - Temperature
  - pH (mobile phase)
  - Flow rate
  - Buffer concentration (ionic strength)
  - Additive concentration
  - Mobile phase composition (percentage of organic modifier)
  - Gradient slope
  - Initial mobile phase composition
  - Final mobile phase composition
  - Injection volume
  - Sample preparation (pH of solutions, reagent concentrations, etc.).

- Ion Chromatography (IC)
  - pH
  - Temperature
  - Flow rate
  - Column type (manufacturer, batch of stationary phase)
  - Sample preparation (pH of solutions, reagent concentrations etc.).
- Spectroscopy
  - Time constant
  - Solvent
  - pH of test solution
  - Temperature
  - Wavelength accuracy
  - Slit width
  - Sample preparation (pH of solutions, reagent concentrations, etc.).

From these lists one point becomes very clear. A test for robustness is an individual test and depends very much on the analytical technique and equipment applied. As a rule of thumb, it is recommended to examine at least those parameters, which are part of the operational qualification of the respective equipment (see Chapter 4). Then the set of parameters investigated in a robustness study can be arbitrarily extended to those specific to the method defined in the operating procedure.

The usual way of performing robustness testing is first to define the parameters with reasonable maximum variation. Then each parameter is successively varied, whereas the others are held constant (at nominal setting). For example, six parameters each at two levels would require twelve experiments, when one parameter is changed and the others are always set to nominal levels. The more parameters that are included, the more experiments must be conducted.

This classical approach is called one-factor-a-time (OFAT) approach. Certainly, this kind of robustness testing has disadvantages, as many experiments, time and resources are needed. In addition, only limited information is made available from such studies, since possible interactive effects, which occur when more than one parameter (factor) is varied, cannot be identified.

Nowadays, an experimental design approach (DOE: design of experiment) is often preferred for robustness testing. The aim of an experimental design is to obtain as much as possible relevant information in the shortest time from a limited number of experiments [129]. Different designs can be used in robustness testing, e.g. including full– and fractional – factorial designs as well as Plackett–Burman designs. The latter have become very popular in method robustness testing during recent years.

The choice of a design depends on the purpose of the test and the number of factors involved. Experimental designs in robustness testing can be employed for all analytical techniques. The general procedure for experimental design employed in robustness testing will be shown in the examples in section 2.7.3. HPLC is taken as an example, since this is still the most widely used analytical technique in pharmaceutical analysis and offers the possibility of applying chromatography modelling software as a further tool for robustness testing.

2.7.2.2 External Parameters

External parameters, such as different laboratories, analysts, instruments and days are an integral part of the ICH approach on the analytical method validation being considered in the determination of precision, comprising the system precision, repeatability, intermediate precision and reproducibility. The design of the final precision studies may vary slightly depending on the pharmaceutical development project itself and the individual planning. Usually, data on the inter-laboratory study, which is called reproducibility or ruggedness (USP), will be taken from the analytical transfer documentation describing and assessing the analytical investigation, which is being conducted at the development and production sites concerned [130].

As for internal parameters, the external parameters can also be examined by applying an experimental design, e.g., within the framework of analytical transfer. Internal and external parameters can also be combined in one experimental design [131].

However, this is not done very often, since it complicates the experimental set-up. In accordance with the ICH guidelines on analytical method validation it is recommended that internal and external parameters be examined separately.

2.7.2.3 Summary

Previously it has been established that a sufficient knowledge of basic parameters is an essential pre-requisite for establishing a new analytical procedure. This analytical work is well defined and is normally carried out after method development and before method validation.

Examinations of external parameters are also well defined, since they are explained in the ICH guidelines on the validation of an analytical method. The relevant data are required for submission to regulatory authorities.

The impact of the internal parameters (robustness parameters) on the performance of an analytical method, must be known and documented, but need not appear in the documentation submitted. These robustness factors are specific for each method and the scope of the investigations can differ depending on the method. The requirements and the extent of work are not comprehensively described in the literature. Robustness testing is time-consuming and resource-intensive, so that guidance in saving time and reducing the extent of work would be very helpful. For this reason a structured procedure for the performance of robustness testing for HPLC methods in pharmaceutical analysis is discussed below.

2.7.3 Examples of Computer-assisted Robustness Studies

In this chapter, two robustness studies, carried out at the Aventis GPD Analytical Sciences department in Frankfurt, Germany (GPD: Global Pharmaceutical Development), are described.

Having clarified that basic parameters are an analytical pre-requisite and that external parameters are in any case covered in validation studies performed in accordance with the ICH guidelines Q2A and Q2B (system precision, repeatability, inter-

mediate precision, reproducibility / ruggedness), real robustness studies on one exemplary HPLC method will be the focus of the following explanations.

In connection with this, two very helpful tools in robustness testing of HPLC methods will be discussed. One of these tools is chromatography modelling (e.g. DryLab [132], ChromSword [133], ACD [134]) and the other is the statistical design of experiments (e.g. MODDE [135], MINITAB [136], STATGRAPHICS [137]). Studies described here were conducted using DryLab and MODDE.

In the course of pharmaceutical development, robustness testing of analytical methods should start as early as possible. Normally, it is initiated in the pre-clinical phase, that is the time between the decision to further develop a new drug candidate (EDC decision) and the decision to start clinical phase I (phase I_IIa decision).

In this stage of pharmaceutical development it is useful to integrate robustness testing in a structured method development procedure based on a chromatography modelling software, such as DryLab.

Each time further analytical development is needed along the pharmaceutical development value chain, irrespectively whether methods for drug substance or drug product analysis, DryLab can be employed so that robustness data can be immediately derived from its calculations without significant additional work.

In the later stage of pharmaceutical development, when the analytical methods have been finalised, the robustness results obtained from DryLab should be supplemented by data from statistically designed experiments. This is illustrated in Figure 2.7-1 showing the Aventis value chain from an analytical development perspective.

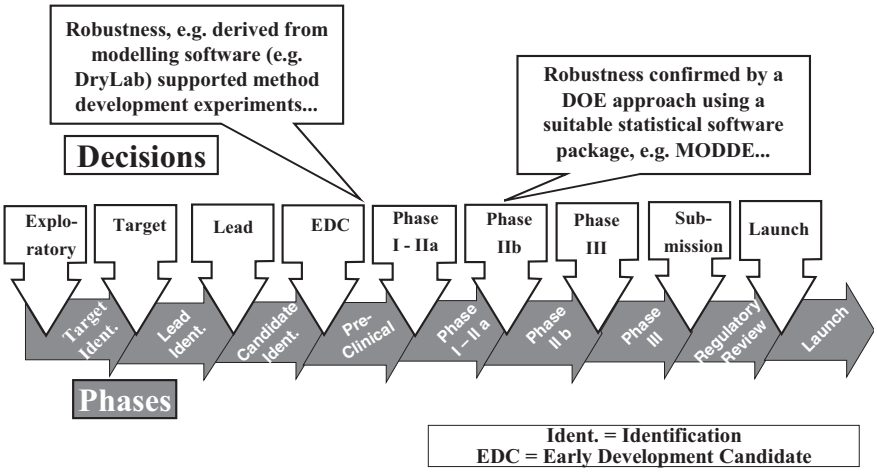


Figure 2.7-1: Utilization of LC modelling software and software for statistical design of experiments within the value chain of pharmaceutical method development.

2.7.3.1 Robustness Testing Based on Chromatography Modelling Software

When robustness testing is combined with analytical development in an early stage of pharmaceutical development (the pre-clinical phase) representative, reliable and predictive data of stress studies is needed [4]. This data is the basis of analytical development. To better understand the process that combines method development with robustness testing, the Aventis method development philosophy is briefly described. Along the Aventis value chain, analytical departments are responsible for drug substance and drug product analytical data from the initial candidate identification phase to the final regulatory review phase. Analytical method development begins in the Early Characterization Laboratories (EC Laboratories) and as soon as the EDC decision is taken it is continued in the Early Development Laboratories (ED Laboratories). The whole process is shown in Figure 2.7-2.

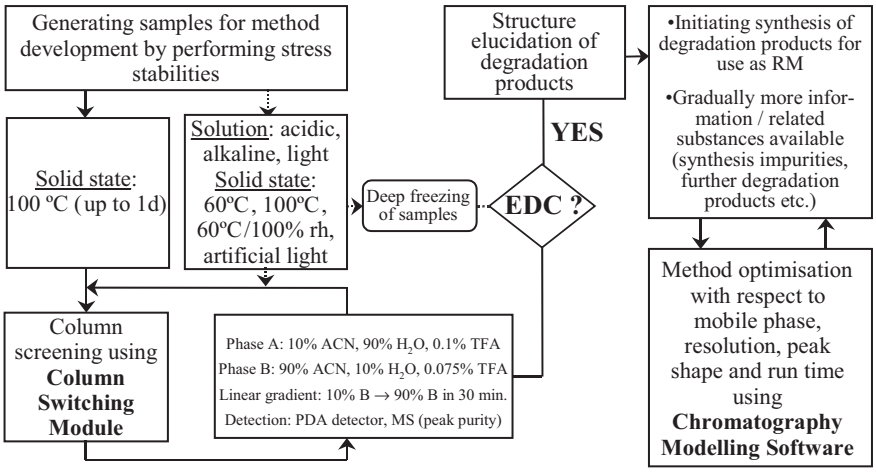


Figure 2.7-2: The Aventis analytical development process. Abbreviations: EDC = Early Development Compound, ACN = Acetonitrile, TFA = Trifluoroacetic Acid, RM = Reference Material.

Stress studies are carried out in the ECL with each new drug substance under development, utilizing stress conditions as indicated above. Additionally, mild oxidizing conditions are often applied. In any case it has to be ensured that the degradation does not significantly exceed 10 % in order to avoid the occurrence of secondary degradation products.

A column screening follows, using specific chromatographic conditions and different analytic columns for analysing 100 °C solid-state samples. A column-switching device is applied, which is capable of selecting up to six columns. Such experiments can be performed to run 24 hours per day, 7 days a week. This leads to more flexibility in the laboratory and helps to save valuable development time.

Once the EDC decision is taken, the structures of degradation products obtained from the stressed samples are elucidated and the LC methods are transferred from the EC to the ED Laboratories and, step-by-step, these degradation products are

available as reference materials. At this point method optimisation with respect to mobile phase, resolution, peak shape and run time is initiated. Optimisation studies are carried out by means of DryLab, which allows the simulation of chromatographic runs on the computer by varying variables such as the eluent composition, gradient time, pH, additive concentration, column dimension, flow rate, column temperature and particle size.

Furthermore, it permits the estimation of the robustness of a particular method. However, scouting analytical runs are a pre-requisite for DryLab calculations (simulations), since they 'calibrate the software' (a term used by the manufacturer, [132]).

Scouting runs are performed under the starting conditions listed below:

- Column: 150 × 4.6 mm, 5 µm; C18 or C8 stationary phase (defined after column screening).
- Buffer (solvent A): 5–20 mM phosphate, pH 2.0–3.5 (...and: pH = main component $pK_a \pm 1.5$ at least); alternatively 0.1 % TFA (trifluoroacetic acid or 0.1 % formic acid).
- Organic solvent (solvent B): Acetonitrile.
- Temperature: 35 °C.
- Flow rate: 1–2 ml/min.
- Additive: None.
- Recommendation:
  - Gradient HPLC method: start with 5 %–100 % solvent B in 30 min. (linear gradient)
  - Isocratic HPLC method: start with 90 % or 100 % solvent B *(only applied in exceptional cases)*.

These scouting chromatographic runs have been standardised within the Analytical Sciences department at Aventis in Frankfurt, making it possible for the laboratories to solve at least 90 % of the separation problems. In Figure 2.7-3 this procedure is illustrated. Only in a few cases does a more complicated approach need to be applied. For instance, a more complicated approach would require optimisation of the ionic strength of the buffer and / or the concentration of an additive.

As it is depicted in Figure 2.7-3 ten analytical runs, which are combined in two two-dimensional optimisation experiments, are generally sufficient for optimising HPLC methods for neutral compounds as well as for ionic compounds.

Even one-dimensional optimisations of pH and ternary solvent mixtures are often suitable and have the added advantage of reducing the number of experiments. Hence, only seven experiments in total would be enough to obtain reliable predictions.

It should be emphasised that pH optimisation is very important in pharmaceutical development due to the fact that most drug substances are salts, which exhibit better solubility and crystallinity than free acids or bases.

Four runs in a two-dimensional optimisation study are needed, when each parameter shows a linear relationship with the retention factor k, which is the case for gradient time and temperature. Two gradient runs allow the calculation of isocratic retention times as a function of mobile-phase composition and two runs at two dif-

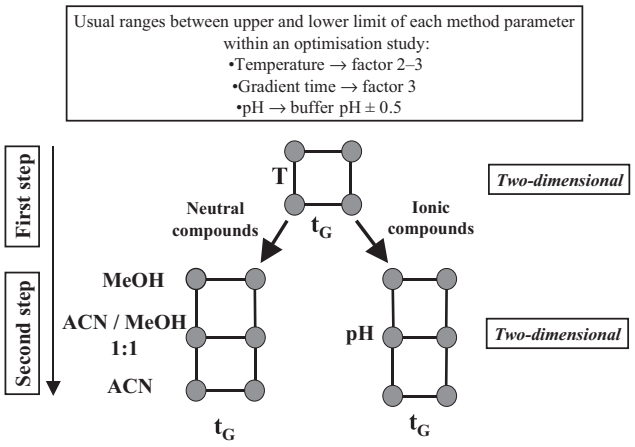


Figure 2.7-3: Standardised method development procedure using DryLab (Aventis, Frankfurt; GPD Analytical Sciences); the two-dimensional experiments of the second step can also be carried out one-dimensionally to save three experiments.

ferent temperatures allow the calculation of retention times as a function of temperature, as given in the equations below:

$$\log k = \frac{a}{T} + b \tag{2.7-2}$$

$$\log k = \log k_w - S\Phi \tag{2.7-3}$$

In these equations a and b are arbitrary constants, k_w is the retention factor for water as the mobile phase, Φ is the volume-fraction of the organic solvent in the mobile phase and S is a constant, that is a function of the molecular structure of each compound and the organic solvent.

If the relationship between the retention factor k and a certain method parameter is non-linear, DryLab applies quadratic and cubic spline fits, for example, for optimisation of ionic strength and additive concentration and for optimisation of pH and ternary solvent composition, respectively. For such studies three (quadratic fit) or at least three (cubic spline fit) scouting runs are necessary.

With the input data of the scouting runs DryLab can begin the calculation. The software evaluates the resolution R as a function of one (one-dimensional optimisation) or two (two-dimensional) chromatographic parameters for each peak pair. A so-called resolution map for the critical pair, which not only reveals the optimum chromatographic conditions but also the robust regions of an HPLC method, is produced. The resolution map of a one-dimensional optimisation is a common two-dimensional graph, whereas the resolution map of a two-dimensional optimisation takes the form of a three-dimensional contour plot, in which the third dimension is colour-coded. More detailed studies on HPLC method development are extensively discussed in the literature [138 – 140]. In this context the excellent and

comprehensive studies carried out by Snyder, Glajch and Kirkland are recommended for further reading [139].

2.7.3.1.1 Experimental Conduct and Results of a Robustness Study Based on Chromatography Modelling Software

The following example illustrates the application of the combined method development and robustness study with a drug substance in the pre-clinical phase. The drug substance is a salt of a carbonic acid with about 400 g/mol molar mass. The pK_a of the active moiety is 6.6. After column screening a Purospher STAR RP18 (125 mm length, 4.0 mm diameter, 5 µm particles) was selected. The HPLC method was developed to separate the drug substance (MC), relevant starting materials and intermediates (SP1, SP2, SP3) as well as the counter ion (CI; no quantitative determination) and a degradation product (DP1) of the active moiety.

HPLC method development was conducted in accordance with the procedure given in Figure 2.7-3. In addition to the experiments performed to optimise the gradient time, temperature and pH, experiments were also conducted to optimise the buffer concentration. The conditions for the scouting runs were as follows (mobile phase A: buffer/acetonitrile = 9/1; mobile phase B: water/acetonitrile = 1/9):

- Gradient time / temperature → 15 min. / 25 °C, 45 min. / 25 °C; 15 min. / 45 °C, 45 min. / 45 °C; pH 3.0, 4.2 mM buffer concentration.
- pH at pH 2.4, pH 3.0, pH 3.6 (buffer concentration: 4.2 mM / gradient time: 20 min. / segmented gradient / temperature: 35 °C).
- Buffer concentration at 5 mM, 10 mM and 20 mM phosphate (pH: 3.5 / gradient time: 20 min. / segmented gradient / temperature: 35 °C).

Each set of experiments was founded on the results of each previous experimental set. For each of those sets (gradient time / temperature; pH; buffer strength) a resolution map was obtained allowing the identification of the optimum chromatographic parameters, which were subsequently confirmed experimentally. In Figure 2.7-4 the resolution map of the two-dimensional optimisation of gradient time and

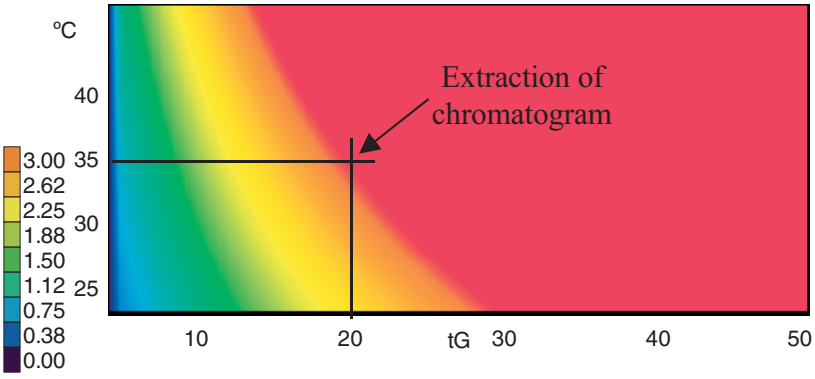


Figure 2.7-4: Resolution map of a two-dimensional optimisation of gradient time and temperature. The colour as the third dimension of the map represents the resolution for the critical peak pair, as shown in the legend left. The extraction of the optimised chromatogram is indicated at the intersection of the lines for gradient time and temperature.

temperature is shown. The x-axis represents gradient time and the y-axis represents temperature. The critical resolution for each gradient time / temperature coordinate is given by different colours. In Figure 2.7-4 blue indicates a bad resolution whilst yellow and red indicate a good and very good resolution at a level above 2, respectively.

In this resolution map an optimised gradient shape has already been implemented, which means that the linear gradient has been modified into a segmented gradient. A satisfactory separation is obtained, when the gradient time is 20 minutes and the column temperature is 35 °C. This chromatogram calculated by DryLab is shown in Figure 2.7-5 in comparison to that obtained experimentally. This comparison demonstrates the good agreement between calculated and experimental data. There are no marked differences in retention times and peak areas. Therefore, these parameters were implemented in the design for pH optimisation. The three experi-

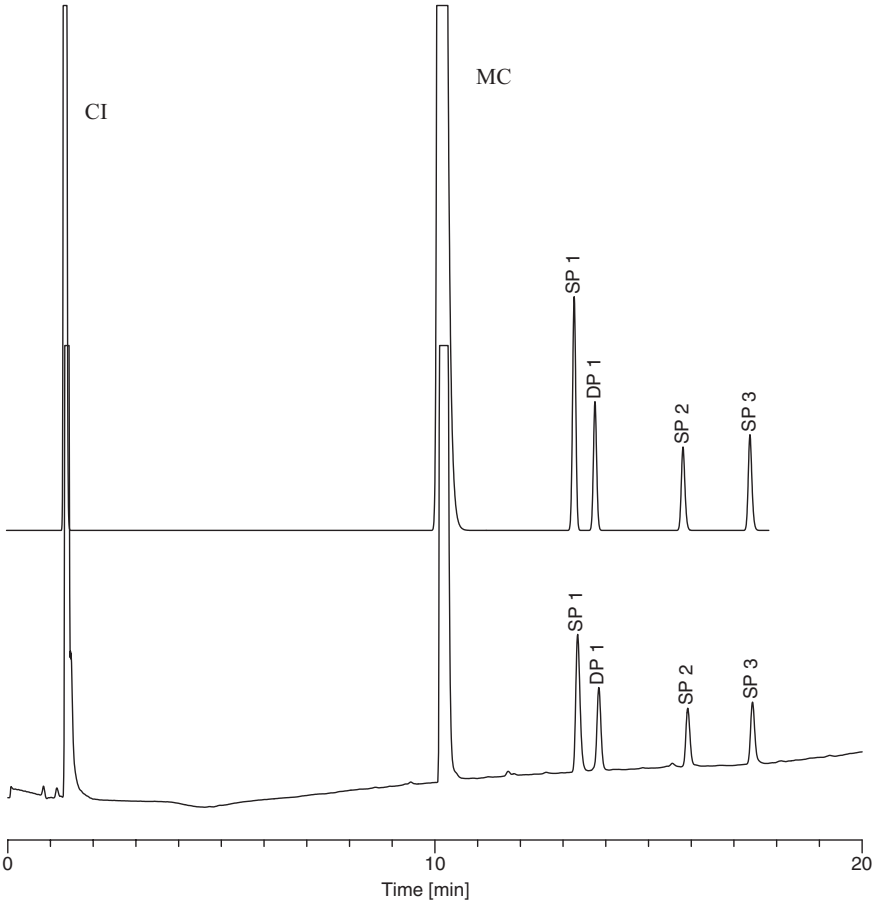


Figure 2.7-5: DryLab prediction obtained for 20 minutes gradient time and 35 °C column temperature (upper chromatogram) in comparison with the experimental run at these conditions (lower chromatogram)

ments mentioned above were performed, a chromatogram was extracted at pH 3.5 and the prediction obtained was experimentally confirmed. Then the final optimisation of the buffer concentration was done. From the corresponding resolution map shown in Figure 2.7-6 a 10 mM buffer concentration leads to further improvement in peak resolution. With buffer concentrations above 10 mM the probability that the buffer substance precipitates in certain parts of the HPLC equipment during routine use (for examples in fittings) would increase.

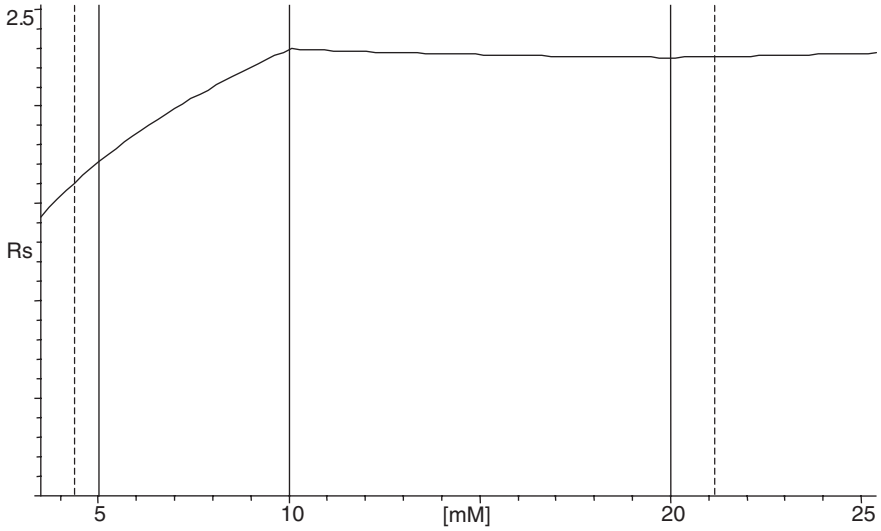


Figure 2.7-6: Resolution map of a one-dimensional optimisation of buffer concentration. Solid lines represent the concentrations for the experiments done (5, 10, 20 mM); the dashed lines represent the range in which predictions are allowed.

The chromatogram extracted for a 10 mM buffer concentration as well as that experimentally obtained are given in Figure 2.7-7.

These chromatograms correspond very well with respect to retention times and peak areas. Comparing the chromatograms in Figure 2.7-7 with those in Figure 2.7-5 it is striking that in Figure 2.7-7 an additional peak can be observed, Unknown 1, representing a low amount of an impurity of SP2. This additional peak had already been observed after pH optimisation demonstrating the importance of pH experiments, especially when a salt of a drug substance is being examined.

This example clearly confirms that it is possible to develop a reliable analytical method with only seven (without optimisation of buffer concentration) to ten (with optimisation of buffer concentration) scouting runs.

Generally, it is feasible to approximately estimate the robustness of an analytical method from the resolution maps obtained during the method development process, but for accurate evaluation all resolution maps must be calculated based on runs conducted at the final chromatographic conditions.

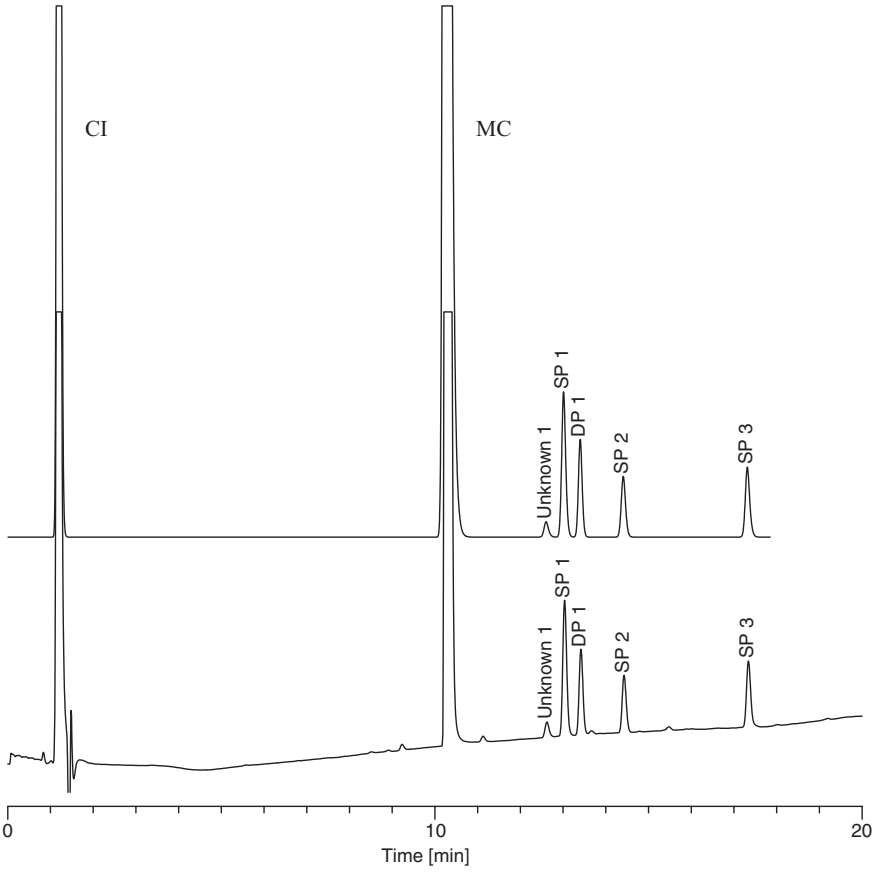


Figure 2.7-7: DryLab prediction obtained for 10 mM buffer concentration (upper chromatogram) in comparison with the experimental run under this condition (lower chromatogram).

In the example described, these final conditions were:

- Gradient time (segmented gradient) → 20 minutes.
- Column temperature → 35 °C.
- Mobile phase pH → 3.5.
- Buffer concentration: 10 mM.

Consequently, only the resolution map for optimisation of the buffer concentration was final and could be used for predictions on robustness. For gradient time and column temperature, as well as for pH, the resolution maps had to be created.

Therefore, four experimental runs at 3.5 mobile phase pH and 10 mM buffer concentration were carried out for creating a three-dimensional resolution map for gradient time and temperature (15 min. / 25 °C, 45 min. / 25 °C; 15 min. / 45 °C, 45 min. / 45 °C).

Additional experiments had to be carried out to generate a final resolution map for predictions on pH robustness. For that purpose two runs were sufficient, since the third run, necessary for calculation, could be taken from the experiments conducted for optimisation of buffer concentration (pH 3.5, 10 mM buffer concentration). The pH conditions chosen were pH 3.1 and pH 3.9. The corresponding resolution map is illustrated in Figure 2.7-8. This resolution map impressively illustrates the strong dependency of peak resolution on changes of mobile phase pH. Within the pH range examined the predictions for the critical resolution varied from zero to approximately 2.4. None of the other chromatographic parameters had such a significant impact on peak resolution.

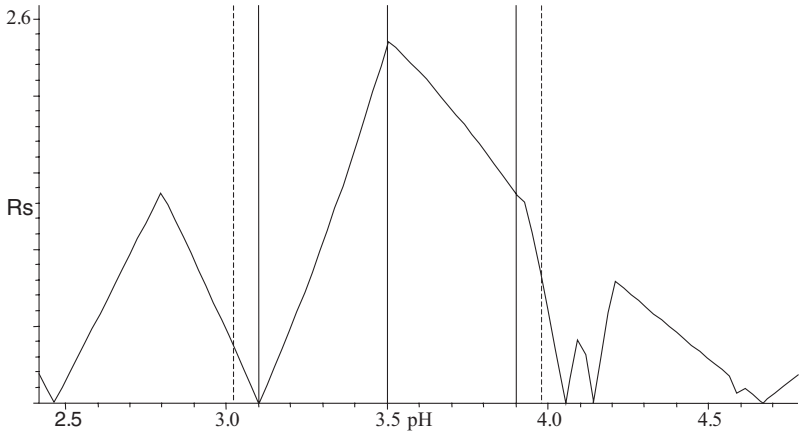


Figure 2.7-8: Final resolution map of a one-dimensional optimisation of mobile phase pH. The solid bars represent the pHs for the experiments done (3.1, 3.5, 3.9); the dashed bars represent the range in which predictions are allowed.

From all these final resolution maps, chromatograms can be extracted, representing runs obtained after small but deliberate changes in the chromatographic parameters of column temperature (gradient time is assumed to be correct), and pH and buffer concentration. Furthermore, the influence of the mobile phase flow rate and the particle size of the stationary phase on the peak resolution can be calculated without the need for specific scouting runs. This is possible due to the fact that on the basis of an existing resolution map (for example, temperature / gradient time) a further map can be calculated enabling the analyst to evaluate the changes in resolution.

Tables 2.7-1 and 2.7-2 summarise the DryLab predictions on robustness of the analytical method described above and compare them with experimental data.

Examination of data reveals that the theoretical data, i.e., the data from the predictions, match remarkably well with the experimental data with respect to retention time, peak area and critical resolution.

However, it must be noted that differences in peak areas occur when the flow rate is varied. The experimental data confirm the well-known phenomenon that a

decrease / increase in the mobile phase flow rate leads to an increase / decrease in the peak areas [141]. DryLab is not able to simulate this behaviour when using the scouting runs presented, but with DryLab it is possible to predict the critical resolution R_s, when the mobile phase flow rate is varied, which is of greater importance in assessing the robustness of an analytical method.

To make a clear decision on whether the analytical HPLC method is robust with respect to selectivity and in particular with regard to the parameters flow rate, column temperature, pH, buffer concentration and particle size, it is helpful to calculate the data given in Table 2.7-3 (based on data of Tables 2.7-1 and 2.7-2). It is striking that the experimental and theoretical data on critical resolution fit, which indicates that an adequate set of experiments had been selected. Even more convincing than the absolute values obtained for critical resolutions are the differences between the resolutions found at nominal conditions and conditions above or below the nominal parameter settings. These numbers are almost identical.

2.7.3.1.2 **Assessment of the Experimental Results**

General Note: A resolution of 1.5 between two adjacent peaks is regarded to be sufficient for accurate peak integration, if the peak areas are not too much different [141].

Flow rate: The analytical method is certainly sensitive to changes in the mobile phase flow rate, but in these experiments the worst resolution of 1.75 can still be a 'comfortable critical resolution' when the peak areas are not markedly different [141]. Nevertheless, the data show that an accurate flow rate is desirable, which is ensured by adequate equipment qualification and equipment maintenance, which of course is obvious in a GMP-regulated environment. Considering variations of ± 1–2 % which HPLC pumps normally exhibit, this analytical method is regarded as robust, since critical resolutions around 2.3 are guaranteed.

Column Temperature: The analytical method is robust with regard to temperature, since the critical resolution is around two in the temperature range of 32 °C – 38 °C. However, what has been mentioned on equipment qualification and equipment maintenance under the item 'flow rate', also applies to temperature.

pH: Taking into account that the HPLC methods applied for organic salts are generally very sensitive to changes in pH, this method can be considered as robust, since the critical resolution is between 1.7 and 2.2 in a pH range between 3.4 and 3.6. Nevertheless, a pH accuracy of ± 0.05 pH should be ensured to obtain a critical resolution of around two. Sometimes the robust pH range can be directly derived from the resolution map, when a plateau is obtained as shown in Figure 2.7-9. Unfortunately, this is only observed in very rare cases.

Table 2.7-1 Predictions obtained with DryLab for the parameters flow rate, column temperature, pH, buffer concentration and particle size.

Parameter	Below Nominal Nominal Above Nominal	CI		MC		Unknown 1		SP1		DP1		SP2		SP3		Critical R_s
		t_R [min]	Area [mAU·min]	t_R [min]	Area [mAU·min]	t_R [min]	Area [mAU·min]	t_R [min]	Area [mAU·min]	t_R [min]	Area [mAU·min]	t_R [min]	Area [mAU·min]	t_R [min]	Area [mAU·min]	
Flow rate [ml/min]	0.7	1.70	21.7	12.37	40.1	15.07	0.1	15.45	1.0	15.77	0.7	16.78	0.5	19.75	0.5	1.85
	1.0	1.19	21.7	10.28	40.1	12.64	0.1	13.05	1.0	13.42	0.7	14.43	0.5	17.36	0.5	2.46
	1.3	0.92	21.7	9.06	40.1	11.17	0.1	11.59	1.0	12.00	0.7	13.01	0.5	15.91	0.5	2.85
Column	30	1.21	21.7	10.38	40.1	12.88	0.1	13.30	1.0	13.61	0.7	14.64	0.5	17.52	0.5	2.03
Temperature	32	1.20	21.7	10.34	40.1	12.78	0.1	13.20	1.0	13.54	0.7	14.55	0.5	17.45	0.5	2.20
[°C]	**35**	1.19	21.7	10.28	40.1	12.64	0.1	13.05	1.0	13.42	0.7	14.43	0.5	17.36	0.5	2.46
	38	1.18	21.7	10.22	40.1	12.50	0.1	12.89	1.0	13.30	0.7	14.30	0.5	17.26	0.5	2.52
	40	1.18	21.7	10.18	40.1	12.41	0.1	12.79	1.0	13.23	0.7	14.22	0.5	17.20	0.5	2.48
pH	3.1	1.31	21.0	10.36	39.7	13.27	0.1	13.27	1.1	13.74	0.7	15.48	0.5	17.42	0.5	0.00
	3.2	1.28	21.0	10.32	40.0	13.13	0.1	13.22	1.1	13.68	0.7	15.24	0.5	17.39	0.5	0.49
	3.3	1.24	21.0	10.30	40.2	12.98	0.1	13.17	1.0	13.61	0.6	14.98	0.5	17.37	0.5	1.05
	3.4	1.22	21.0	10.28	40.4	12.81	0.1	13.10	1.0	13.52	0.6	14.71	0.5	17.35	0.5	1.66
	3.5	1.19	21.1	10.27	40.6	12.62	0.1	13.03	1.0	13.42	0.6	14.43	0.4	17.34	0.5	2.34
	3.6	1.17	21.3	10.27	40.7	12.41	0.1	12.95	1.0	13.31	0.6	14.13	0.4	17.33	0.5	2.15
	3.7	1.15	21.6	10.28	40.9	12.19	0.1	12.86	1.0	13.18	0.6	13.82	0.4	17.33	0.5	1.91
	3.8	1.13	22.0	10.30	41.0	11.94	0.1	12.76	1.0	13.04	0.7	13.50	0.4	17.34	0.6	1.64
	3.9	1.12	22.4	10.33	41.2	11.68	0.1	12.65	1.0	12.88	0.7	13.16	0.4	17.35	0.6	1.35

Table 2.7-1 Continued.

Parameter	Below Nominal Nominal Above Nominal	CI		MC		Unknown 1		SP1		DP1		SP2		SP3		Critical R_s
		t_R [min]	Area [mAU·min]	t_R [min]	Area [mAU·min]	t_R [min]	Area [mAU·min]	t_R [min]	Area [mAU·min]	t_R [min]	Area [mAU·min]	t_R [min]	Area [mAU·min]	t_R [min]	Area [mAU·min]	
Buffer conc.	5	1.22	19.9	10.16	36.7	12.82	0.1	13.12	1.0	13.53	0.6	14.82	0.4	17.37	0.5	1.71
[mM]	8	1.21	20.8	10.24	38.3	12.69	0.1	13.05	1.0	13.45	0.6	14.59	0.4	17.34	0.5	2.10
	10	1.20	21.2	10.28	39.0	12.64	0.1	13.03	1.0	13.42	0.6	14.49	0.4	17.34	0.5	2.29
	12	1.19	21.5	10.31	39.4	12.60	0.1	13.02	1.0	13.40	0.6	14.42	0.4	17.34	0.5	2.28
	15	1.19	21.7	10.35	39.7	12.57	0.1	13.01	1.0	13.38	0.6	14.33	0.4	17.35	0.5	2.26
	20	1.18	21.3	10.40	38.8	12.53	0.1	13.01	1.0	13.37	0.6	14.24	0.4	17.37	0.5	2.25
Particle Size	7	1.19	21.7	10.28	40.1	12.64	0.1	13.05	1.0	13.42	0.7	14.43	0.5	17.36	0.5	1.99
[µm]	**5**	1.19	21.7	10.28	40.1	12.64	0.1	13.05	1.0	13.42	0.7	14.43	0.5	17.36	0.5	2.46
	3	1.19	21.7	10.28	40.1	12.64	0.1	13.05	1.0	13.42	0.7	14.43	0.5	17.36	0.5	3.20

Table 2.7-2 Experimental data obtained for flow rate, column temperature, pH, buffer concentration and particle size.

Parameter	Below Nominal Nominal Above Nominal	CI		MC		Unknown 1		SP1		DP1		SP2		SP3		Critical R_s
		t_R [min]	Area [mAU·min]	t_R [min]	Area [mAU·min]	t_R [min]	Area [mAU·min]	t_R [min]	Area [mAU·min]	t_R [min]	Area [mAU·min]	t_R [min]	Area [mAU·min]	t_R [min]	Area [mAU·min]	
Flow rate [ml/min]	0.7	1.65	30.49	12.03	55.69	14.75	0.16	15.14	1.45	15.48	0.90	16.53	0.63	19.65	0.77	1.75
	1.0	1.19	21.15	10.27	40.56	12.62	0.11	13.03	1.03	13.42	0.63	14.43	0.45	17.34	0.54	2.30
	1.3	0.92	16.44	9.07	30.30	11.13	0.10	11.56	0.80	11.98	0.50	12.99	0.34	15.89	0.41	2.70
Column	30	1.21	21.69	10.37	39.84	12.86	0.13	13.29	1.04	13.60	0.68	14.64	0.45	17.50	0.56	1.87
Temperature	**35**	1.19	21.15	10.27	40.56	12.62	0.11	13.03	1.03	13.42	0.63	14.43	0.45	17.34	0.54	2.30
[°C]	40	1.18	21.71	10.17	39.96	12.39	0.13	12.78	1.06	13.22	0.66	14.21	0.46	17.17	0.55	2.34
pH	3.1	1.31	21.04	10.36	39.73	–	–	13.27	1.20	13.74	0.71	15.48	0.52	17.42	0.55	0
	3.5	1.19	21.15	10.27	40.56	12.62	0.11	13.03	1.03	13.42	0.63	14.43	0.45	17.34	0.54	2.30
	3.9	1.12	22.38	10.33	41.16	11.68	0.12	12.65	1.03	12.88	0.71	13.16	0.43	17.35	0.56	1.38
Buffer conc.	5	1.22	19.93	10.16	36.73	12.82	0.11	13.12	0.96	13.53	0.63	14.82	0.44	17.37	0.50	1.70
[mM]	**10**	1.19	21.15	10.27	40.56	12.62	0.11	13.03	1.03	13.42	0.63	14.43	0.45	17.34	0.54	2.30
	20	1.18	21.29	10.40	38.81	12.53	0.12	13.01	1.00	13.37	0.63	14.24	0.44	17.37	0.53	2.22
Particle size	**5**	1.19	21.15	10.27	40.56	12.62	0.11	13.03	1.03	13.42	0.63	14.43	0.45	17.34	0.54	2.30
[µm]	3	1.08	21.92	10.07	40.18	12.33	0.12	12.72	1.05	13.10	0.67	14.12	0.46	17.14	0.56	2.86

Table 2.7-3 Comparison of critical resolution data of predictions obtained and experiments performed.

Parameter	Below Nominal (BN) Nominal (N) Above Nominal (AN)	Predicted Critical Resolution R_s(pred.)	Experimental Critical Resolution R_s(exp.)	R_s(pred.) – R_s(exp.)	Experimental (N – BN) and (N – AN)	Predicted (N – BN) and (N – AN)
Flow rate	0.7	1.85	1.75	0.10	***0.55***	***0.61***
[ml/min]	**1.0**	2.46	2.30	0.16	—	–
	1.3	2.85	2.70	0.15	***–0.40***	***–0.39***
Column	30	2.03	1.87	0.16	***0.43***	***0.43***
Temperature	32	2.20	–	–	–	0.26
[°C]	**35**	2.46	2.30	0.16	–	
	38	2.52	–	–		–0.06
	40	2.48	2.34	0.14	***–0.04***	***–0.02***
pH	3.1	0.00	0.00	0.00	***2.30***	***2.34***
	3.2	0.49	–	–	–	1.85
	3.3	1.05				1.29
	3.4	1.66				0.68
	3.5	2.34	2.30	0.04		—
	3.6	2.15	–	–		0.19
	3.7	1.91				0.24
	3.8	1.64				0.70
	3.9	1.35	1.38	–0.03	***0.92***	***0.99***
Buffer conc.	5	1.71	1.70	0.01	***0.60***	***0.58***
[mM]	8	2.10	–	–	–	0.19
	10	2.29	2.30	–0.01		–
	12	2.28	–	–		0.01
	15	2.26		0.03		
	20	2.25	2.22	0.03	***0.08***	***0.04***
Particle size	7	1.99	–		–	0.47
[µm]	**5**	2.46	2.30	0.16		–
	3	3.20	2.86	0.34	***–0.56***	***–0.74***

Buffer concentration: With regard to buffer concentration this analytical method is also robust. In a range from 8 mM to 12 mM the critical resolution is clearly above two, which is a sufficient range with respect to the nominal setting of 10 mM.

Particle size: According to theory, the data shows that smaller particle size of the same stationary phase material leads to better critical resolutions.

Further parameters that can be simulated by DryLab are the influence of different dwell volumes and the column dimensions of length and diameter. However, these parameters were not covered in the study presented here.

2.7.3.1.3 Conclusion

The example discussed in this section impressively demonstrates that chromatography modelling software (e.g., DryLab) is a very helpful tool in assessing robustness of an HPLC method. Based on the extremely good agreement between predicted

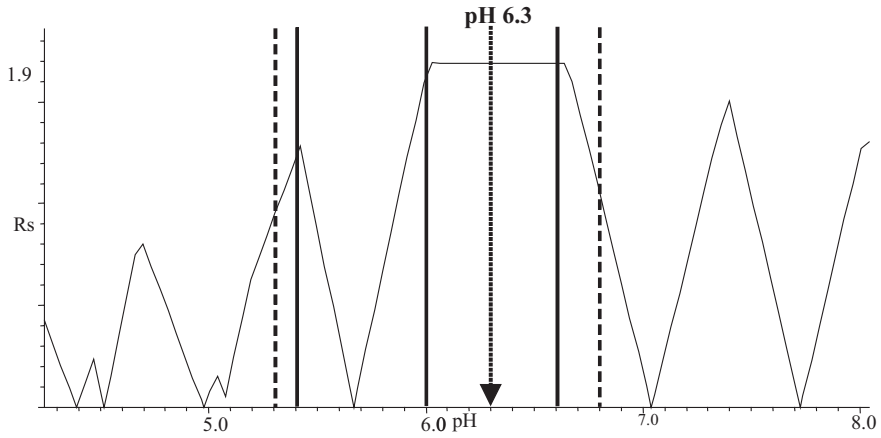


Figure 2.7-9: Final resolution map of a one-dimensional optimisation of mobile phase pH; solid lines represent the pHs for the experiments done (5.4, 6.0, 6.6); the dashed lines represent the range in which predictions are allowed. The arrow indicates the pH optimum.

and experimental data (especially for (N - AN), (N - BN)) one can rely on the DryLab calculations and there is not necessarily a need for experimental confirmation after each variation of a chromatography parameter, when at least the final chromatogram extracted from each final resolution map has been experimentally confirmed (confirmation of the nominal settings).

This procedure is certainly acceptable for the early drug development phase. Should DryLab also be used in a later stage of drug development, when transfer and submission activities are initiated, it is strongly recommended to perform a confirmatory experiment in every case.

It must be emphasised that, after completion of the DryLab robustness investigation, it has to be demonstrated that the results obtained with columns containing stationary phase material of different batches are identical with those of the robustness study with respect to selectivity, peak area and peak shape. Here, the normal variation in chromatographic results has to be taken into consideration. Many column manufacturers offer so-called validation kits for such a purpose.

The use of DryLab in robustness testing is also described in the literature but most publications are mainly focussed on method development aspects. Most of the discussion is related to the effects of varying the chromatographic parameters on the chromatographic performance but little on the range of acceptability of each of the parameters, which would still permit adequate performance [142–145]. In principle, it must be underlined that a good knowledge and experience in HPLC helps to develop robust methods. But to improve that knowledge the use of chromatography modelling software (e.g. DryLab) as a supporting tool, is always a good choice [146–152].

2.7.3.2 Robustness Testing Based on Experimental Design

The use of experimental design (DOE: design of experiments) for robustness studies in pharmaceutical analysis, has become more frequent during recent years. This is mainly related to the fast development in computer technology and the widespread availability of capable and user-friendly statistical software packages, such as MODDE, MINITAB, STATGRAPHICS, etc. However, some theoretical background knowledge on statistical design and data evaluation is useful for accurate interpretation of the calculated results obtained with such software packages. For details on statistical design and data evaluation, the reader is referred to relevant literature [153 – 156], but for a better understanding, a short digression is given below.

Presuming that an analyst wants to investigate k factors (i.e., robustness parameters) at n levels (settings), this would result in

$$N = n^k \tag{2.7-4}$$

where N = the number of experiments. Consequently, for three factors at two levels, eight experiments need to be done for a full evaluation (full factorial design). It is assumed that the relationship between the response and the factors (parameters) can be explained by a polynomial equation that also could contain powers of each term, and mixed terms in which the terms could be represented by any power. Due to the fact that in this example (n= 2; k= 3) two factor levels are considered, only those terms are taken into account, in which each factor is linear (Eq. 2.7-5).

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{23}x_2x_3 + b_{123}x_1x_2x_3 \tag{2.7-5}$$

Provided that interactions between the factors can be neglected, as is mostly the case for robustness studies on analytical methods, eight experiments are sufficient for investigating up to seven factors (Eq. 2.7-6). These investigations may be carried out by means of Plackett–Burman designs (see section 2.7.3.2.2).

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 + b_5x_5 + b_6x_6 + b_7x_7 \tag{2.7-6}$$

Therefore, Plackett–Burman designs enable an analyst to perform robustness studies with only a few experiments. For instance, a full factorial design for seven factors would require 128 experiments ($2^7 = 128$) which would need a lot more working time and resources compared with the eight experiments of a Plackett–Burman design and, with respect to robustness studies, would not reveal any more relevant information. Unquestionably, examples such as this demonstrate the great advantage of reduced experimental designs (e.g., Plackett–Burman).

To improve the interpretability of coefficients, it is common practice to perform a coordinate transformation, i.e., scaling and centring the variables (Eq. 2.7-7). It should be noticed that x_i' is dimensionless in Eq. (2.7-7):

$$x_i' = \frac{x_i - x_{i,c}}{x_{i,h} - x_{i,c}} \tag{2.7-7}$$

In Eq.(2.7-7) the indices c and h represent the centres and the high levels of the factors x_i [157, 158]. In the transformed coordinate system, the experimental design for three factors takes the form of a cube illustrated in Figure 2.7-10 (for $k > 3$ usually tabular overviews are provided).

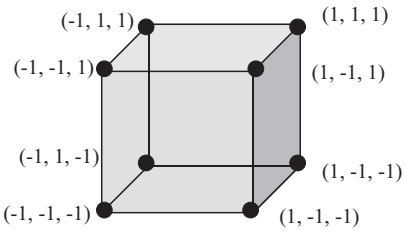


Figure 2.7-10: Experimental design for three factors ($k = 3$) at two levels ($n = 2$) in a transformed coordinate system.

Based on this design, as soon as response values, y_i, are available, the respective system of linear equations for the coefficients, b_i, is given as follows:

$$\begin{aligned}
y_1 &= 1 \cdot b_0 - 1 \cdot b_1 - 1 \cdot b_2 + 1 \cdot b_{12} - 1 \cdot b_3 + 1 \cdot b_{13} + 1 \cdot b_{23} - 1 \cdot b_{123} \\
y_2 &= 1 \cdot b_0 + 1 \cdot b_1 - 1 \cdot b_2 - 1 \cdot b_{12} - 1 \cdot b_3 - 1 \cdot b_{13} + 1 \cdot b_{23} + 1 \cdot b_{123} \\
y_3 &= 1 \cdot b_0 - 1 \cdot b_1 + 1 \cdot b_2 - 1 \cdot b_{12} - 1 \cdot b_3 + 1 \cdot b_{13} - 1 \cdot b_{23} + 1 \cdot b_{123} \\
y_4 &= 1 \cdot b_0 + 1 \cdot b_1 + 1 \cdot b_2 + 1 \cdot b_{12} - 1 \cdot b_3 - 1 \cdot b_{13} - 1 \cdot b_{23} - 1 \cdot b_{123} \\
y_5 &= 1 \cdot b_0 - 1 \cdot b_1 - 1 \cdot b_2 + 1 \cdot b_{12} + 1 \cdot b_3 - 1 \cdot b_{13} - 1 \cdot b_{23} + 1 \cdot b_{123} \\
y_6 &= 1 \cdot b_0 + 1 \cdot b_1 - 1 \cdot b_2 - 1 \cdot b_{12} + 1 \cdot b_3 + 1 \cdot b_{13} - 1 \cdot b_{23} - 1 \cdot b_{123} \\
y_7 &= 1 \cdot b_0 - 1 \cdot b_1 + 1 \cdot b_2 - 1 \cdot b_{12} + 1 \cdot b_3 - 1 \cdot b_{13} + 1 \cdot b_{23} - 1 \cdot b_{123} \\
y_8 &= 1 \cdot b_0 + 1 \cdot b_1 + 1 \cdot b_2 + 1 \cdot b_{12} + 1 \cdot b_3 + 1 \cdot b_{13} + 1 \cdot b_{23} + 1 \cdot b_{123}
\end{aligned} \tag{2.7-8}$$

The coefficients b_i can be determined by adequate addition and subtraction of the equations above:

$$\begin{aligned}
b_0 &= 1/8\left(y_1 + y_2 + y_3 + y_4 + y_5 + y_6 + y_7 + y_8\right) \\
b_1 &= 1/8\left(\left(y_2 + y_4 + y_6 + y_8\right) - \left(y_1 + y_3 + y_5 + y_7\right)\right) \\
b_2 &= 1/8\left(\left(y_3 + y_4 + y_7 + y_8\right) - \left(y_1 + y_2 + y_5 + y_6\right)\right) \\
b_{12} &= 1/8\left(\left(y_1 + y_4 + y_5 + y_8\right) - \left(y_2 + y_3 + y_6 + y_7\right)\right) \\
b_3 &= 1/8\left(\left(y_5 + y_6 + y_7 + y_8\right) - \left(y_1 + y_2 + y_3 + y_4\right)\right) \\
b_{13} &= 1/8\left(\left(y_1 + y_3 + y_6 + y_8\right) - \left(y_2 + y_4 + y_5 + y_7\right)\right) \\
b_{23} &= 1/8\left(\left(y_1 + y_2 + y_7 + y_8\right) - \left(y_3 + y_4 + y_5 + y_6\right)\right) \\
b_{123} &= 1/8\left(\left(y_2 + y_3 + y_5 + y_8\right) - \left(y_1 + y_4 + y_6 + y_7\right)\right)
\end{aligned} \tag{2.7-9}$$

In literature the effects (Eff) of factors on responses are defined as:

$$Eff\left(x_i\right) = 2b_j \tag{2.7-10}$$

If an experimental design is more complicated than the 2^3-design shown above, a lot more time and effort has to be put into the calculation of the coefficients and effects. In such a case the aforementioned software packages are needed. These software packages mostly use multiple linear regressions for evaluation of experimentally designed studies. In a regression analysis a minimum for the inexplicable error ε is obtained by means of the least-squares fit procedure.

A general form of a linear regression function is given in Eq.(2.7-11):

$$y = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_k x_k + \varepsilon \tag{2.7-11}$$

If the variables x_{1i}, x_{2i}, ..., x_{ki} and y_i are known from adequately designed experiments, the regression coefficients b_1, b_2, ..., b_k can be determined in accordance with Eq.(2.7-12), which shows a system of normal equations [157]:

$$\begin{aligned}
&b_0 Q_0 + b_1 Q_{01} + b_2 Q_{02} + \ldots + b_k Q_{0k} = Q_{0y} \\
&b_0 Q_{01} + b_1 Q_{x_1} + b_2 Q_{x_1 x_2} + \ldots + b_k Q_{x_1 x_k} = Q_{x_1 y} \\
&b_0 Q_{02} + b_1 Q_{x_1 x_2} + b_2 Q_{x_2} + \ldots + b_k Q_{x_2 x_k} = Q_{x_2 y} \\
&\cdots = \cdots \\
&b_0 Q_{0k} + b_1 Q_{x_1 x_k} + b_2 Q_{x_2 x_k} + \ldots + b_k Q_{x_k} = Q_{x_k y}
\end{aligned} \tag{2.7-12}$$

The sum of squares can be calculated using Eq.(2.7-13):

$$\begin{aligned}
Q_0 &= N \\
Q_{0j} &= \sum_{i=1}^{N} \left(x_{ij} - \bar{x}_j \right) \\
Q_{x_j} &= \sum_{i=1}^{N} \left(x_{ij} - \bar{x}_j \right)^2 \\
Q_{x_j x_{j'}} &= \sum_{i=1}^{N} \left(x_{ij} - \bar{x}_j \right) \left(x_{ij'} - \bar{x}_{j'} \right) \\
Q_{x_j y} &= \sum_{i=1}^{N} \left(x_{ij} - \bar{x}_j \right) (y_i - \bar{y}) \\
Q_{0y} &= \sum_{i=1}^{N} (y_i - \bar{y})
\end{aligned} \tag{2.7-13}$$

For a scaled and centred coordinate system and an orthogonal design like the two-level factorial, the following equations are also valid (Eq. 2.7-14):

$$\begin{aligned}
Q_{x_j} &= \sum_{i=1}^{N} \left(x_{ij} - \bar{x}_j \right)^2 = N \\
Q_{x_j x_{j'}} &= \sum_{i=1}^{N} \left(x_{ij} - \bar{x}_j \right) \left(x_{ij} - \bar{x}_j \right) = 0
\end{aligned} \tag{2.7-14}$$

When the general equation of the mathematical model is expressed as a matrix term, it takes the form of Eq.(2.7-15):

$$Y = Xb + e \tag{2.7-15}$$

The variables in Eq.(2.7-15) represent the terms shown in Eq.(2.7-16):

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \dots \\ y_N \end{pmatrix} \quad X = \begin{pmatrix} 1 & x_{11} & x_{21} & \dots & x_{k1} \\ 1 & x_{12} & x_{22} & \dots & x_{k2} \\ 1 & x_{13} & x_{23} & \dots & x_{k3} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{1N} & x_{2N} & \dots & x_{kN} \end{pmatrix} \quad e = \begin{pmatrix} e_1 \\ e_2 \\ e_3 \\ \vdots \\ e_N \end{pmatrix} \quad b = \begin{pmatrix} b_0 \\ b_1 \\ b_2 \\ \vdots \\ b_k \end{pmatrix} \tag{2.7-16}$$

As a solution of the normal equations, the vector (b) can be obtained from Eq.(2.7-17):

$$X^T \cdot X \cdot b = X^T \cdot Y \tag{2.7-17}$$

For experiments that are statistically designed, the product of the matrices (X^TX) can be inverted, for orthogonal designs it is even diagonal, and then vector (b) is calculated by applying Eq.(2.7-18):

$$b = \left(X^T \cdot X\right)^{-1} \cdot X^T \cdot Y \tag{2.7-18}$$

A definite estimation of the regression parameters b_0, b_1, ..., b_k is obtained and when C, the matrix of the inverse elements c_{ij} or information matrix, is defined as

$$C = \left(X^T \cdot X\right)^{-1} \tag{2.7-19}$$

it follows for the variances of the regression coefficients $s^2_{b_j}$:

$$s^2_{b_j} = s^2_y \cdot c_{ii} \tag{2.7-20}$$

By applying Eq.(2.7-21) the standard error of residuals can be estimated:

$$s^2_y = \frac{1}{N-k-1} \sum_{i=1}^{N} \left(y_i - b_0 - \sum_{j=1}^{k} b_j \cdot x_{ij} \right)^2 \tag{2.7-21}$$

Terms for Eqs (2.7-10) to (2.7-20):

y = response
x_i = factor or monomial-term like x_i^2 *or* $x_i x_j$
b_0 = regression constant
b_j = regression coefficient
ε = inexplicable error

Q = sum of squares
x_{ij} = values of the factors
y_i = values of the responses
$\bar{x}_j$ = mean of factors
$\bar{y}$ = mean of responses
Y = response vector
X = factor matrix
b = vector of the regression coefficients
e = vector of the experimental error under the assumption that the variances are homogeneous
X^T = transposed factor matrix.
C = matrix of the inverse elements c_{ij}
s_y^2 = variances of the residual error
c_{ij} = elements of the inverse matrix C
c_{ii} = diagonal elements of the inverse matrix C
c_{ii} = $1/\sqrt{N}$ for an orthogonal experiment
s_y^2 = variance of standard error of residuals
N = numer of value sets
k = number of factors

A detailed view on analysis of variances is provided in the relevant literature [154–156].

After these short explanations, which are intended to impart some basics on mathematical evaluation of data obtained from statistically designed experiments, a gradual procedure will be described on design, conduct, analysis and interpretation of robustness studies (Figure 2.7-11).

Amongst an increasing number of publications in this area the extensive fundamental work of Massart et al. is recommended, e.g., [159]. This work is a comprehensive guide for any analyst starting with statistical design for robustness testing of analytical methods and provides valuable information on this topic.

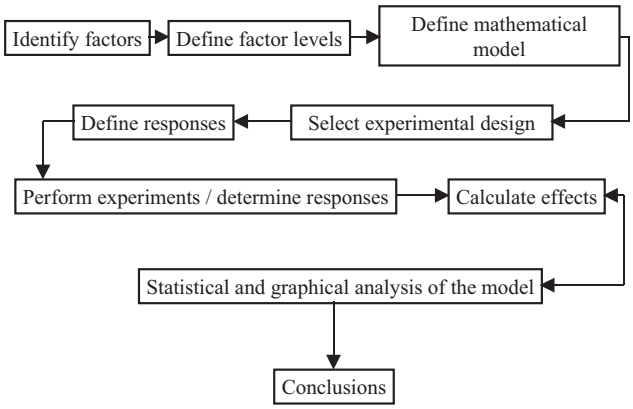


Figure 2.7-11: General procedure for statistical design, conductance, analysis and interpretation of experiments.

2.7.3.2.1 Identification of Factors and Definition of Factor Levels

Identification of factors and definition of factor levels are the first two working steps in this procedure. The factors to be examined in a robustness study are related to the analytical procedure (internal factors) and to the environmental conditions (external factors). The internal factors ensue from the description of the analytical method (operating procedure), whereas the external factors are usually not mentioned explicitly in the analytical method. Factors can be quantitative (continuous), qualitative (discrete) or they can be mixture-related. Under section 2.7.2.1 a representative selection of factors for different analytical methods has already been introduced. Certainly, this selection is not exhaustive, but it gives a picture of the factors typically tested. Of course, sophisticated sample preparation steps (extraction or filtration steps, pre- or post-column derivatisation) that may be necessary when a particular analytical method is applied, need also to be included in the robustness study.

The selected factors should be those, which are most likely to vary when a method is used daily under different conditions, and that potentially could impact on the performance (indicated by changes in the responses) of the analytical method.

Quantitative Factors

Examples of quantitative factors are the pH of a solution or the mobile phase, the temperature or the concentration of a solution, the column temperature, the buffer concentration, etc. In principle, there are different ways to enter factors in an experimental design, which may lead to information of more or less significance. Therefore, the definition of factors should be considered well. For instance, the composition of the widely used buffer $[NaH_2PO_4]/[H_3PO_4]$ can be defined in two different ways. The preferred way to prepare this buffer is to dissolve a defined amount of salt (NaH_2PO_4) and then to adjust the pH by adding the respective acid (H_3PO_4) or base (NaOH). In this case, the pH and the salt concentration (representing ionic strength μ) should be investigated as factors.

Another way to define the composition of this buffer is to prescribe the amount and the volume, respectively of its acidic (A) and its basic (B) components. The preparation of the buffer is then carried out by mixing a specified amount of NaH_2PO_4 in [g] and a certain volume of H_3PO_4 in [ml] per litre of buffer. With regard to this method of buffer preparation or the mixing of the two components, respectively, two approaches are possible to examine NaH_2PO_4 and H_3PO_4. On the one hand they can be considered as two factors (approach 1) and on the other hand they can be considered as one combined factor B/A (approach 2) representing pH or ionic strength μ. Focussing on robustness only, approach 1 might be chosen. However, it must be taken into account that, in the case where one of the two factors appears important, the other factor needs also to be controlled carefully. So, approach 1 seems to be useful only in exceptional cases.

But when detailed information is needed, it is necessary to define factors that always correspond to a clear analytical, chemical, or physical meaning. In that case approach 2 is superior to approach 1. For examination of ionic strength in accordance with approach 2, the pH is kept constant by taking care that the ratio B/A is

always unchanged, and changing the concentrations of A (H_3PO_4) and B (NaH_2PO_4) varies the ionic strength. When the ratio B/A is varied a change in pH results.

The situation becomes more complicated when a buffer system is used in which, not only one component contributes to the ionic strength μ, but a change in the B/A ratio would lead to a change in pH and ionic strength. This would be observed for a buffer, such as $[Na_2HPO_4]/[NaH_2PO_4]$.

So, from this example it becomes very clear, that the aforementioned preferred way (dissolution of the buffer salt and pH adjustment) is always the better choice. Buffer preparations that require proceedings in robustness studies according to approach 1 or approach 2 should be an exception.

Generally, factor levels are set symmetrically around the nominal level defined in the analytical procedure. The range between the upper and lower level represents the limits between which the factors are expected to vary when the analytical method is routinely applied. The decision on how to set the factor levels can be taken on the basis of the experience gained with a certain technique or procedure, which is the most common way to proceed. The selection of the levels can also be based on the precision or the uncertainty. The determination of uncertainty in analytical measurements is detailed in an EURACHEM guideline [21].

Knowing the uncertainty of an analytical measurement, it is possible to express the interval between the upper and lower level as a multiple of the uncertainty. Since the calculation of uncertainties can be time-consuming, a pragmatic alternative, is to take the last number given by a measuring instrument or to take the value specified by the manufacturer as uncertain [160]. Such numbers could be, for instance, 0.01 mg for an analytical balance, 0.05 for a pH meter or 0.1 ml for a 100.0 ml volumetric flask.

Defining the robust range of factors, it should also be considered that the analytical procedure is validated within these ranges. Consequently, variations required in the long-term application can be regarded as adjustments. Outside the validated range, we have to assume formally a change with all consequences, such as change control, revalidation, etc. (see Chapter 9).

Qualitative Factors

Qualitative factors for chromatographic methods are factors that are related to the column, such as the column manufacturer, the batch of the column (especially of the stationary phase) and also different columns of one batch. An analyst investigating qualitative factors should always remember that the absence of a significant effect does not necessarily mean that this factor never has any impact on the method performance. By testing a limited number of 'samples' (here: columns) a conclusion about the total population cannot be drawn.

Only conclusions regarding the robustness of the method with respect to the selected samples can be made.

Mixture-related Factors

Mixtures of solvents are ubiquitous in the daily use of analytical methods. Mobile phases in chromatography or buffers in electrophoresis are examples for such sol-

vent mixtures. A mixture comprised of m components, only allows m-1 components to be changed independently. Apart from the aqueous phase, the mobile phase in HPLC analysis can consist of one to three organic modifiers, resulting in mixtures of two to four components. An easy way to combine mixture-related factors and method factors (e.g., temperature, flow rate, etc.) in one experimental design is to include at maximum, m-1 components that are to be tested as factors. These m-1 factors are mathematically independent and so can be treated as method factors. Normally, the contributions of the different components in the mixture are given as volume fractions. The components can be arranged such a way that the m^{th} component is that one with the highest volume fraction and therefore, it usually serves as an adjusting component. The value of the adjusting component is calculated from the respective levels of the mixture-related factors [160].

In the case where one component is found to be relevant, then the mixture composition in total is important. Consequently, the composition of the mixture must be strictly controlled. Regarding the definition of levels for mixture-related factors the same reflections are also valid as for quantitative factors.

Adequate software packages (e.g., MODDE) guide the user through the design of the experiments, which can be very helpful, especially for studies including those mixture-related factors.

2.7.3.2.2 Mathematical Model and Experimental Design

The factors are tested by means of statistically designed experimental protocols, which are selected as functions of the number of factors to be examined. The experimental designs usually applied in robustness studies are two-level screening designs, which enable an analyst to screen a relatively large number of factors in a relatively small number of experiments. Such designs are fractional factorial or Plackett–Burman designs [122], [161–163]. In a robustness study an analyst is normally interested in the main effects of factors. For this purpose Plackett–Burman designs (PB-designs) guarantee satisfactory results. Typically, in PB-designs the two-factor interaction effects, among higher-order interaction effects, are confounded with the main effects, so that these effects cannot be evaluated separately [159, 160]. However, it has already been discussed in the literature, that two-factor interactions occurring in a robustness study can be neglected [164]. Since PB-designs are easier to build than fractional factorial designs, they became the first choice in robustness testing.

Three is the smallest number of factors to be investigated in an experimental design. Due to statistical considerations, mainly regarding the interpretation of effects, designs with less than eight experimental runs are not used, whereas those with more than twenty-four are too time-consuming [160]. For PB-designs the first lines with N = 8–24 experiments are listed below,

N = 8: + + + − + − −
N = 12: + + − + + + − − − + −
N = 16: + + + + − + − + + − − + − − −
N = 20: + + − − + + + + − + − + − − − − + + −
N = 24: + + + + + − + − + + − − + + − − + − + − − − −

where N represents the number of experiments and (+) and (–) the factor levels [163]. For construction of the complete design the following N-2 rows are obtained by shifting step-by-step each line by one position to the right. This procedure is repeated N-2 times until all but one line is formed. The last row (N^{th}) then consists of minus signs only. An equivalent procedure can be applied, when the first column of a PB-design is given. From the list above it can be derived that a PB-design can examine up to N-1 factors. It is not recommended to assign at least two columns of such a design to any factor, since these columns may indicate the magnitude of the occurring random error and the two-factor interactions [157].

The mathematical models applied for PB-designs are linear as shown in Eq.(2.7-22) for two factors and in Eq.(2.7-6) for seven factors. Besides linear models, interaction and quadratic models also play a certain role in the statistical design of experiments depending on the studies and their objectives (Eq. 2.7-5, Eq. 2.7-22).

$$\begin{aligned} &\textit{Linear model}: && y = b_0 + b_1 x_1 + b_2 x_2 \\ &\textit{Interaction model}: && y = b_0 + b_1 x_1 + b_2 x_2 + b_{12} x_1 x_2 \\ &\textit{Quadratic model}: && y = b_0 + b_1 x_1 + b_2 x_2 + b_{12} x_1 x_2 + b_{11} x_1^2 + b_{22} x_2^2 \end{aligned} \tag{2.7-22}$$

2.7.3.2.3 Definition of Responses

Generally, responses measured in a robustness study can be divided into two groups, related either to determination of a quantitative characteristic or to a qualitative characteristic. Taking HPLC as an example, this means that peak area, peak height and content are quantity-related characteristics, whilst resolution, relative retention, capacity factor, tailing factor and theoretical plates are quality-related characteristics.

2.7.3.2.4 Experiments and Determination of Responses

Before conducting the experiments in a robustness study some essential points need to be considered:

- Aliquots of the same test sample and standard (in the case of evaluating quantitative characteristics) are investigated under different experimental conditions.
- Ideally, the experiments are performed randomly.
  - If blocking, which means sorting by factors, is unavoidable due to practical reasons, a check for drift is recommended. Running experiments under nominal conditions as a function of time could perform this check.
  - Since certain designs cannot be carried out within one day, blocking by external factors not tested in the design such as, for example, days, is also allowed [160, 165].
- As already indicated for PB-designs, replicated experiments at nominal levels (centre points) conducted before, at regular time intervals between, and after the robustness study, are helpful for several reasons [160]:
  - A check of the method performance at the beginning and the end of the experiments.

  - An estimation of the pure error.
  - A first estimation of potential time effects and correction of results for possible time effects.
- Instead of correcting for time effects, sophisticated experimental designs enable an analyst to minimise time effects by confounding them with interaction effects or dummy factors (columns in a PB-design that are not assigned to any factor) [160, 165].

2.7.3.2.5 Calculation of Effects and their Statistical and Graphical Evaluation

Effects can be calculated in accordance with Eqs. (2.7-9) and (2.7-10) or with Eq. (2.7-18). An equivalent form of the equations (2.7-9) and (2.7-10) is given by Eq. (2.7-23):

$$Eff(X) = \frac{\sum Y(+)}{N/2} - \frac{\sum Y(--)}{N/2} \tag{2.7-23}$$

X = factor
$\Sigma(Y+)$ = sum of responses, where X is at the extreme level (+)
$\Sigma(Y-)$ = sum of responses, where X is at the extreme level (–)
N = number of experiments of the design.

The interpretation of effects can be done graphically and /or statistically. The graphical interpretation of important effects is typically applied with a normal probability plot [162]. The statistical interpretation is based on identifying statistically significant effects usually derived from the *t*-test statistic [166]. A more detailed description of the evaluation of statistically designed experiments is given in the relevant literature [153–155, 166]. However, some further statistical characteristics will be discussed here in conjunction with data from the following example.

2.7.3.2.6 Conclusion

At the conclusion of a statistically designed robustness study the main effects can be discussed, assessed and summarised. SST-limits (SST: System Suitability Test) can be derived from the results of a robustness test, taking the worst combinations of factor levels, which still give a satisfactory performance.

2.7.3.2.7 Example of an Experimentally Designed Robustness Study – Experimental Conduct, Interpretation of Results, Assessment and Conclusion

To allow for comparison, the study presented here has been carried out with the same HPLC method and with the same drug substance already discussed in section 2.7.3.1, even though the drug substance is only in the pre-clinical development phase and it is more meaningful to apply DOE for robustness testing in a later stage of development. The study has been planned in accordance with the procedure shown in Figure 2.7-11. The nominal conditions of the respective HPLC method are given in Table 2.7-4:

The solution used in the robustness test contained the drug substance MC at a concentration of 0.2 mg/ml (including the counter ion CI) and the related impurities SP1, SP2 (including its impurity U1) and SP3, as well as the degradation product DP1 at a concentration of 0.002 mg/ml. An analytical reference standard had

Table 2.7-4 Nominal method conditions of the robustness study.

Condition	Settings			
Apparatus:	Liquid chromatographic gradient pump system with UV/VIS detector, column thermostat, autosampler and data acquisition system.			
Column:	Merck – Purospher STAR RP18, length 125 mm, diameter 4.0 mm.			
Buffer pH 3.5:	Water de-ionised			1000 ml
	Sodium dihydrogen phosphate, anhydrous (10 mM)			1.2 g
	Phosphoric acid (85%) (for adjustment of pH)			
Mobile phaseA:	Buffer pH 3.5			900 ml
	Acetonitrile R			100 ml
Mobile phaseB:	Water de-ionised			100 ml
	Acetonitrile R			900 ml
Gradient (linear):	Time	A	B	
	0 min.	100%	0%	
	0 –15 min.	35%	65%	
	15 – 20 min.	0%	100%	
Run time:	20 min.			
Injection volume:	10 µl			
Column temperature :	+ 35 °C			
Flow:	1.0 ml/min.			
Wavelength :	227 nm			
Sample temperature :	+10 °C			

not been established for MC at that early stage of development. This study was focussed on internal factors, since external factors are generally covered by intermediate precision studies.

One qualitative (Col) and seven quantitative (pH, Conc., WL, CT, F, %BAS, %BAE) internal factors were selected. The levels defined for these factors are summarised in Table 2.7-5. They were set on the basis of technical data of the HPLC equipment used and also based on experience already gained with the DryLab-supported robustness study.

It should be noted that for the qualitative factor 'Column (batches of stationary phase material)' the nominal column was assigned to level (-1), since it is more meaningful to compare it with another one than to compare two columns that are both different from the nominal column. For quantitative factors and linear models the nominal levels can be interpolated by statistical software packages, but this is not possible for qualitative factors. Addition of a third column would require the application of a three-level design instead of a two-level design.

The experimental design and the evaluation of the data obtained were performed by means of the statistical software package MODDE.

The factors were investigated in a Plackett–Burman design for eleven factors, i.e., $N = 12$ experiments. The resolution – a term describing the degree to which estimated main effects are confounded with estimated two-level interactions, three-level interactions etc. – of such a design is III. This means that two-factor interactions could not be evaluated [167].

Table 2.7-5 Factor levels compared to nominal conditions.

#	Factor	Abbreviation	Nominal	Units	Limits	Level (–1)	Level (+1)
1	Buffer pH	pH	3.5	–	± 0.1	3.4	3.6
2	Buffer concentration	Conc	10	mM	± 2.5	7.5	12.5
3	Detection wavelength	WL	227	nm	± 2	225	229
4	Column temperature	CT	35	°C	± 3	32	38
5	Flow rate	F	1.0	ml/min	± 0.1	0.9	1.1
6	Column[a)]	Col	A	–	–	A	B
7	%B(start)[b)]	%BAS	10	%	± 1	9	11
8	%B(end)[c)]	%BAE	90	%	± 1	89	91

a) Batches of stationary phase material
b) Percentage of organic solvent in the mobile phase at the start of the gradient
c) Percentage of organic solvent in the mobile phase at the end of the gradient

However, as already discussed above, two-factor and higher order interactions in robustness studies can usually be neglected. Plackett–Burman designs are orthogonal and they are limited to linear models.

The factor correlation matrix of the orthogonal Plackett–Burman design applied is illustrated in Table 2.7-6. The value zero indicates that there is no correlation, which is expected for the factors of the robustness study described here, and unity indicates that maximal correlation is observed, which of course is the case between the factors themselves. The responses determined in this study were the critical resolutions between U1/SP1 (R_{U1_SP1}) and SP1/DP1 (R_{SP1_DP1}), the tailing factor of the main component T_{MC} and the relative peak areas of CI, MC, U1, SP1, DP1, SP2 and SP3 (%CI, %MC, %U1, %SP1, %DP1, %SP2, %SP3).

The relative peak area of CI has been included in the list of responses to gain additional information. The method discussed here only serves for a rough estimation of CI. In a later stage of development CI will not be determined by means of HPLC, but ion chromatography will be used to evaluate its content in the drug substance.

Table 2.7-6 Correlation matrix.

	pH	Conc	WL	CT	F	Col	%BAS	%BAE
pH	**1**	0	0	0	0	0	0	0
Conc	0	**1**	0	0	0	0	0	0
WL	0	0	**1**	0	0	0	0	0
CT	0	0	0	**1**	0	0	0	0
F	0	0	0	0	**1**	0	0	0
Col(B)	0	0	0	0	0	**1**	0	0
%BAS	0	0	0	0	0	0	**1**	0
%BAE	0	0	0	0	0	0	0	**1**

In addition to the 12 runs required by the selected Plackett–Burman design, three nominal experiments were conducted. For each of the 15 runs, three injections of the sample solution were carried out. Each third injection was used for calculation, provided that the second and the third injection revealed identical chromatograms. This step was taken in order to ensure that the data selected for calculation were obtained from an adequately equilibrated system. In addition, blank solutions were injected at the start and at the end of each triple injection. The run order of the experiments set by MODDE was fully randomised. However, for practical reasons the experiments have been sequenced in relation to the factors, buffer concentration and buffer pH. Furthermore, the experiments at nominal level were set to the positions 1, 8 and 15, and the whole set was finally sorted by run order as shown in the respective worksheet in Table 2.7-7, which also presents the experimental results obtained for the ten responses studied.

Fit and Review of Fit

After fitting the data shown in Table 2.7-7 by means of Multiple Linear Regression (MLR) it is helpful to have a first look on the replicates plot, which shows the responses as a function of the experiment number labels. The replicates plot provides the analyst with an idea of the experimental error, the so-called pure error, which follows from the replicated experiments at the nominal levels of the factors investigated (no. 13, 14, 15). In Figure 2.7-12 a typical example is shown for the response "Relative Peak area MC". The numbers 13, 14, and 15 indicate good repeatability (reproducibility in the sense of the MODDE terminology) and a small pure error. Besides such typical examples occasional examples with smaller and larger errors were also found in this investigation. Such findings were obtained for the peak areas of SP3 and U1 and also for the peak resolution between U1 and SP1.

It should be noticed that, apart from the replicates in Figure 2.7-12, two groups of response values can be observed, which are correlated to measurements at two different detection wavelengths (225 nm and 229 nm).

The coefficients of the model and the effects, which the factors have on the different responses will be described below. Before this, a check of the summary of fit, which is shown in Figure 2.7-13 is necessary. Such a plot provides an overview of the characteristics 'R^2, Q^2, Model Validity and Reproducibility'. With these characteristics the analyst can assess how good is the fit for each response [155, 156, 158, 168–170].

R^2 is the percentage of the *variation of the response given by the model*. It is a measure of fit and demonstrates how well the model fits the data. A large R^2 is a necessary condition for a good model, but it is not sufficient. Even poor models (models that cannot predict) can exhibit a large R^2. However, a low value of R^2 will be obtained in case of poor 'reproducibility' (poor control over the experimental error) or poor model validity (the model is incorrect). If R^2 is 1 the model fits the data perfectly. Q^2 is the percentage of the *variation of the response predicted by the model* according to cross-validation. Q^2 tells an analyst how well the model predicts new data. A useful model should have a large Q^2. A low Q^2 indicates poor 'reproducibility' (poor control over the experimental error) and/or poor model validity (the model

Table 2.7-7 Worksheet – Robustness study on eight factors; ten responses were monitored; experiments were performed in accordance with the run order: 1–7 (day 1), 8–11 (day 2), 12–15 (day 3).

Exp No	Exp Name	Run Order	Buffer pH	Buffer concentration	Detection wavelength	Column temperature	Flow rate	Column batch	%B at the start	%B at the end	Tailing factor MC
13	N13	1	3.5	10	227	35	1	A	10	90	1.48
8	N8	2	3.4	7.5	229	38	1.1	A	11	91	1.61
9	N9	3	3.4	7.5	225	38	1.1	B	9	91	1.67
12	N12	4	3.4	7.5	225	32	0.9	A	9	89	1.56
3	N3	5	3.4	12.5	229	32	1.1	A	9	89	1.45
7	N7	6	3.4	12.5	229	38	0.9	B	11	89	1.54
11	N11	7	3.4	12.5	225	32	0.9	B	11	91	1.51
14	N14	8	3.5	10	227	35	1	A	10	90	1.49
1	N1	9	3.6	7.5	229	32	0.9	A	11	91	1.55
4	N4	10	3.6	7.5	229	38	0.9	B	9	89	1.61
10	N10	11	3.6	7.5	225	32	1.1	B	11	89	1.65
2	N2	12	3.6	12.5	225	38	0.9	A	9	91	1.43
5	N5	13	3.6	12.5	225	38	1.1	A	11	89	1.5
6	N6	14	3.6	12.5	229	32	1.1	B	9	91	1.51
15	N15	15	3.5	10	227	35	1	A	10	90	1.49

Table 2.7-7 Continued.

Exp No	Exp Name	Run Order	Relative peak area CI	Relative peak area MC	Relative peak area SP1	Relative peak area DP1	Relative peak area SP2	Relative peak area SP3	Relative peak area U1	Resolution U1_SP1	Resolution SP1_DP1
13	N13	1	33.651	61.896	1.615	1.135	0.687	0.831	0.184	2.35	2.29
8	N8	2	29.884	66.001	1.141	1.218	0.724	0.833	0.198	1.51	2.97
9	N9	3	37.134	57.865	2.186	1.067	0.706	0.848	0.194	1.45	3.01
12	N12	4	37.145	57.801	2.212	1.086	0.699	0.857	0.2	1.38	2.07
3	N3	5	29.975	65.953	1.107	1.222	0.723	0.825	0.194	1.9	2.27
7	N7	6	30.062	65.866	1.116	1.201	0.729	0.84	0.184	1.69	2.63
11	N11	7	37.153	57.832	2.185	1.09	0.701	0.854	0.184	1.75	2.01
14	N14	8	33.812	61.683	1.63	1.145	0.705	0.835	0.191	2.35	2.27
1	N1	9	30.143	65.789	1.114	1.219	0.709	0.847	0.179	2.66	1.82
4	N4	10	30.213	65.719	1.115	1.213	0.711	0.852	0.177	2.59	2.32
10	N10	11	37.152	57.941	2.134	1.075	0.685	0.838	0.174	2.98	2.11
2	N2	12	37.229	57.867	2.133	1.075	0.663	0.845	0.188	3.06	2.13
5	N5	13	37.088	58.044	2.126	1.073	0.649	0.843	0.177	3.48	2.54
6	N6	14	30.243	65.705	1.122	1.218	0.709	0.818	0.185	3.31	1.91
15	N15	15	33.891	61.621	1.621	1.142	0.699	0.835	0.191	2.34	2.29

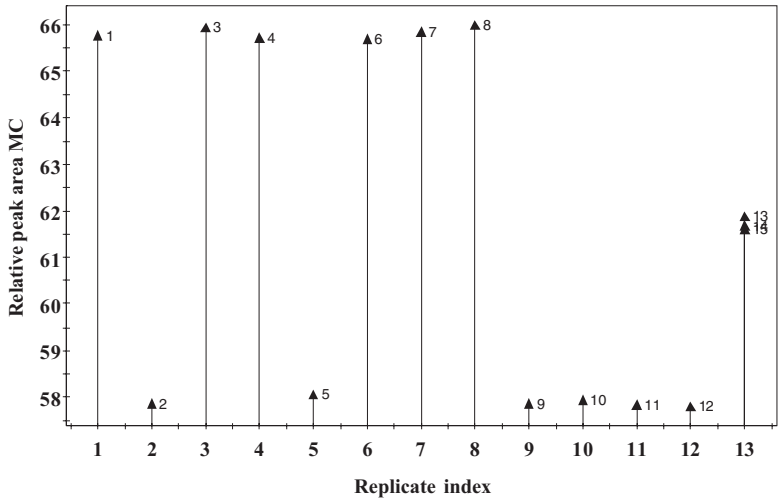


Figure 2.7-12: Replicates plot for the relative peak area of the main component MC (relative peak area MC as a function of the experiment number labels). The different levels indicate the absorptions at 225 nm, 229 nm and at the nominal level of 227 nm.

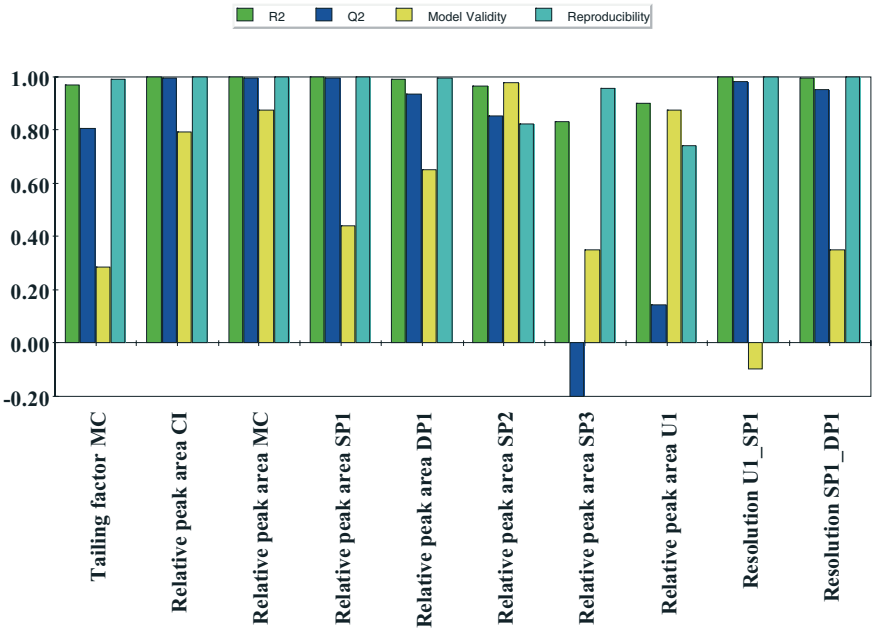


Figure 2.7-13: Summary of fit for all responses defined in this robustness study. R^2 is the percentage of the variation of the response explained by the model. Q^2 is the percentage of the variation of the response predicted by the model. The model validity measures the error of fit. Reproducibility is a comparison of the variation of the response under the same conditions with the total variation of the response.

is incorrect). Assuming that there is a good R^2, moderate model validity, and a design with many degrees of freedom of the residuals, then a low Q^2 is usually due to insignificant terms in the model. Such insignificant terms might be removed from the model [168]. The *model validity* measures the error of fit and compares it with the pure error. If the model validity bar is larger than 0.25, there is no *lack of fit* of the model [168]. This means that the model error is in the same range as the pure error ('reproducibility'). A model validity bar of unity represents a perfect model. When the model validity is below 0.25 a significant lack of fit exists. This indicates that the model error is significantly larger than the pure error. There are many parameters that could cause a lack of fit and therefore poor model validity. However, in many cases the cause is artificial and can simply be a very low pure error that tends to zero [168]. *'Reproducibility'* is a comparison of the variation of the response under the same conditions, with the total variation of the response. The variation of the response under the same conditions corresponds to the pure error and it is often determined at centre points. When the reproducibility is unity, the pure error is zero. This means that, under the same conditions, the values of the response are identical. When the reproducibility bar is zero, the pure error equals the total variation of the response [168]. It must be noted that 'reproducibility' is here used according to MODDE terminology, not as a precision level.

In addition to the explanations above it has to be mentioned that for robustness studies a large R^2 is not necessarily needed. This depends on the study itself and especially on the range between the lower and upper factor levels. When the pure error is small a small R^2 is also sufficient. However, Q^2 should then be positive and not much smaller than R^2.

A small R^2 may just indicate that the applied model does not fit the variation of the responses very well. This could simply mean that the method is not sensitive to changes and therefore it is robust (best case in robustness testing, i. e. insignificant model and results within specification!) [168].

If a large R^2 is obtained in a robustness study this indicates strong correlation and normally that the method is sensitive to changes and that therefore the respective factors have to be carefully controlled. To decide then whether the method is robust or not depends on the range over which the response varies. When the response varies in a range that is not critical, the method nevertheless might be regarded as robust (second best case in robustness testing, i. e. significant model and results within specification!) [168].

The data of Figure 2.7-13 implies that seven out of ten responses show nearly perfect goodness-of-fit results. R^2 and Q^2 values are above 0.8, reproducibilities are above 0.9 and the values for model validity are above 0.25. However, for the three responses 'Relative Peak Area SP3', 'Relative Peak Area U1' and 'Resolution U1_SP1' smaller values were obtained. For the relative peak areas of SP3 and U1 the Q^2 alues are too small and for the resolution between U1 and SP1 the model validity term is too small. According to the remarks on the statistical characteristics above, there are reasonable explanations for these findings. The small Q^2 values of the relative peak areas of SP3 and U1 can be explained by the poor 'reproducibility' (compared to the others) and mainly by the coefficients that appear to have little sig-

nificance or are even insignificant (not relevant) for these responses. The poor model validity term in the case of the resolution between U1 and SP1 is simply explained by the extremely high reproducibility (0.99993) corresponding to a pure error that tends to zero. Therefore, these findings are not relevant and can be neglected.

Diagnostics

Besides the interpretation of the summary of fit, a further evaluation step is necessary before starting the calculation of coefficients and effects. It has to be checked whether the residuals are random and normally distributed. For this purpose the normal probability plot of residuals is usually applied using a double logarithmic scale. Such a plot allows the identification of outliers and the assessment of normality of the residuals.

If the residuals are random and normally distributed, then they lie on a straight line between –4 and +4 studentised standard deviation, whilst outliers lie outside the range of –4 to +4 standard deviation [168]. Figure 2.7-14 illustrates a typical normal probability plot obtained in this study (tailing factor of MC).

For seven out of the ten responses measured, the normal probability plots corresponded to Figure 2.7-14 indicating random and normally distributed residuals. However, three exceptions were observed. These were the "Relative Peak Area SP1", the "Resolution U1_SP1" and the "Resolution SP1_DP1", corresponding to the experimental runs N12, N3 and N6, and N1 and N9, respectively.

These outliers may be statistically significant but they are certainly not relevant, which is demonstrated in Figure 2.7-15, which shows the linear relationship between observed and predicted data for the resolution of U1 and SP1. It can be clearly seen that the difference between N3 or N6 and the straight line is marginal and therefore not of relevance.

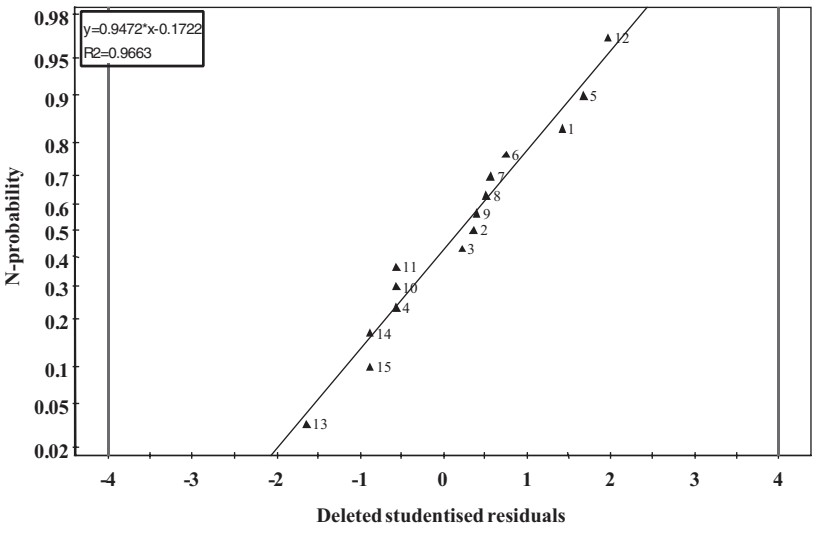


Figure 2.7-14: Normal probability plot for the tailing factor of the main component.

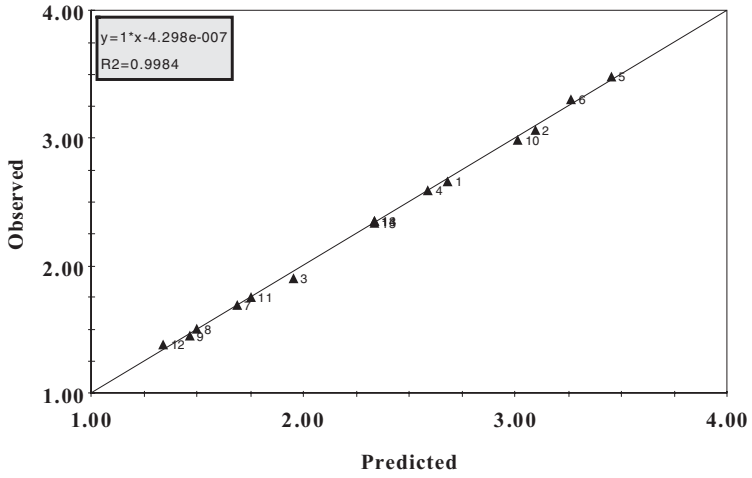


Figure 2.7-15: Observed versus predicted data for response 'Resolution U1_SP1' and each experiment.

In addition to the summary of fit, the test on the normal distribution, and the comparison of observed and predicted data also an the analysis of variances (ANOVA) was performed, which revealed that the standard deviation of the regression was larger than the standard deviation of the residuals with its upper confidence level. Thus, the results discussed above were also confirmed by ANOVA. Furthermore, no relevant lack of fit was detected [168].

Interpretation of the Model

As it was clear that the model was correct, the coefficients and effects of the different factors could then be evaluated. It is advisable to construct the coefficient overview plot (Fig. 2.7-16), which displays the coefficients for all responses defined in the study.

Normally the different responses have different ranges. In order to make the coefficients comparable, they are normalised by dividing the coefficients by the standard deviation of their respective response. So, this plot allows an analyst to see how the factors affect all the responses. From Figure 2.7-16 it is obvious that in case of the relative peak areas the impact of the detection wavelength is really dominant, which is due to the almost identical absorption properties of the compounds under investigation. This is illustrated by the section of the uv/vis-spectrum of MC where there appears to be no absorption maximum at the nominal detection wavelength of 227 nm (Fig. 2.7-17).

The exception among the known related impurities for dominating detection wavelength is the relative peak area of SP3, which seems to be also affected by the mobile phase flow rate and the buffer concentration. For the resolution between U1 and SP1 the dominating factor is the pH, and for the resolution between SP1 and DP1 the column temperature and the pH are most significant (relevant). In case of the tailing factor, the buffer concentration and the batch of the stationary phase are most important.

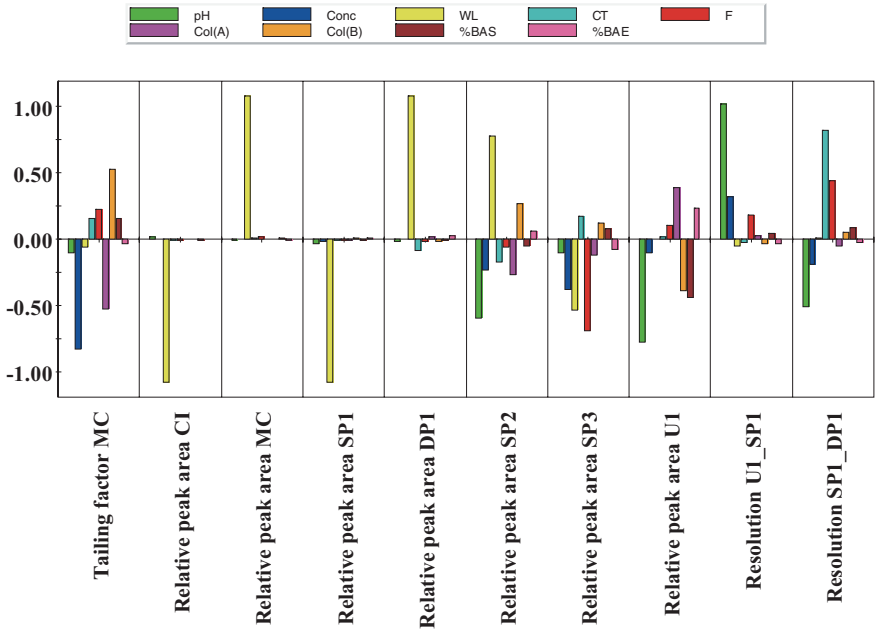


Figure 2.7-16: Coefficient overview plot for all responses defined in the study (all factor settings for qualitative factors included). The factors are colour-coded as shown in the legend above (see also Table 2.7-5).

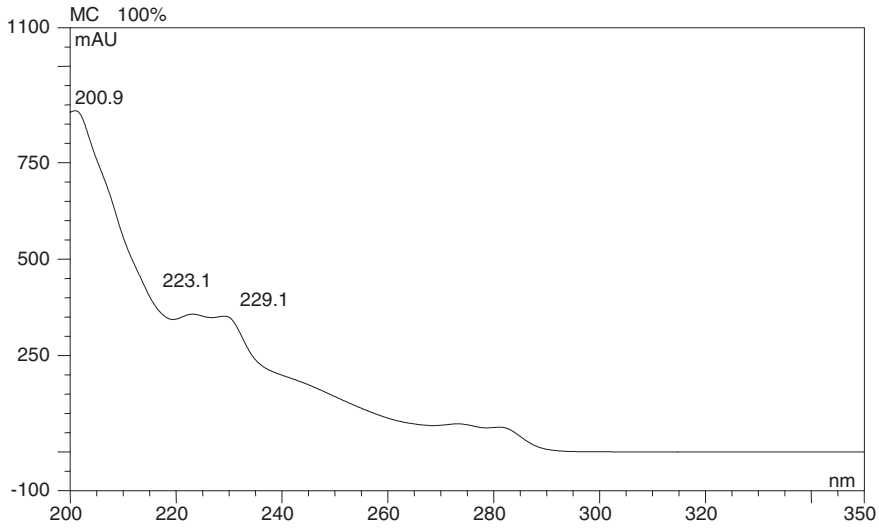


Figure 2.7-17: Section of the uv/vis-spectrum of the main component MC. Absorption [m AU] vs. wavelength [nm]

For a more detailed analysis of the extent to which the examined factors influence the responses, coefficient plots are very useful. Such coefficient plots display the regression (MLR) coefficients with confidence intervals and they are applied to interpret the coefficients. Normally, the coefficient plot is for data that are centred and scaled to make the coefficients comparable.

The size of a coefficient represents the change in the response when a factor is varied from zero to unity (in coded units), while the other factors are kept at their averages. A coefficient is significant (different from the noise), when the confidence interval does not include zero. For Plackett–Burman designs, the coefficients are half the size of the effects.

A typical coefficient plot obtained in this robustness study is shown in Figure 2.7-18 for the resolution SP1_DP1. The pH, the buffer concentration, the column temperature and the flow rate are the significant (relevant) factors impacting the peak resolution between SP1 and DP1.

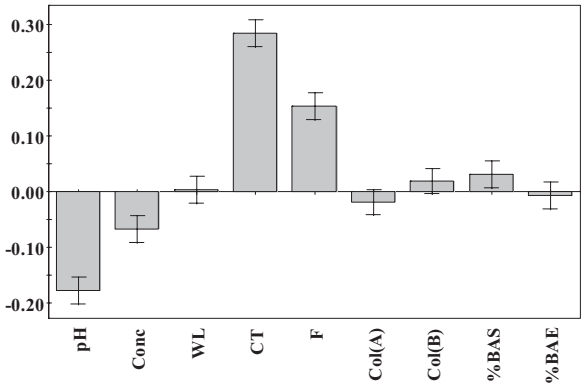


Figure 2.7-18: Coefficient plot for the resolution between SP1 and DP1; scaled and centred data including all settings for the qualitative factor Col (batch of stationary phase material).

The significant factors influencing the responses are summarised in Table 2.7-8.

Table 2.7-8 Overview of the responses and their significant factors derived from coefficient plots, or from the coefficient list calculated by MODDE.

Response	Significant Factor(s)
Tailing Factor	Conc, Col, F*
Relative Peak Area CI	WL
Relative Peak Area MC	WL
Relative Peak Area U1	pH, %BAS*, Col*
Relative Peak Area SP1	WL, pH*
Relative Peak Area DP1	WL
Relative Peak Area SP2	WL, pH, Col*, Conc*
Relative Peak Area SP3	F, WL*
Resolution U1_SP1	pH, Conc, F, WL*
Resolution SP1_DP1	CT, pH, F, Conc, %BAS*

*: To a minor extent.

A further way to visualise the effects of the factors of a robustness study is by examination of the 'effect plot'. This plot is especially useful for screening designs such as Plackett–Burman. In this plot the effects are ranked from the largest to the smallest.

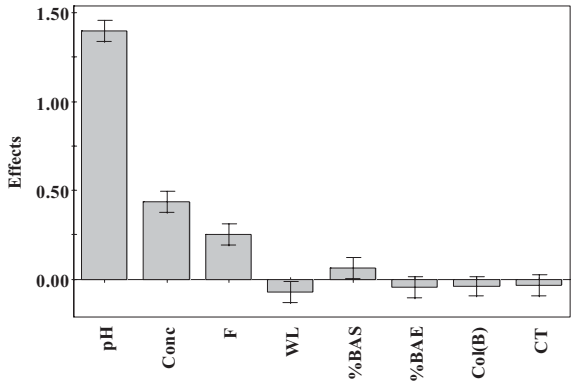


Figure 2.7-19: Effect plot for response 'Resolution U1_SP1'.

The effect plot displays the change in the response when a factor varies from its low level to its high level. All other factors are kept at their averages. It should be noted that the effects are twice the coefficients, as the coefficients are the change in the response when the factors vary from the average to the high level. Insignificant (not relevant) effects are those where the confidence interval includes zero. Small effects are those which are of minor importance (they affect to only a small extent). In Figure 2.7-19 an effect plot is shown for the resolution between the peaks U1 and SP1. This illustrates that the pH is the most important factor for the peak resolution between U1 and SP1. The buffer concentration and the flow rate also play a certain role but all the other factors are insignificant. In other words, they are not relevant and can be neglected. For this study, Figure 2.7-19 is a representative example of an effect plot. All effects obtained in this robustness testing are summarised in Table 2.7-9.

Table 2.7-9 List of effects obtained in this study.

Tailing Factor MC	Effect	Conf. int (±)	Relative Peak Area CI	Effect	Conf. int (±)
Conc	–0.118333	0.0276095	**WL**	–7.0635	0.199474
Col(B)	0.0750001	0.0252039	**pH**	0.119165	0.199475
F	0.0316668	0.0276095	**F**	–0.0781655	0.199474
CT	0.0216668	0.0276095	**%BAS**	–0.0761682	0.199475
%BAS	0.0216666	0.0276095	**CT**	–0.0334971	0.199474
pH	–0.0150002	0.0276095	**%BAE**	0.0251666	0.199475
WL	–0.00833339	0.0276095	**Conc**	0.0131673	0.199474
%BAE	–0.00500001	0.0276095	**Col(B)**	0.0130576	0.182094

Table 2.7-9 Continued.

Relative Peak Area MC	Effect	Conf. int (±)	Relative Peak Area SP1	Effect	Conf. int (±)
WL	7.94717	0.192223	WL	–1.0435	0.026238
F	0.105836	0.192223	pH	–0.0338335	0.026238
%BAS	0.0938276	0.192224	Conc	–0.0188333	0.026238
CT	0.0568372	0.192223	%BAE	0.0118334	0.026238
%BAE	–0.0441673	0.192224	%BAS	–0.00983348	0.026238
pH	–0.0421727	0.192224	F	–0.00983327	0.026238
Col(B)	–0.029217	0.175475	Col(B)	0.00977797	0.0239519
Conc	0.0251697	0.192223	CT	–0.00949982	0.026238
Relative Peak Area DP1	**Effect**	**Conf. int (±)**	**Relative Peak Area SP2**	**Effect**	**Conf. int (±)**
WL	0.1375	0.0114332	WL	0.0336667	0.00878793
CT	–0.0104999	0.0114332	pH	–0.0260001	0.00878794
%BAE	0.00283334	0.0114332	Col(B)	0.0115	0.00802225
Col(B)	–0.00211096	0.0104371	Conc	–0.00999996	0.00878793
pH	–0.00183346	0.0114332	CT	–0.0073333	0.00878793
F	–0.0018333	0.0114332	F	–0.00266667	0.00878793
%BAS	–0.00083341	0.0114332	%BAE	0.00266664	0.00878794
Conc	0.000166742	0.0114332	%BAS	–0.0023334	0.00878794
Relative Peak Area SP3	**Effect**	**Conf. int (±)**	**Relative Peak Area U1**	**Effect**	**Conf. int (±)**
F	–0.015	0.009622	pH	–0.0123334	0.00541719
WL	–0.0116667	0.009622	%BAS	–0.00700002	0.00541719
Conc	–0.00833331	0.009622	Col(B)	–0.00611109	0.0049452
CT	0.00366671	0.009622	%BAE	0.00366666	0.00541719
Col(B)	0.00266677	0.00878364	F	0.00166667	0.00541719
pH	–0.00233344	0.009622	Conc	–0.00166666	0.00541719
%BAE	–0.00166667	0.009622	CT	0.00033335	0.00541719
%BAS	0.0016666	0.009622	WL	–6.68E-04	0.00541719
Resolution U1_SP1	**Effect**	**Conf. int (±)**	**Resolution SP1_DP1**	**Effect**	**Conf. int(±)**
pH	1.4	0.059333	CT	0.568334	0.0481869
Conc	0.436667	0.059333	pH	–0.355	0.0481869
F	0.25	0.059333	F	0.305	0.0481869
WL	–0.0733336	0.059333	Conc	–0.135	0.0481869
%BAS	0.0633334	0.059333	%BAS	0.0616665	0.0481869
%BAE	–0.0466668	0.059333	Col(B)	0.0372224	0.0439884
Col(B)	–0.0416666	0.0541633	%BAE	–0.015	0.0481869
CT	–0.0333331	0.059333	WL	0.00833327	0.0481869

One way to illustrate the impact of a particular factor of a certain response is an examination of the main effect plot. For screening designs, the main effect plot displays the fitted values of the response with the confidence interval at the low and high value of the selected factor, and at the centre point. The other factors are kept at their average values.

Interesting main effect plots are depicted in Figure 2.7-20 and Figure 2.7-21 where the contrary effects of pH on the resolution of the peak pairs U1 / SP1 and SP1 / DP1 are obvious.

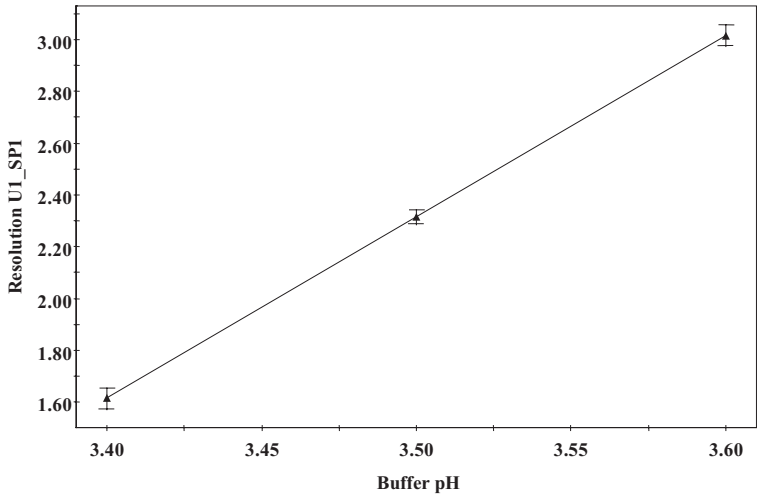


Figure 2.7-20: Main effect plot of the response Resolution U1_SP1 as a function of the factor pH.

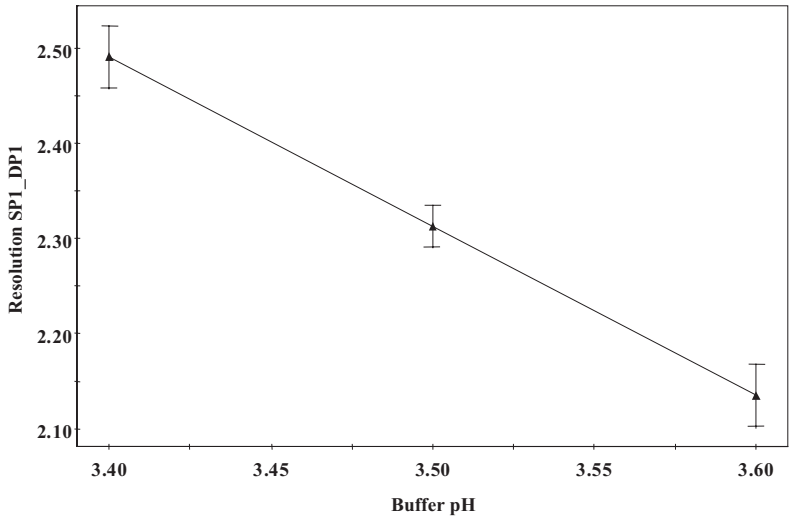


Figure 2.7-21: Main effect plot of the response Resolution SP1_DP1 as a function of the factor pH.

Use of the Model

After interpretation, the model can be applied to predictions. For that purpose MODDE provides several features that are useful for robustness studies, such as the Prediction Plot, the Response Prediction Plot, the Contour Plot, and the so-called SweetSpot Plot [168].

The Prediction Plot shows a two-dimensional overlay of all responses, each as a function of one factor. The other factors can be set to the levels LOW, CENTRE, HIGH and CUSTOM. The Response Prediction Plot illustrates the functional dependency of a certain response on one factor that is varied, while the other factors are kept at their centre levels. Additionally, the respective confidence interval is shown for each plot. Confidence levels of 90 %, 95 % or 99 % can be chosen. The Prediction Plot and the Response Prediction Plot are both tools that help the analyst to assess the robustness of an analytical method. However, the more powerful tools are the Contour Plot and, especially the SweetSpot Plot. The Contour Plot can be compared to the three-dimensional resolution maps that are calculated by DryLab. It presents a certain magnitude of the response (colour-coded range of the response), as a function of two factors shown on the x- and the y-axis. The other factors are usually set to their centre levels. Typical examples obtained in this study are shown in Figure 2.7-22 and Figure 2.7-23.

In Figure 2.7-22 the functional dependency of the relative peak area of the main component MC (%MC) on the buffer concentration and the detection wavelength, is shown. All the other factors are kept at their centre levels. The qualitative factor Col (batch of the stationary phase) was set to batch B. From the graph it can be seen that %MC is only influenced by the detection wavelength, whilst the impact of the buffer concentration can be neglected.

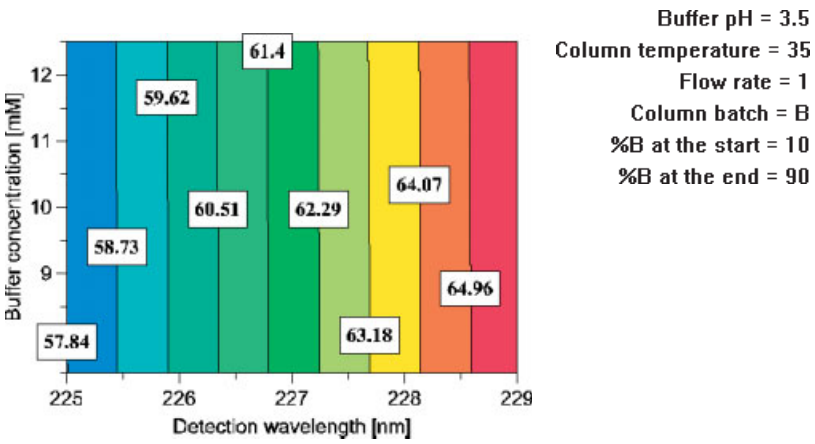


Figure 2.7-22: Contour Plot of the relative peak area of MC as a function of the buffer concentration and the detection wavelength; all other factors are kept at the centre levels; columns stationary phase material B selected.

In Figure 2.7-23 the resolution between the peaks of U1 and SP1 is given as a function of the buffer concentration and the buffer pH. It is demonstrated that both factors impact on the resolution. The higher the value of the factors, the higher is the resolution between U1 and SP1. Certainly, the effect of the buffer pH is more pronounced than the effect of the buffer concentration.

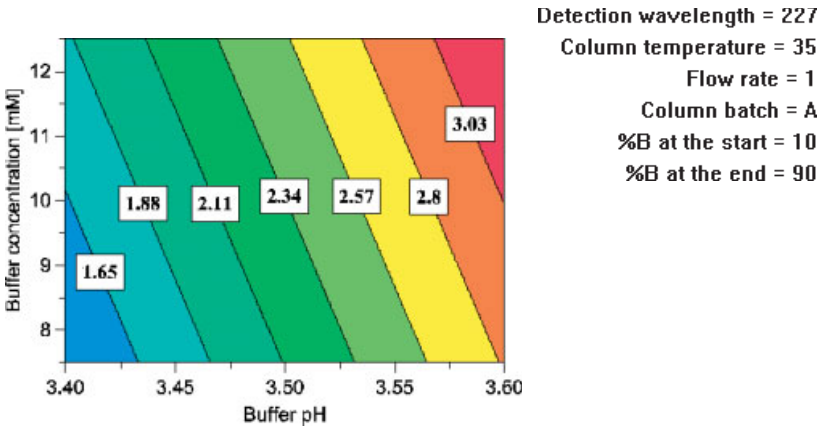


Figure 2.7-23: Contour Plot of the resolution between U1 and SP1 as a function of the buffer concentration and the buffer pH; all other factors are kept at the centre levels; columns stationary phase material A selected.

The SweetSpot Plot is a very powerful feature for assessing the robustness of an analytical method. Before creating a SweetSpot Plot the requirements to be fulfilled by a certain analytical method to ensure its reliable and accurate performance, should be considered. These requirements are those defined for the responses. The SweetSpot Plot is a three-dimensional graph that is similar to the Contour Plots and the DryLab resolution maps. The third dimension (z-axis) is colour-coded and visualises the regions where all or none of the requirements are met and the first (x-axis) and second (y-axis) dimension represent two factors. The other factors are held constant at their levels LOW, CENTRE, HIGH or CUSTOM. For calculating the regions where all or none of the requirements are met, MODDE uses a Nelder Mead simplex method [168, 171]. With respect to the HPLC robustness study discussed here, the following aspects were considered. The study was conducted with a development compound at the pre-clinical phase and – as it has already been remarked above – analytical reference standards were not available at this early stage of development. Therefore, relative peak areas have been included as responses in the design. Since the definition of requirements for relative peak areas is not very useful, the respective results obtained in this study were assessed qualitatively and they will be considered in robustness studies that will be conducted at a later stage of development. In such studies the relative peak areas will be replaced by the assay (calculated versus external reference standard(s)), i.e., the contents of the main component and the related impurities and then acceptance criteria will definitely be required. Conse-

quently, requirements were only set for the responses, tailing factor of the main component and peak resolutions between U1 and SP1 as well as between SP1 and DP1. Taking into account the experimental data obtained, a tailing factor of 1.7 (MC belongs to a compound class that always tends to a slight tailing) was considered as the upper limit. The range defined for calculation was 0.8 – 1.7 ($0.8 \leq T_{MC} \leq 1.7$). For the peak resolutions a minimum value of 1.8 was set ($R_{U1\_SP1} \geq 1.8$, $R_{SP1\_DP1} \geq 1.8$). The x- and y-axis in such a plot should represent the factors that appeared to have the most significant impact on the responses, which were the factors buffer concentration and buffer pH. Figure 2.7-24 illustrates a SweetSpot Plot obtained under those conditions. Besides the factors buffer concentration and buffer pH, the factors were kept at the level CENTRE. The stationary phase material was batch A.

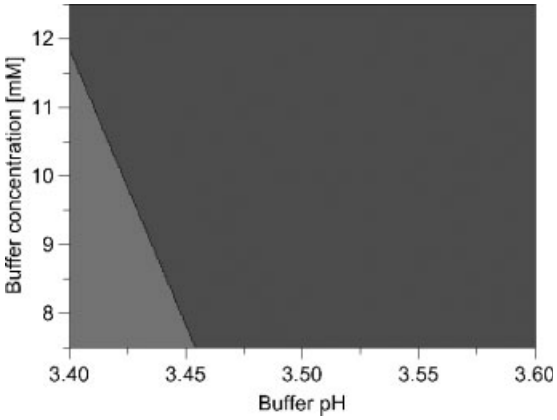


Figure 2.7-24: SweetSpot Plot – T_{MC}, R_{U1_SP1}, R_{SP1_DP1} vs. buffer pH and buffer concentration; SweetSpot: $0.8 \leq T_{MC} \leq 1.7$, $R_{U1\_SP1} \geq 1.8$, $R_{SP1\_DP1} \geq 1.8$; CENTRE level for other factors; batch A of stationary phase material. Dark grey indicates that all criteria are met. Light grey represents the area where only two criteria are fulfilled ($1.4 < R_{U1\_SP1} < 1.8$).

In Figure 2.7-24 the large dark grey area in the graph indicates the SweetSpot area, which is the area where all criteria are fulfilled. The light grey area is where two criteria are met ($1.4 < R_{U1\_SP1} < 1.8$). The black and white areas are those where only one or none of the criteria is met. However, these cases were not observed in this study. Even if the factors that were set at the CENTRE level in Figure 2.7-24 are varied between their settings LOW and HIGH the plot does not significantly change. Therefore, the SweetSpot Plot shows on the one hand that the method is robust and on the other hand that the requirements defined for T_{MC}, R_{U1_SP1}, and R_{SP1_DP1} can be adopted for the system suitability test.

However, for robustness studies at a later stage of drug development it is advisable to perform experiments at the borderline (between the light and dark grey area in Figure 2.7-24) in order to confirm that the predictions obtained fit with the experimental data.

Once a robust range of factors is found, it should also be considered that the analytical procedure is validated within these ranges. Consequently, variations required in the long-term application can be regarded as adjustments. Outside the validated range, it has to be assumed formally that there is a change with all consequences, such as change control, revalidation, etc. (see Chapter 9).

Conclusion

The example discussed in this section demonstrates that DOE software, such as MODDE can be a very powerful tool in assessing robustness of analytical methods. In this study eight factors each at two levels were examined, which means that for a complete evaluation $2^8 = 256$ experiments would have been needed. But by means of a Plackett–Burman design consisting of 12 experimental runs plus three runs at the centre point, the scope could be reduced to 15 experiments without loss of information relevant for the evaluation of the robustness of the analytical method studied. From this example it follows that by using experimental design, the savings of working time, resources and costs can be enormous.

Probably, mainly due to that fact DOE has become more and more popular in HPLC and Capillary Electrophoresis during recent years, both in the area of robustness testing of analytical methods [172–174] and also in analytical method development [175–177].

Acknowledgement

I would like to thank Ralf Hohl for his involvement and care in performing the experimental work of this robustness study, for his helpful and interesting comments, and also for the inspiring discussions.

I also would like to express my thanks to Prof. Dr. Andreas Orth for reviewing the section on experimental design, for the instructive discussions, and for his invaluable advice on the design of experiments.

2.8 System Suitability Tests

John H. McB. Miller

2.8.1 Introduction

Once an analytical procedure has been validated the method can be transferred for routine use. However, to do so and to verify the validation status each time the analytical procedure is applied, system suitability tests are to be included in the written procedure. These tests are to be developed during the validation of the analytical method and are to measure parameters, which are critical to the conduct of the method, and the limits set should ensure the adequate performance of the analytical procedure. In the pharmacopoeias, general methods, when referenced in an individual monograph, have been demonstrated to be applicable to that substance. Revalidation is not required but only compliance to any prescribed system suitability criteria. These requirements are usually given in the general method unless otherwise prescribed in the individual monograph. System suitability criteria are limits applied to various analytical procedures designed to ensure their adequate performance. These criteria are to be fulfilled before and/or during the analyses of the samples. Failure to comply with system suitability criteria during an analytical run will render the results obtained un-useable and an investigation into the cause of the poor performance must be undertaken. Corrective action is to be taken before continuing the analysis. Compliance with the criteria of the system suitability tests will ensure transferability of the method and increase the reliability of the results obtained by better control of the procedure.

2.8.2 Non-chromatographic Techniques

In this section, some examples of system suitability tests to be applied to various techniques will be described, but emphasis will be given to chromatographic separation techniques in the next section. These are only some examples of system suitability criteria set for a number of different analytical procedures. All analytical methods should have performance indicators built into the procedure. This is particularly true for separation techniques, which have so many variables.

2.8.2.1 Infra-red Spectrometry

In the preparation of an infrared spectrum for identification there are two system suitability criteria which must be met in order to obtain a spectrum of adequate quality to permit interpretation. The criteria [178] are that the transmittance at 2000 cm^{-1} is at least 70 percent and that transmittance of the most intense band is not less than 5 percent.

2.8.2.2 Coulometric Micro-determination of Water

In the method for the micro-determination of water [179] by a coulometry system, suitability criteria are given for accuracy depending on the content of water in the sample. Water or a standard solution of water, at approximately the same amount as expected in the sampl,e is added to the apparatus and determined. The recoveries should be in the ranges of 97.5–102.5 per cent and 90.0–110.0 per cent for the additions of 1000 µg and 100 µg of water, respectively. The recovery experiment is to be performed between two successive sample titrations.

2.8.2.3 Heavy Metal Test

Recent proposals [180] for the conduct of the Heavy Metals in the European Pharmacopoeia test have included a monitor solution as a system suitability requirement. The test is invalid if the reference solution does not show a brown colour compared to the blank or if the monitor solution is at least as intense as the reference solution. The monitor solution is prepared as for the test solution but with the addition of the prescribed volume of the lead standard solution. Thus, the approaches taken by both the United States Pharmacopoeia (USP) and the European Pharmacopoeia are similar for this type of test.

2.8.2.4 Atomic Absorption Spectrometry

Recent proposals [181] for the revision of the general chapter on Atomic Absorption Spectroscopy have also included system suitability criteria for sensitivity, accuracy, precision and linearity. The sensitivity of the method is set by assuming the absorbance signal obtained with the most dilute reference solution must at least comply with the instrument sensitivity specification. The accuracy (or recovery) is between 80.0 and 120 percent of the theoretical value. The recovery is to be determined on a suitable reference solution (blank or matrix solution) to which a known quantity of analyte (middle concentration of the validation range) is added. Alternatively, an appropriate certified reference material is to be used. The repeatability is acceptable if the relative standard deviation at the test concentration is not greater than 3 percent. When calibration is performed by linear regression then the correlation coefficient is to be equal to or greater than 0.97.

2.8.2.1 Volumetric Solutions

System suitability criteria are set in the section for volumetric solutions [182] where it is stated that the prescribed strength of the titrant to be used is to be within ± 10 percent of the nominal strength and the repeatability (relative standard deviation) of the determined molarity does not exceed 0.2 percent.

2.8.3 Separation Techniques

Given the importance of separation techniques in modern pharmaceutical analysis, the bulk of this section is focussed on liquid chromatographic system suitability criteria and in particular selectivity. The system suitability criteria [15] applied to sepa-

ration techniques for assay, include selectivity, symmetry, repeatability to which is added a requirement for sensitivity when performing a test for related substances (organic impurities). In general terms selectivity and efficiency are related to the stationary and mobile phases, whilst sensitivity and precision are principally limited to the performances of the injector and detector. However, it is evident that selectivity and efficiency also contribute to sensitivity and precision.

2.8.3.1 Selectivity

A number of different approaches are valid to ensure the selectivity of the method. The column performance (apparent efficiency) with regard to selectivity may be calculated as the apparent number of theoretical plates (N).

$$N = 5.54\left(\frac{t_R}{w_R}\right)^2 \tag{2.8-1}$$

t_R = retention time (or volume or distance) along the baseline from the point of injection to the perpendicular dropped from the maximum of the peak corresponding to the analyte.

w_R = width of the peak at half-height.

The disadvantage of this method of indicating selectivity is that it varies depending on the stationary phase employed and the retention time of the analyte. It also varies with the usage or the extent of use of the stationary phase so this term is not very reliable as a measure of the separation to be achieved. The variation in apparent efficiency with column type and age of the column is shown in Table 2.8-1.

Table 2.8-1 Variation in column performance with usage.

Spherisorb ODS-2			Hypersil ODS			Kromasil C-18			Nucleosil C-18		
Rs	Rt	N/m	Rs	Rt	N/m	Rs	Rt	N/m	Rs	Rt	N/m
12.64	6.70	67492	11.20	7.78	59484	18.4		74486	14.7	7.8	90372
10.67	6.31	67360	11.56	7.83	62888	17.3	12.1	63716	13.7	8.7	94324
12.54	6.67	45152	10.17	6.41	61828	17.1	10.7	71580	16.5	8.7	100400
10.10	6.65	40588	11.88	7.81	62500	15.1	10.3	63324	15.8	8.3	100056
9.70	6.78	72824*	10.18	6.19	66350	16.1	10.1	67844	12.9	7.6	83308
12.60	6.62	76432	11.28	6.45	77312	17.1	10.4	74548	15.4	9.5	91920
			9.72	5.95	67326	18.2	10.8	77168	11.4	7.4	65224
			10.18	6.69	62148	17.0	10.2	71904	15.8	9.6	101820
			10.07	6.48	63072	17.6	10.5	77248	13.2	8.2	82700
			9.89	6.61	61208	18.3	11.9	72526	12.5	7.4	81688
			8.72	5.26	45268*	17.9	11.1	73279	13.2	7.9	86292
			10.47	6.62	65304	17.3	10.6	71933	13.0	7.6	80980
			7.5	7.65					7.8	6.7	344992*
									8.6*	7.0	38936
									12.7	6.9	86052

* column regeneration

The resolution is calculated from the following formula :

$$R = \frac{1.18(t_2 - t_1)}{w_{1h} + w_{2h}} \qquad (2.8\text{-}2)$$

t_1 and t_2 = retention times along the baseline from the point of injection to the perpendicular dropped from the maximum of two adjacent peaks.

w_{1h} and w_{2h} = peak widths at half-height of two adjacent peaks.

The resolution is calculated by using two closely eluting peaks (critical pair) usually of similar height, preferably corresponding to the substance itself and an impurity. When the elution times of the peaks are very different and the resolution is large (> 5.0) the use of the resolution factor as a performance test has little value. It is preferable to use another impurity or another substance, perhaps chemically related to the original substance, which will give a more meaningful resolution. For example, in the monograph of doxorubicin hydrochloride in the test for related substances doxorubicin and epirubicin are employed to determine the resolution since there is no impurity eluting close to the substance itself (Fig. 2.8-1). Ideally the resolution test should be chosen after a test for robustness has been performed (see Section 2.7). When studying the effect on the selectivity of separation between closely eluting impurities, using the variation of four variables by applying a full-fraction factorial design at two levels, not only was the robustness of the method demonstrated, but the choice of the resolution criterion can be made, i.e., provided that there is a minimum resolution, in which case there is adequate separation of all the impurities from each other and from the substance itself. Such studies have been reported for the robustness of the liquid chromatographic methods in order to control the impurities in amoxicillin [183] and in ampicillin [184] and the choice of resolution between cefadroxil and amoxicillin and cefradine and ampicillin, respectively. It is important to ensure that the test for selectivity guarantees that the separation is adequate to control all the potential impurities.

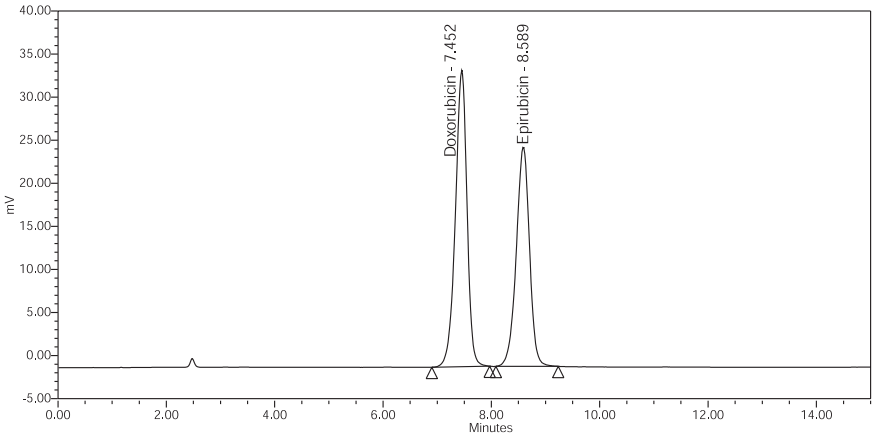


Figure 2.8-1: System Suitability Test – resolution between doxorubicin and epirubicin.

When it is not possible to identify a critical pair of substances for a resolution or if the impurity is unavailable or not available in sufficient quantities, it may be necessary to devise a selectivity test by degrading the substance in solution. Thus 'in situ' degradation [185] offers an alternative approach to defining the selectivity of the chromatographic system provided that the substance can be degraded in mild stress conditions within a reasonably short time in order to produce decomposition products which can be used to determine a resolution or peak-to-valley ratio. An example is the decomposition of rifabutin in mildly alkaline conditions resulting in partial decomposition. The chromatogram is shown in Figure 2.8-2. The requirement is that the resolution is at least 2.0 between the second peak of the three peaks, due to degradation products, and the peak due to rifabutin (retention time about 2–5 minutes).

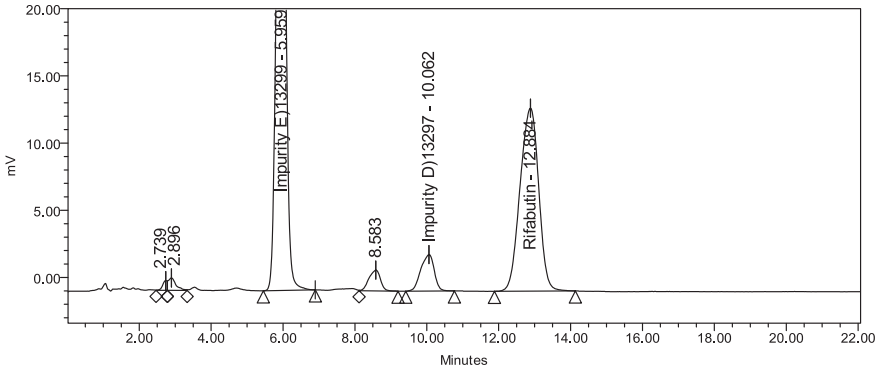


Figure 2.8-2: Chromatogram of rifabutin after alkaline hydrolysis. To 10 mg dissolved in 10 ml methanol, was added 1 ml dilute sodium hydroxide solution which was allowed to stand for 4 min. 1ml of dilute hydrochloric acid was added and diluted to 50 ml with the mobile phase.

When the chromatography is such that there is incomplete separation of the impurity and the availability of the impurity is restricted, the peak-to-valley ratio (p/v) can be employed to define the selectivity

$$p/v = {}^{H_p}/_{H_v} \tag{2.8-3}$$

H_p = height of the peak of the impurity from the extrapolated baseline.
H_v = height above the extrapolated baseline at the lowest point of the curve separating the peaks of the impurity and the analyte.

An example is shown in Figure 2.8-3 for clazuril where there is a peak-to-valley requirement – a minimum of 1.5 for impurity G and the principal peak, but also the chromatogram obtained with the CRS is concordant with the chromatogram supplied with the CRS. Another example is shown for loperamide hydrochloride for system suitability CRS, where there are two peak-to-valley criteria to be fulfilled – a minimum of 1.5 between impurities G and H and impurities A and E (Fig. 2.8-4).

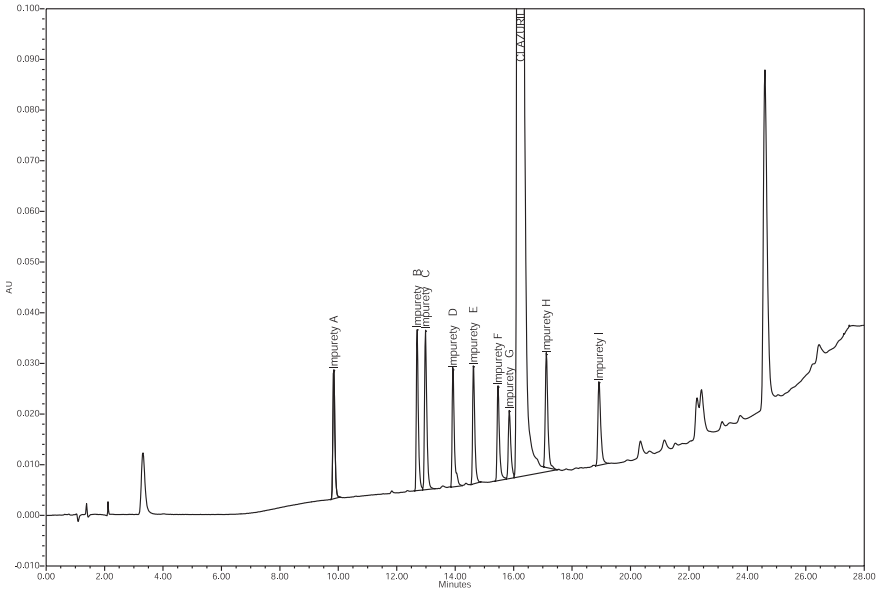


Figure 2.8-3: Chromatogram of clazuril for system suitability CRS.

Another possible approach which can be used, when it is difficult to isolate or obtain an impurity (in sufficient quantity) eluting close to the main peak, is to prepare a reference standard of a mixture of the impurity (ies) with or without the substance itself. In this case a chromatogram is supplied with the reference standard so that a selectivity criterion is included and also the peaks of the impurities may be identified. An example of such an approach is shown in Figure 2.8-5 closantel sodium dihydrate where the baseline separation between impurity G and the main peak are shown, but the impurities can also be identified by their relative retentions.

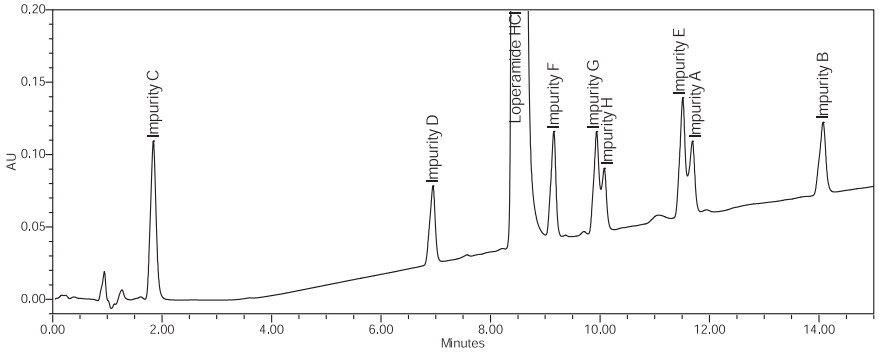


Figure 2.8-4: Chromatogram of loperamide hydrochloride for system suitability CRS.

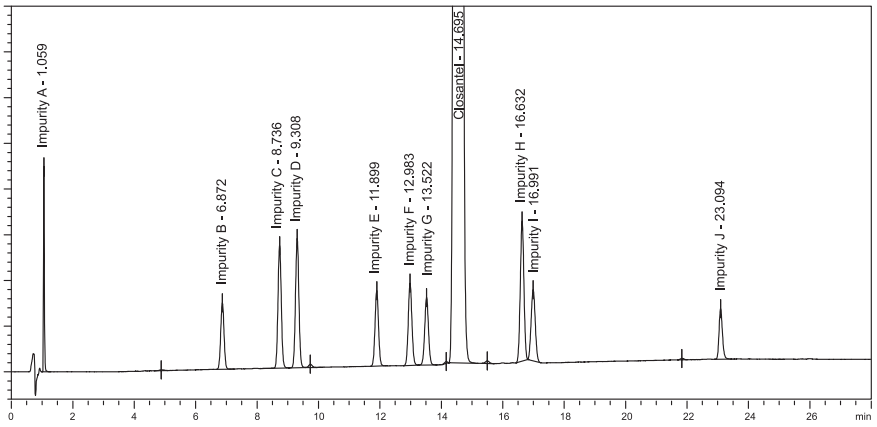


Figure 2.8-5: Chromatogram of closantel sodium dihydrate for system suitability CRS.

2.8.3.2 **Symmetry**

Peak shape is an important contributor to the selectivity of the method. It is therefore necessary to include a system suitability criterion for symmetry, which is generally applicable (Eq. 2.8-4) [15] to the appropriate reference solution either in the assay method or in the procedure for the test for related substances. In the latter case, the symmetry factor does not apply to the principal peak of the test solution since it will either be asymmetric due to overloading or it cannot be calculated because of detector saturation.

Unless otherwise stated in the monograph, the symmetry factor should fall between 0.8 and 1.6 (a value of 1.0 signifies ideal symmetry).

$$As = {}^{W_{0.5}}/_{2d} \tag{2.8-4}$$

$W_{0.5}$ = width of the peak at 1/20 of the peak height.
d = distance between the perpendicular dropped from the peak maximum to the leading edge of the peak at 1/20 of the peak height.

2.8.3.3 **Retention Time/Relative Retention**

Although not generally given as suitability requirements in monographs, for guidance the approximate retention time of the substance as well as the relative retention of the impurities, should be indicated. Nonetheless, it has been shown that, even when the selectivity requirements are fulfilled, there can be a wide variation in the retention time of the main component depending on the stationary phase employed (Table 2.8-2). In a collaborative study [53] to assess a liquid chromatographic method for the assay, and a related substances test for dicloxacillin sodium the retention times of the peak due to dicloxacillin were in the range 7–39 minutes! This is dramatic and will have an effect on the selectivity of the method. Such disparity in retention times between different columns when using an isocratic elution indicates

Table 2.8-2 Data obtained from the collaborative study to evaluate a LC method for dicloxacillin sodium. (Reproduced from reference [53].)

Lab	Column (commercial source)	Dimensions (mm)	Symmetry	Resolution	Retention time (min)	Repeatability (RSD) of retention time
1	Hypersil-ODS (5 µm)	4.6 × 250	1.3	5.1	17.16	0.55
2	Kromasil C-18 (5 µm)	4.6 × 250	1.4	10.4	18.03	0.64
3	Kromasil 100A C-18 (5 µm)	4.6 × 250	1.6	9.0	24.95	0.14
4	Nucleosil C18 (5 µm)	4.6 × 250	1.2	8.0	16.81	1.03
5	Lichrospher 100 RP18 (5 µm)	4.6 × 250	1.2	9.5	24.69	1.15
6	Hichrom C-18 (5 µm)	4.6 × 250	1.0	6.7	7.78	0.59
7	Lichrospher 100 RP18 (5 µm)	4.6 × 250	1.0	10.2	28.55	0.13
8	Altima C18 (5 µm)	4.6 × 250	1.5	10.4	39.26	0.45
9	Hypersil-ODS (5 µm)	4.6 × 250	1.9	6.2	12.76	0.31

that a resolution criterion is insufficient on its own and should be supplemented by an indicated retention time which should be within pre-defined limits (e.g. ± 10 percent).

Although not considered as system suitability criteria, the expected retention time of the main component and the relative retentions of other compounds (e.g. impurities) should be given for information.

2.8.3.4 Adjustment of Chromatographic Conditions

Differences in the efficiency of stationary phases (particularly of the reverse phase type) which vary between batches either from the same manufacturer or from one manufacturer to another, can lead to the concept of permitted adjustments to the chromatographic conditions [186, 187]. The extent to which the various parameters of a chromatographic test may be adjusted to satisfy the system suitability criteria for selectivity, without fundamentally altering the method to such an extent that revalidation is required, was published in the 4th Edition of the European Pharmacopoeia [15]. Permitted maximum modifications to various parameters were given for thin-layer and paper chromatography, liquid chromatography, gas chromatography and supercritical fluid chromatography. However, it should be noted that the modifications cited for liquid chromatography were only to be applied to isocratic methods. If a column is used whose dimensions are different from those described in the method, then the flow rate will need to be adjusted to achieve a similar retention time. The retention time, column temperature and the flow rate are related in the following way:

$$Q = (Rt \times f)/(l \times d^2) \tag{2.8-5}$$

where Q is a constant and Rt is the retention time, f is the flow rate, l is the length of the column and d is the internal diameter of the column.

For example, changing from a typical US manufactured column (300 x 0.39 cm) to a European manufactured column (250 × 0.46 cm) it is necessary to change the flow rate to obtain the same retention time. If in the original method the retention time of the analyte is 10 minutes with a flow rate of 1.0 ml/min then using the expression above, the flow rate must be increased to 1.2 ml/min. Alternatively, a minor adjustment of the mobile phase composition is permitted for isocratic elution. The amount of minor solvent component may be adjusted by ±30 percent relative or ±2 percent absolute, whichever is the larger. No other component is to be altered by more than ±10 percent absolute. For any such change the system suitability criterion for selectivity must be fulfilled. However, with reverse-phase liquid chromatographic methods, adjustment of the various parameters will not necessarily result in satisfactory separation. It will be necessary therefore to change the column with another of the same type, which exhibits similar chromatographic behaviour.

2.8.3.5 **Column Selection**

With the wide variety of reverse-phase stationary phases commercially available, it would be very useful if they could be classified so as to select a suitable column for a particular separation. This is not always evident when applying liquid chromatographic methods in either the United States Pharmacopoeia or the European Pharmacopoeia, since the description of the stationary phase is often not sufficiently precise and neither compendia publish, in the individual monographs, the commercial name of the column employed. The European Pharmacopoeia describes [188] seven types of octadecylsilyl silica phases in general terms (Table 2.8-3). Any column falling into a category would be expected to give a satisfactory separation in the test described in the individual monographs. Nonetheless, some separations can only be achieved using one particular stationary phase, in which case a detailed description of the physical properties of the stationary phase is included in the description of the test.

Table 2.8-3 Reverse-phase (C18) stationary phases listed in the European Pharmacopoeia (octadecylsilyl silica gel for chromatography).

BvReference	Description	
1077500	R	3–10 µm
1110100	R1	Ultra pure < 20 ppm metals
1115300	R2	Ultra pure, 15 nm pore size, 20 percent carbon load
1077600	Base-deactivated	3–10 µm, pre-washed and hydrolysed to remove superficial siloxane groups
115400	End-capped	3–10 µm, chemically modified to react with the remaining silanol groups
1108600	End-capped and base deactivated	3–10 µm, 10 nm pore size, 16 percent carbon load, pre-washed and hydrolysed to remove superficial silox groups and chemically modified to react with the remaining silanol groups

The European Pharmacopoeia does, however, provide the commercial names of the columns found to be satisfactory on its website [189]

It has been proposed that stationary phases should be classified based on chemical rather than physical properties. Chemical properties can be measured chromatographically, such as column efficiency, hydrophobicity, steric selectivity, silanol activity, ion-exchange capacity, steric selectivity, level of metal impurities and polar interactions [190].

Evaluation of the retention behaviour of a large number of compounds using such techniques as principal component analysis (PCA), cluster analysis (CA) and radar plots which have been applied [191–194] to the results obtained from a limited number of chromatographic tests designed to assess different properties of the stationary phases [193]. Using these approaches 30 commercial columns were evalu-

Table 2.8-4 Chromatographic testing procedures to categorise liquid reverse-phase commercial columns (http//www.farm.kuleuven.ac.be/pharmchem).

Method	Test solution			Mobile phase
	Substance	Parameter (to be measured)	Composition (% m/m)	
M4 (silanol activity)	Phenol	rK'ba/ph	0.05	MeOH:H_2O:0.2MKH_2PO_4 buffer @ pH 2.7 (34:90:10, m/m/m)
	Benzylamine			
M6 (silanol activity and metal impurities)	Uracil		0.001	MeOH:H_2O (34:100, m/m)
	Caffeine		0.003	
	Theobromine		0.002	
	Theophylline		0.01	
	Phenol		0.035	
	Pyridine		0.01	
	2,2'dipyridyl	$K_{2,2}$'d	0.03	
	2,3 dihydroxynapthalene		0.03	
M8 (hydrophoticity and steric selectivity)	Uracil		0.001	MeOH:H_2O (317 : 100, m/m)
			0.006	
	Phenol		0.025	
	Toluene		0.025	
	Ethylbenzene		0.025	
	Butylbenzene		0.070	
	Amylbenzene	(K'amb)	0.002	
	o-terphenyl	$rK'_{tri/ter}$	0.0002	
	Triphenylene			

rK'ba/ph relative retention factor of benzylamine/phenol
K'amb retention factor of amylbenzene
$rK'_{tri/ter}$ relative retention factor of triphenyl/terphenyl
$K_{2,2}$'d retention factor of 2,2'dipyridyl

ated showing that large differences such as column length and type of silica can be identified, so that such analyses may be useful to identify columns with similar qualities and also to categorise columns. Reviews of the methods employed have been recently published [196–198]. Although cluster analysis and radar plots can be useful in visually distinguishing different stationary phases, the use of principal component analysis has been favoured as a means of categorising commercial reverse-phase columns. A column characterisation database containing 135 different stationary phases, including C_{18}, C_8, cyano, phenyl, etc., has been established based on the results of the chromatographic tests subjected to PCA [199]. Thus, different silica-based reverse-phase columns can be grouped into non-C_{18} phases, acidic phases, new generation phases, polar embedded phases, cyano phases and perflurophenyl phases. Another group of workers have also taken a similar approach and have evaluated the chromatographic tests by determining the repeatability, reproducibility and correlation of 36 test parameters [200]. Subsequently a reduced testing regime has been proposed [201, 202] which has been used to classify the stationary phases.

The separation of acetylsalicylic acid and its impurities was performed according to the related substances test described in the European Pharmacopoeia using the stationary phases, which had been previously characterised chromatographically [203]. The chromatographic tests applied are given in Table 2.8-4 and the classification of the stationary phases is shown in Table 2.8-5. Three major groups were observed, including a principal group containing the majority of columns tested, a group with high silanol activity and a group specifically conceived to analyse polar compounds. Group 1 was further sub-divided into Groups 1a and 1b, essentially based on their hydrophobicity. Of the 31 columns in Group 1a, nine failed the resolution test and nine failed the requirement for the symmetry factor (0.8–1.5). Four stationary phases failed to conform to either requirement. However, the resolution criteria was not met by the columns whose lengths were shorter than that prescribed in the monograph (0.25 m). Nine of the 18 columns of the correct column length complied with the criteria for resolution and symmetry. Some columns, although not complying with the system suitability requirement, still gave baseline separations, whilst other columns meeting the system suitability requirements did not give baseline separations. These columns were not found in Group 1a or were short columns. The chromatographic response function (CRE, Eq. 2.8-6) was used to assess the separation. It equals 1.0 when there is complete separation of the components (Table 2.8-5).

$$\prod_{t=1}^{n-1} f_i/g_i \tag{2.8-6}$$

$\prod$ = total number of solutes (components).
g = interpolated peak height (distance between the baseline and the line connecting two peak tops, at the location of the valley.
f = depth of valley, increased from the line connecting two peak tops.

Major differences in selectivity were observed with some columns, evidenced in the changing order of elution, but none of these stationary phases were in Group 1a.

Table 2.8-5 Results for the resolution (SST), chromatographic response factor (CRF), and the symmetry factor for ASA (SF) on RP-LC stationary phases. (Reproduced from reference [203].)

Group	No	Column	SST	CRF	SF
Ia	*3*	*Alltima 3*	*7.8*	*0.00*	*2.5*
	4	Alltima 5	9.6	1.00	1.5
	7	*Aqua 5*	*7.1*	*1.00*	*1.2*
	13	*Genesis C18-3*	*4.9*	*0.74*	*1.3*
	22	Kromasil NM	10.0	1.00	1.3
	23	Krmasil EKA	7.0	1.00	4.2
	26	*Luna 5*	*6.9*	*1.00*	*1.5*
	29	Necleosil HD	3.2	1.00	3.2
	31	OmniSpher	6.0	1.00	3.9
	37	*Prodigy 3*	*5.4*	*0.81*	*1.4*
	39	Purospher endcapped	6.9	1.00	1.1
	40	Purospher Star	10.6	1.00	1.2
	47	Superspher	5.1	1.00	4.4
	49	Symmetry	8.8	1.00	2.3
	50	*Tracerexcel 3*	*8.0*	*1.00*	*1.3*
	51	Tracerexcel 5	10.4	1.00	1.2
	52	*TSKgel ODS-80TS*	*6.1*	*0.84*	*1.4*
	54	*Uptispher HDO3*	*5.8*	*1.00*	*1.2*
	55	Uptispher HDO5	9.8	1.00	1.1
	56	*Uptispher ODB3*	*5.9*	*0.94*	*1.2*
	57	Uptispher ODB5	8.3	1.00	1.1
	58	Validated C18	4.5	1.00	3.4
	59	*Wakosil HG 5-10*	*4.2*	*0.66*	*1.6*
	60	*Wakosil HG 5-25*	*7.8*	*1.00*	*1.7*
	61	*Wakosil RS 3-10*	*5.3*	*0.89*	*1.5*
	64	YMC Hydospher C18	7.3	1.00	1.1
	65	YMC-Pack-Pro C18-3	7.6	1.00	1.4
	66	YMC-Pack-Pro C18-5	7.7	1.00	1.3
	67	Zorbax Eclipse XDB	8.6	1.00	1.4
	68	Zorbax Extend C18	2.4	1.00	2.6
	69	Zorbax SB-C18	6.6	1.00	3.1
Ib	*1*	*ACE C18-3*	*7.1*	*0.93*	*1.4*
	2	ACE C18-5	9.0	1.00	1.2
	9	*Brava BDS 3*	*5.5*	*0.63*	*1.4*
	10	Brava BDS 5	5.8	0.67	1.4
	12	Discovery	6.6	0.96	1.3
	16	Hypersil BDS	6.6	0.91	1.3
	19	*HyPuritiy Elite 3*	*5.1*	*0.80*	*1.4*
	20	*HyPurity Elite 5*	*4.4*	*0.66*	*1.4*
	33	*Platinum 3*	*4.6*	*0.40*	*1.5*
	34	Platinum 5	3.0	0.24	3.0
	38	Purospher	a	0.00	b
	45	*Supelcosil LC-18 DB 3*	*a*	*0.00*	*b*

Table 2.8-5 Continued.

Group	No	Column	SST	CRF	SF
	46	Supelcosil LC-18 DB 5	a	0.00	b
IIa	*32*	*Pecosphere*	*3.0*	*0.50*	*2.8*
	42	Spheri	3.4	0.96	1.4
IIb	25	LiChrospher	5.2	1.00	3.8
	28	Nucleosil NM	6.4	1.00	1.4
	43	Spherisorb ODS2	3.8	1.00	3.5
IIc	6	Apex ODS	5.1	0.67	2.8
	18	Hypersil ODS	5.2	0.88	3.8
	44	Supelcosil LC 18	4.2	0.00	4.9
III	30	Nucleosil C18 Nautilus	a	0.00	1.0
	35	*Platinum EPS 3*	*2.8*	*0.09*	*1.5*
	36	Platinum EPS 5	3.5	0.42	1.3
Outlier	5	Apex Basic	a	0.00	b

a) Changed selectivity
b) Peak co-eluted or not observed
Italics are used for columns shorter than 250 mm

Overall, it seems that this approach can be potentially useful to identify stationary phases with similar chromatographic behaviour, which would facilitate the choice of an alternative column for chromatography. It seems from the study of acetylsalicylic acid and its impurities that a resolution criteria alone is not sufficient to ensure suitability of the column employed, but that the situation could be improved by a better description of the columns to be used.

The Impurities Working Party of the USP is examining another approach [204–207] where it has been established that five column properties, hydrophobicity (H), steric hindrance (S), hydrogen bonding activity (A), hydrogen bond basicity (B) and cation ion-behaviour (C) are necessary to characterise the stationary phase. It is maintained that the measurement of all five properties is necessary to show similarity between columns.

$$\log K = \log K_{EB} \pm n'H \pm \sigma'S \pm \beta'A \pm \alpha'B \pm k'C \tag{2.8-7}$$

A column-matching factor *Fs* has been defined which is a function of the differences in values *H, S, A, B* and *C* for two columns.

$$Fs = \left[(H_2 - H_1)^2 + (S_2 - S_1)^2 + (A_2 - A_1)^2 + (B_2 - B_1)^2 + (C_2 - C_1)^2\right]^{\frac{1}{2}} \tag{2.8-8}$$

However, weighting factors are applied to each term.

If $Fs < 3$ for two columns, then those columns should have similar selectivity, i.e., the separations are equivalent. The procedure has been developed for columns based on Type B silica only. The advantage of this method is that the same chromatographic conditions are employed for two test solutions each containing four test components. Again a database has been established.

It seems that, in general, insufficient attention is being given to the description of the stationary phases to be employed in the tests for related substances, and the present descriptions in the reagents section are inadequate. As previously advocated [54] and indicated in the Technical Guide for the Elaboration of Monographs [208] a more detailed description of the stationary phase is often necessary.

The classification of stationary phases by chemical rather than physical attributes shows promise but is not totally reliable as seen from the study of acetylsalicylic acid. As yet this approach has not been tested extensively either for other acids or for bases. Neither has it been tested for gradient chromatographic separations.

The column classification as a PCA score plot or a column-matching factor based on the chromatographic parameters, may facilitate the choice of stationary phases with similar selectivity but improved descriptions of the physical characteristics of the stationary phases would lead to improvement in selection of columns

2.8.3.6 **Gradient Elution and Dwell Volume**

For gradient elution in liquid chromatography, the configuration of the equipment may significantly alter the resolution, retention time and relative retention described in the method. The two principal factors affecting the elution in gradient methods are the stationary phase employed and the type of gradient pump used.

In the pharmacopoeia, retention times (volumes) or relative retentions of analytes are given in the individual monographs, as an aid to interpreting the chromatograms, but which do not take into account the configuration of the chromatographic equipment. It is assumed that the *extra-column volume,* V_{ext}, [209] which includes:

- injection volume
- volume of tubing from the injection device to the head of the column
- volume of tubing from the bottom of the column to the detector
- the detector cell volume,

is minimal, if not negligible, and will have no significant effect on the chromatographic process. The extra-column volume, determined during the qualification of the chromatographic equipment (see Chapter 4), should be reduced to a minimum otherwise there will be significant deterioration of chromatographic performance as witnessed by longer retention times and altered relative retentions, whilst system suitability criteria such as resolution and symmetry, may be adversely affected. Thus the experimentally measured retention volume (time) is *gross retention volume*, $V_R{}^G$, [209].

$$V_R^G = V_R^I + V_m + V_{ext} \text{ or } V_R + V_{ext} \tag{2.8-9}$$

V_m = *hold-up volume,* which is the volume of the mobile phase passing through the column with an unretained analyte. In this context, the use of the terms dead volume and void volume should be avoided, since these terms have other definitions.

$V_R{}^i$ = *adjusted retention volume,* which consists of two parts including the volume of the mobile phase used to transport the analyte along the column, and the volume of the mobile phase that left the column while the analyte was retained by the stationary phase.

Provided that the hold-up volume is the same and the extra-column volume is reduced to the minimum and is considered to be negligible, then the recorded retention volumes using different chromatographic systems, but with the same mobile and stationary phases, should be similar.

A number of procedures [210–213] have been described for the determination of the hold-up volume but the simplest to perform include the following:

- The volume required to detect an unretained substance, which may be organic or inorganic, is measured. Commonly substances used for this purpose include sodium nitrate or nitrite and uracil.
- The volume required to detect a solvent (system) peak is measured.
- The column is weighed after passing two different solvents.

$$V_m = \frac{m_1 - m_2}{p_1 - p_2} \tag{2.8-10}$$

m_1 and m_2 = weights of the column containing each of the two different solvents

p_1 and p_2 = their respective densities.

The extra-column volume can be determined by injecting a solution of a substance into the injection device and then measuring the time (volume) for its detection without the column in the chromatographic system. The effect of a large extra-column volume manifests itself with adverse effects on retention time, plate number, symmetry and resolution. An example is given in Table 2.8-6 where the performance of a chromatographic system was compromised by the installation of tubing, whose internal diameter was too great, from the bottom of the column to

Table 2.8-6 The effects on chromatographic parameters by variation in the internal diameter of capillary tubing used in the chromatographic system. Capillary tubing from bottom of column to detector (A) before replacement, 0.02 ins (B) after replacement, 0.009 ins.

	A				B			
	Peaks				Peaks			
	1	2	3	4	1	2	3	4
Retention times								
mean	1.46	2.38	3.42	5.22	1.37	2.25	3.23	4.91
RSD	0.17	0.53	0.81	1.02	0.29	0.27	0.26	0.26
Peak area RSD	1.36	0.36	0.20	0.11	0.61	0.20	0.14	0.05
Symmetry	1.95	1.93	1.77	1.60	1.52	1.48	1.43	1.39
Plate number	496	1215	2105	3605	1682	3400	5119	7107
Resolution	3.50	3.70	5.60		6.13	5.87	8.16	

the detector. Replacement of this tubing with tubing of smaller internal diameter led to a remarkable improvement in performance.

When consideration is given to gradient elution in liquid chromatography, yet another parameter is to be considered – the volume between the solvent-mixing chamber and the head of the column. This volume is sometimes referred to as the *dwell volume*, V_D (other terms employed include: effective system delay volume, dead volume and delay volume). If the mixing of solvents takes place before (low pressure) or after (high pressure) the pump work is dependent on the configurations of the pumping system (Fig. 2.8-6) including the dimensions of the capillary tubing, the solvent mixing chamber and the volume of the injection loop [209].

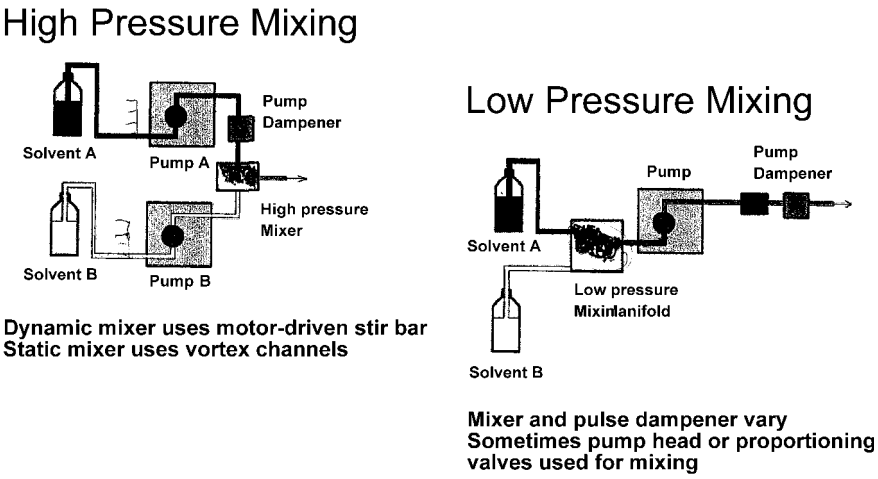


Figure 2.8-6: Configuration of high-pressure and low-pressure pumps used in liquid chromatography.

In gradient elution the factor Q should remain constant to avoid changes in selectivity [148].

$$Q = tg/ld^2 \tag{2.8-11}$$

t_g = gradient time (chromatographic run time)
f = flow rate
l = column length
d = internal diameter of the column.

This holds true only if the dwell volume is negligible with respect to the retention volume. It has been proposed [214] that the dwell volume of the chromatographic system used during the development of a gradient method for a monograph should be indicated, so that an appropriate correction can be made to published retention times or relative retentions, when the dwell volume differs.

However, the correction may be justified only if there are large differences in the dwell volumes of equipment as there will be an effect on the measured resolution with significant changes in the retention times, resulting in the failure of the system suitability test for selectivity. The calculated dwell volumes for a number of different gradient chromatographic systems are given in Table 2.8-7. Thus, provided that the ratio

$$V_D/V_m$$

is constant, then the relative retention

$$t_R/t_O$$

will also be constant (where t_0 is the retention time of an unretained peak). It can be seen from this table that the specification of various analytical pumps from different manufacturers, whether eluent mixing is from the low-pressure or high-pressure side of the pumping system, are similar. The measured dwell volumes for some of these pumping systems are also given and have been determined as previously described [214] (Fig. 2.8-7). As can be seen from Table 2.8-7 the measured dwell volumes, using the same method for each system, vary and do not correspond to the manufacturers specification in all cases. Thus, a number of gradients were applied to two chromatographic systems whose dwell volumes were different, in order to separate a test mixture of phenol, naphthalene and anthracene. In some systems, linear gradients were applied and, in another, one of the linear gradients was preceded by an isocratic phase. In the latter example an isocratic phase of five minutes was applied to the first system and a correction was applied to the second system based on the difference in dwell volume (3.4 – 0.6 = 2.8 ml) so that the isocratic phase was programmed for 2.8 ml.

Table 2.8-7 Dwell volumes of various pumping systems from different manufacturers.

	Manufacturer	Pump	Type	Dwell volume (ml)	
				(Specification)	(Experimental)
1	Waters	Alliance	L	< 0.65	1.3
2	Waters		L	< 0.65	3.4
3	Waters		L		3.9
4	TSP	P400	L	< 0.8	0.7
5	TSP	P1000XR	L	< 0.6	0.6
6	TSP	P1000XR	L	< 0.6	< 0.5
7	Jasco	PU980	L	na	1.3
8	Jasco	PU980	L	na	1.4
9	Beckman		H	< 0.6	na

L = Low-pressure mixing
H = High-pressure mixing
na = Not available

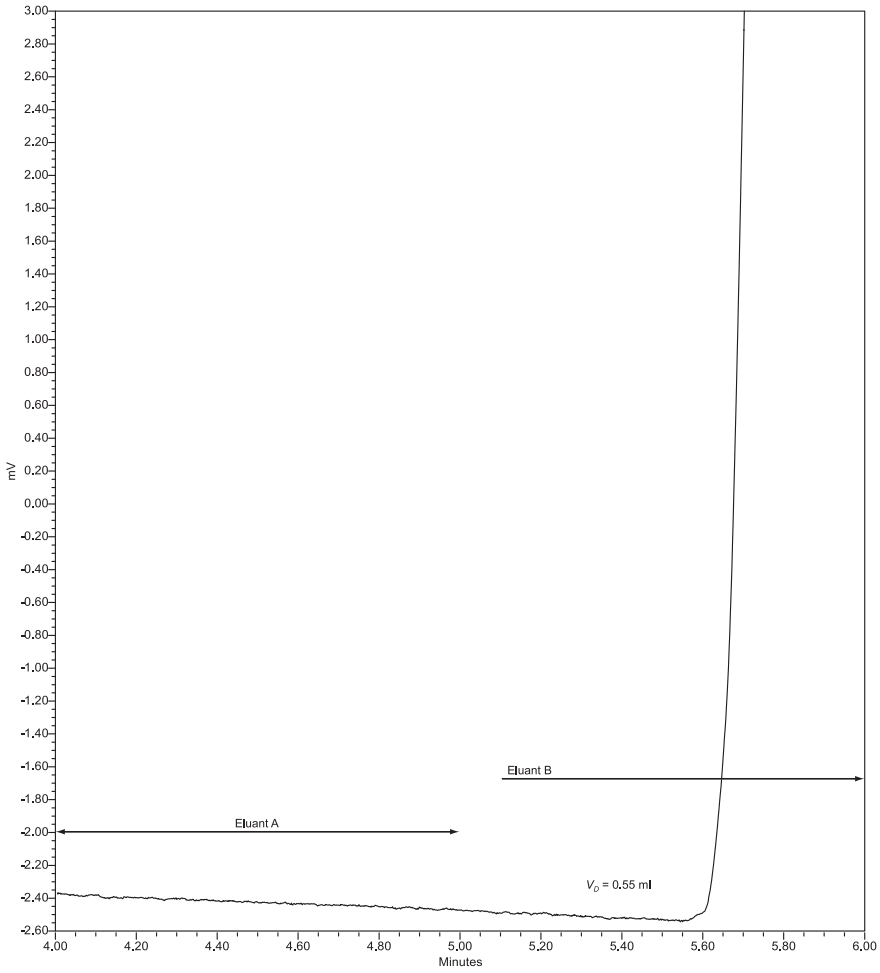


Figure 2.8-7: Determination of the dwell volume. Mobile phase A (water) for 5 minutes then mobile phase B (0.5 % v/v acetone in water) between 5 and 10 minutes. Flow rate 1.0 ml/min, wavelength of detection 254 nm.

A comparison of the two chromatographic systems has been made with relation to retention time, relative retention and resolution and the results are presented in Table 2.8-8 from which it can be seen that:

- The steeper the gradient the greater is the difference in retention time, relative retention and resolution when comparing chromatographic gradient systems with markedly different dwell volumes.
- The greater the dwell volumes, the greater the resolution.
- The more gradual the gradient, the less different are the retention times, relative retentions and resolution.

– The introduction of an initial isocratic step and correction for the differences in dwell volume between the two systems results in similar values for the three parameters determined.

Table 2.8-8 Comparison of different chromatographic gradients on selectivity for chromatographs with different dwell volumes.

Gradient A: Acetonitrile (70 : 30, v/v) to (90 : 10, v/v) in 5 mins (linear gradient).
Gradient B: Acetonitrile in water (70 : 30, v/v) to (90 : 10, v/v) in 20 mins (linear gradient).
Gradient C: Acetonitrile in water (60 : 40, v/v) to (80 : 20, v/v) in 20 mins (linear gradient).
Gradient D: Acetonitrile in water (70 : 30, v/v) in 5 mins (isocratic), then acetonitrile in water (70 : 30, v/v) to (90 : 10, v/v) in 10 mins (linear gradient), then acetonitrile in water (90 : 10, v/v) in 10 mins (isocratic).
Column: 4.6 × 250 mm Hypersil ODS (5 µm), flow rate: 1.0 ml/min, detection wavelength: 240 nm.

Gradient	Chromatographic	Phenol			Naphthalene			Anthracene	
	system	Rt	Rr	Rs	Rt	Rr	Rs	Rt	Rr
A	1	3.13	1.00	13.24	5.88	1.88	11.86	8.07	2.58
	2	3.34	1.00	14.12	6.72	2.01	13.95	9.97	2.98
	Δ%	+6.3	–	+6.3	+12.5	+6.5	+15.0	+19.0	+13.4
B	1	3.13	1.00	13.47	6.21	1.99	14.12	9.66	3.09
C	1	3.46	1.00	20.14	8.86	2.56	18.77	14.50	4.19
	2	3.68	1.00	20.66	9.66	2.63	19.58	16.26	4.42
	Δ%	+6.0	–	+2.5	+8.3	+2.7	+4.1	+10.8	+5.2
D	1	3.13	1.00	13.56	6.39	2.05	16.09	10.43	3.34
	2	3.34	1.00	13.67	6.72	2.01	16.13	10.89	3.26
	Δ%	+6.3	–	+0.8	+4.9	–1.2	+0.25	+4.2	–2.5

Over the years, the European Pharmacopoeia has seen a rapid increase in the number of monographs containing liquid chromatographic methods for the control of related substances. Since many active pharmaceutical substances are synthesised by a number of synthetic routes, the list of potential impurities to be limited has increased considerably and the analytical challenge to separate them is great. Isocratic liquid chromatographic methods may not be sufficiently selective so that there is an increasing propensity to employ gradient liquid chromatographic methods. Thus, it is important to be aware of the potential pitfalls due to significant differences in dwell volume in the context of method transfer.

In an effort to gauge the real effect on the chromatography of systems with different dwell volumes, it was decided to examine the separation of the impurities of substances for which a gradient elution is described and where the system suitability for selectivity is critical – either a resolution requirement of less than 3.0 or a peak-to-valley ration is prescribed. To this end, a comparison of two systems has been made with the test for related substances in the monograph of trimethazidine hydrochloride as an example of where a minimum peak-to-valley ration of 3 is required. The superimposed chromatograms are shown in Figure 2.8-8.

It seems, therefore, that the contribution of the dwell volume of an analytical gradient chromatographic system to variation in retention times when changing from one system to another is very small, and a more important factor is probably the performance of the stationary phase employed.

It seems simpler to prescribe, for pharmacopoeial purposes, an upper limit for the dwell volume if gradient chromatographic systems are employed. Thus there would be no relevant effect on the aforementioned chromatographic parameters if the dwell volume were less than the given limit.

It was shown [212] that the greatest effect of differing dwell volumes on retention times was for those substances which were not strongly retained. Thus, gradient systems should be conceived in such a way that analytes should be strongly retained at the beginning of the gradient. It is best if less strongly retained components are eluted with an initial isocratic phase followed by a gradient for elution of the more strongly retained analytes. By such a strategy, the effect of differences in dwell volumes is minimised and can also be corrected for by adjusting the isocratic part, if required.

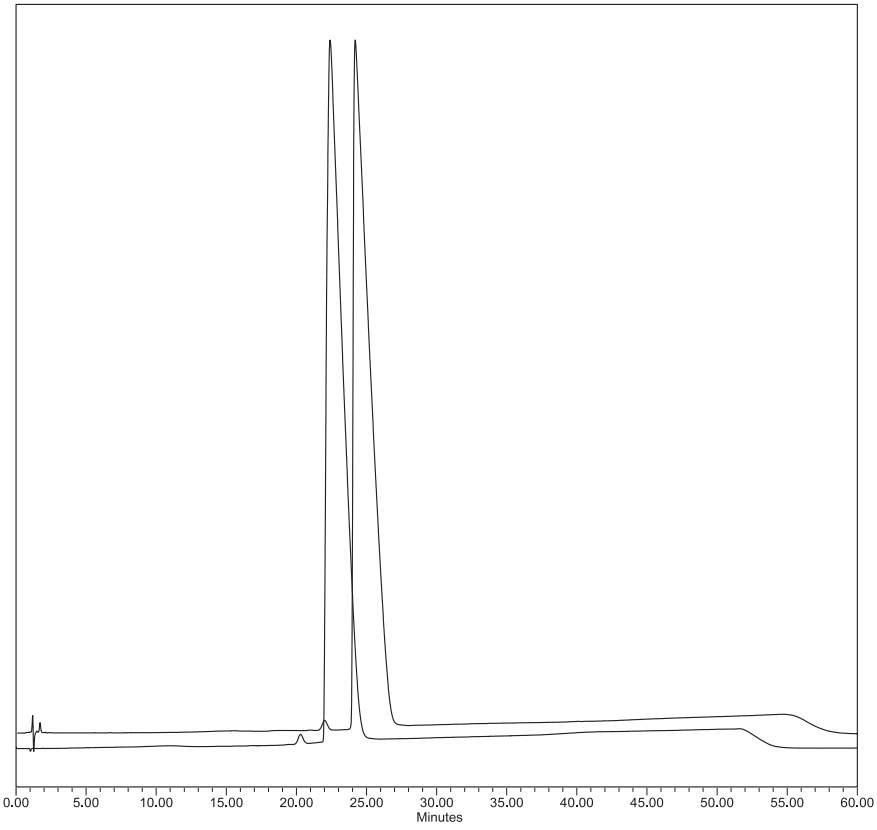


Figure 2.8-8: Chromatograms of 0.1 % m/v trimethzidine for system suitability CRS with pumps of 0.6 ml and 3.4 ml dwell volumes.

2.8.3.7 Sensitivity

According to ICH [1c] there is a reporting threshold below which peaks are not to be quantified. The reporting threshold is equivalent to the disregard limit, which is employed by the European Pharmacopoeia in tests for related substances. The limit of detection of the peak (corresponding to a signal-to-noise ratio of three) is below the disregard limit (reporting threshold) in the test for related substances. It follows that the limit of quantitation of the peak (corresponding to a signal-to-noise ratio of ten) is equal or less than the disregard limit in the test for related substances. The signal-to-noise ratio (S/N) influences the precision of the quantification and is calculated as follows:

$$S/N = 2H/h \tag{2.8-12}$$

H = height of the peak of the analyte in the chromatogram obtained with the prescribed reference solution, measured from the maximum peak to the extrapolated baseline of the signal, obtained over a distance equal to 20 times the width of the peak at half-height.

h = range of the background noise in a chromatogram obtained after injection or application of a blank, observed over a distance equal to 20 times the width of the peak in the chromatogram obtained with the prescribed reference solution and, if possible, situated equally around the place where this peak would be found.

The reporting threshold depends on the maximum daily dose. For drug substances where the maximum daily dose exceeds 2 g the reporting threshold is 0.03 percent, otherwise the reporting threshold is 0.05 percent (see Table 2.8-9)

Table 2.8-9 Control of impurities in substances for pharmaceutical use.
Reporting, identification and qualification of organic impurities in active substances.

Use	Max. daily dose	Reporting threshold	Identification threshold	Qualification threshold
Human or human and veterinary use	≤ 2 g/day	> 0.05%	>0.10% or a daily intake of > 1.0 mg (whichever is the lower)	> 0.15% or a daily intake of > 1.0 mg (whichever is the lower)
Human or human and veterinary use	> 2 g/day	> 0.03%	> 0.05%	> 0.05%
Veterinary use	Not applicable	> 0.1%	> 0.2%	> 0.5%

2.8.3.8 Precision

System suitability criteria for system precision are included in the description of the assay procedures prescribed in specifications or in individual monographs. Often the test requires six replicate injections of the same solution and the RSD should

not exceed 1.0 or 2.0 percent. It was demonstrated [215], however, that there was no apparent link between the precision requirement and the limits set for the assay. The repeatability requirements were often incompatible with the assay limits given. As a result, the European Pharmacopoeia [15] introduced system precision criteria which were dependent on the number of replicate injections performed and the reproducibility of the method.

It had been proposed [14] that maximum permitted relative standard deviations can be calculated for a system suitability requirement for precision, taking into account repeatability reported by laboratories participating in inter-laboratory studies in a similar manner as had been described for setting assay limits [18]. It was shown that [41] a maximum RSD of 0.6 after six replicate injections was required to set a limit of ± 1.0 percent for direct methods such as volumetric titration. For comparative assay methods this value is to be divided by $\sqrt{2}$. To assure the same level of precision if less replicates (n) are performed it is necessary to adjust the calculation:

$$RSD_{\max} = \frac{0.6}{\sqrt{2}} \times \frac{t_{90\%,5}}{\sqrt{6}} \times \frac{\sqrt{n}}{t_{90\%,n-1}} \times B = \frac{0.349B\sqrt{n}}{t_{90\%,n-1}} \tag{2.8-13}$$

B = the upper limit of the assay (provided it represents the reproducibility of the method minus 100 percent.

n = number of replicate injections

$t_{90\%,n-1}$ = student- t value at the 90 percent probability level.

The relationship between the number of replicate injections, the maximum permitted relative standard deviation and the upper content limit is given in Table 2.8-10.

Table 2.8-10 Relationship between the number of injections, the maximum permitted relative standard deviation and the content limit (reproduced from reference 41).

*A**	Maximum permitted related standard deviation (RSD_{max})				
(%)	$n = 2$	$n = 3$	$n = 4$	$n = 5$	$n = 6$
1.0	**0.08**	0.21	0.30	0.37	0.42
1.5	**0.12**	0.31	0.44	0.55	0.64
2.0	**0.16**	0.41	0.59	0.73	0.85
2.5	**0.20**	0.52	0.74	0.92	1.06
3.0	**0.23**	0.62	0.89	1.10	1.27

* *A*: Upper specification limit (%)– 100

Any decision to accept or reject analytical results should include an assessment of the system suitability criteria to ensure that adequate precision is achieved. The maximum number of replicate injections to be performed is six but fewer may be performed provided the system precision RSD is equal to or less than RSD_{max} given in the table for the appropriate number of injections. These levels of precision can be easily achieved with modern chromatographic equipment, provided that the concentration of the analyte is sufficiently above the quantitation limit, to reflect the

injection precision (see Section 2.1.3.1). Table 2.8-11 shows examples of precision of replicate injections achieved by participants in a collaborative trial to establish reference substances. The importance of complying with the system suitability criteria for injection repeatability is illustrated in Table 2.8-12 which tabulates the results obtained by a number of laboratories which had participated in the establishment of a reference standard for buserelin. It is clear that laboratory F failed to meet the cri-

Table 2.8-11 System Suitability: Injection repeatability ($n = 6$) of reference solution. (Data extracted from CRS establishment reports.)

Substance	No. of Labs	Concentration (mg/ml)	Mean RSD	Range of RSD
Allopurinol	5*	0.5	0.13	0.04 – 0.24
Ampicillin anhydrous	5	0.03	0.38	0.10 – 0.71
Budnesonide	8*	0.5	0.20	0.08 – 0.43
Cloxacillin sodium	10*	1.0	0.44	0.02 – 0.87
Cloxacillin sodium	10*	0.01	0.56	0.34 – 0.89
Crotamiton	5	0.5	0.26	0.08 – 0.58
Doxorubicin HCl	4	0.05	0.46	0.35 – 0.58
Doxorubicin HCl	4	0.5	0.46	0.0 – 0.77
Elgocalciferol	6	1.0	0.52	0.16 – 0.94
Flucloxacillin sodium	5	1.0	0.29	0.08 – 0.57
Liothyronine sodium	7	0.2	0.31	0.04 – 0.65
Lovastatin	5	0.4	0.25	0.02 – 0.55
Roxithromycin	5	0.4	0.44	0.25 – 0.63
Simvastatin	5*	1.5	0.35	0.23 – 0.38

* one outlier laboratory

Table 2.8-12 Repeatability and assay results from each of the laboratories participating in a collaborative trial to assign a content to a lyophilised standard of buserelin.

		Repeatability RSD (n=3)			Assay result
Laboratory	Reference solution	Test 1	Test 2 (assay)	Test 3	mg/vial
A	0,29	0,22	0,32	0,70	5,04
B	0,03	0,12	0,27	0,68	4,97
C	0,86	0,73	0,53	0,34	5,05
D	0,53	3,10	0,40	1,10	4,87
E	0,11	1,03	0,71	0,55	4,67
F*	6,85	6,46	9,07	1,64	3,02
				Mean	4,92
				Standard Deviation (σ)	0,16
				Variance (σ^2)	0,026

* Outlier

teria for replicate injections of the reference solution and unacceptable precision was also then obtained for the test solutions. This is reflected in the value of the final result reported which is clearly an 'outlier'.

Presently there is no system suitability criterion for injection repeatability imposed in the tests for related substances. However, the general method [216] for the control of residual solvents and the general method [217] for the limitation of ethylene oxide levels are applied quantitatively when there are precision requirements. In both cases the relative standard deviation should be less than 15 percent for three replicate injections. In the pharmacopoeias the control of organic impurities has traditionally been the application of TLC limit tests whereby the spot of the impurity in the chromatogram of the test solution does not exceed in size, colour and intensity, the spot in the chromatogram given by the reference solution. Even with the introduction of other separation techniques, which are easily quantifiable (gas chromatography and liquid chromatography), the same principle was upheld, i.e., the peak area of the impurity in the test solution should not exceed the area of the peak in the chromatogram given by the reference solution when external standardisation was applied. Now, it is common to limit not only individual impurities but also the sum of impurities. There is a tendency, particularly in the pharmaceutical industry, to report the results numerically and now according to ICH guidelines, the control of impurities is to be performed quantitatively with the results reported numerically [1c]. In such a situation, it will be necessary to introduce a system suitability criterion for precision, since quantification implies the performance of replicate determinations and the decisions on compliance will depend on the mean result obtained and its uncertainty. Table 2.8-13 tabulates some data generated from a number of recently conducted collaborative trials to establish assay standards. The relative standard deviation of the peak areas of replicate injections (n=6) of the reference solution at the limiting concentration for impurities is given for each of the participating laboratories. The concentrations of the reference solutions vary considerably but it can be assumed that the detector responses are similar. However, as

Table 2.8-13 Intra-laboratory precision of replicate injections of the reference solution (n=6) in tests for related substances.

Cefapirin 2 μg/mL	Roxithromyin 20 μg/mL	Cefaclor 50 μg/mL	Ciprofloxacin 1 μg/mL	Cefalexin 10 μg/mL		Clarithromycin 7.5 μg/mL	Propofol 100 μg/mL			Fenofibrate 1 μg/mL			Amlodipine besilate 3 μg/mL
										A	B	G	
1.48	0.74	0.08	0.39	0.07	0.08	1.41	0.59	1.06	0.28	0.32	0.49	0.21	1.48
2.36	1.2	0.95	0.37	0.25	0.20	1.60	2.51	0.53	0.23	0.75	1.90	1.33	1.49
0.82	0.35	0.73	0.54	0.91	0.46	3.36	0.86	1.05	1.70	0.40	1.80	0.80	1.38
2.53	0.88	0.13	1.50	0.18	0.24	4.2	0.32	0.29	0.07	0.10	0.28	0.12	
1.87	0.38	0.12	2.74	0.26	0.16	1.82	1.03	0.60					
3.0		0.70					0.28	2.62					
0.92							0.35	0.83					
							0.31						

previously indicated (Section 2.1.3.1) integration errors exert a considerable effect on the precision and accuracy at low concentrations of analyte. Nonetheless, it would seem that a maximum permitted relative standard deviation for the areas of the principal peak of replicate injections of the prescribed reference solution, could be set at 2.0 percent.

Acknowledgement

My sincere thanks are due to Miss Emma Polland for her patience, tolerance and dedication during the preparation of the texts.

3
Case Study: Validation of an HPLC-Method for Identity, Assay, and Related Impurities

Gerd Kleinschmidt

3.1
Introduction

A typical validation study on an HPLC method applied for the items 'Identity', 'Assay' and 'Related Impurities' for drug product release, is described here. The drug product is a lyophilisate with a dosage of 180 mg. Some details of the analytical procedure are given in Table 3-1. For assay of the active pharmaceutical ingredient (main component, MC) a multiple-point calibration is applied. Related impurities were validated using the specified degradation product DP1. The analytical procedure was validated with respect to its specificity for all three test items. Linearity, precision, accuracy, range, detection and quantitation limit and robustness were validated for 'Assay' of MC and for 'Related Impurities' [1a,b]. The quantitation limit for MC was validated, because the HPLC assay is also applied for the analysis of placebo batches.

Table 3-1 Specification of the HPLC method used in the case study.

Method parameter	Description	
Test method:	Liquid chromatography (Ph.Eur.)	
Equipment:	Liquid chromatograph e.g. Dionex LC system, consisting of gradient pump M 480, Autosampler GINA 160, Detector UVD 320 or equivalent	
Column:	Material:	stainless steel
	Length:	125 mm
	Internal diameter:	4 mm
Stationary phase:	Superspher 60 RP-select B, 4 µm, or equivalent	
Mobile phase A:	Water	850 ml
	Acetonitrile R	150 ml
	Phosphoric acid R	1 ml
	Sodium chloride	1 g
Mobile phase B:	Water	450 ml
	Acetonitrile R	550 ml
	Phosphoric acid R	1 ml
	Sodium chloride	1 g

Method Validation in Pharmaceutical Analysis. A Guide to Best Practice. Joachim Ermer, John H. McB. Miller (Eds.)

ISBN: 3-527-31255-2

Table 3-1 Continued.

Method parameter	Description
Preparation of mobile phases:	Mix the water with acetonitrile R, add phosphoric acid R and dissolve the amount of sodium chloride. Then adjust the pH with 10N NaOH to 3.6.
Gradient:	Time (min) / % Phase A / % Phase B 0 / 100 / 0 10 / 0 / 100 20 / 0 / 100 21 / 100 / 0 25 / 100 / 0
Preparation of test solution:	Take four vials of the lyophilisate to be examined and dissolve each in 36.0 ml of water. Mix the samples in a beaker. 2.0 ml of this solution are diluted to 100.0 ml, using acetonitrile 30 % as solvent. The clear solution obtained after shaking is used as test solution. Concentration obtained: 0.1 mg/ml. Prepare at least two test solutions and inject each twice.
Injection volume:	10 µl
Flow:	1.0 ml/min
Run time:	20 min
Detection:	UV at 246 nm
Identification:	t_R(sample) = t_R(reference) ± 5 %
Acceptance limits and quantification	
Acceptance limits Assay:	95.0 to 105.0 % label claim
Acceptance limits Related Impurities:	
Degradation product DP1:	≤ 1.0 %
Any other individual unspecified impurity:	≤ 0.10 %
Total impurities:	≤ 1.5 %
Quantification of MC (main component):	Multi-point calibration Prepare three standard solutions as follows and inject them at least twice: SS1-3: Dissolve 7.5, 10.0, and 12.5 mg of MC reference standard in 100.0 ml of acetonitrile 30 % to obtain a solution of 75, 100, and 125 µg/ml, respectively. The calibration curve is calculated via linear regression without intercept ($y = b \cdot x$) using a suitable software system (e.g. Chromeleon) based on the weights and the corresponding areas of the standard solutions SS1-3. The assay is calculated in mg/vial using the calibration curve and taking into account the dilution factor of the test sample.
Quantification of Related Impurities:	The amount is calculated by peak area normalization (100 % standard). Each impurity peak is related to the sum of all peaks, apart from mobile phase generated ones. For DP1, the response factor of 1.3 is taken into account.

Prior to the start of the validation experiments, the design and the acceptance criteria were defined in a validation plan. A tabular overview is provided in Tables 3-2 to 3-4.

3.2 Experimental

In the validation study, analyst 1 used an LC system (no. 5) with Chromeleon acquisition software (Version 6.2) (DIONEX, Germering, Germany) and a Superspher 60 RP-select B column, no. 048909 (MERCK, Darmstadt, Germany). Analyst 2 utilised also a DIONEX system (no. 1) and the same column type (no. 432040). The batches used were internally characterised reference standards of the drug substance (MC), the degradation product DP1, and potential process related impurities SP1, SP2, and SP3 with contents of 99.9%, 97.0%, 98.4% and 92.3%, respectively. The excipient P1 was purchased from Riedel de Haen, with a content of ≥ 99.7%.

Test solutions were prepared according to the analytical procedure, placebo preparations in accordance with drug product composition.

All calculations were performed using the software MVA 2.0 [28].

3.3 Validation Summary

All parameters defined in the validation plan met their acceptance criteria. A tabular summary is shown in Tables 3-2 to 3-4.

Table 3-2 Validation protocol and summary of the test items 'Identity', 'Assay', and 'Related Impurities'.

Validation characteristic	Results			
	Acceptance criteria	Complies	Does not comply	Remarks
Specificity	Complete separation of MC, DP1, SP1, SP2, SP3 and no interfering placebo peaks	☒	☐	List relative retention times rrt and resolutions R_s.

Table3-3 Validation protocol and summary of the test item 'Assay' (active pharmaceutical ingredient MC).

Validation characteristic	Acceptance criteria	Result			Remarks
		Test result	Complies	Does not comply	
	No deviation from linear response function (residual plot)	Random scatter of residuals	☒	☐	Unweighted linear regression (40 to 130 % label claim) $y = a + b \cdot x$.
Linearity	CI[1] of intercept a includes 0 (acceptable deviations to be justified)	a = –0.408522 CI: –0.98 to 0.16	☒	☐	
	Coefficient of correlation ≥ 0.999	r = 0.99979	☒	☐	
	Test according to Mandel (acceptable deviations to be justified)	Yes	☒	☐	No significant better fit by quadratic regression.
Accuracy	Mean recovery: 98.0 to 102.0 %	100.7 %	☒	☐	3 × 3 concentrations, percent recovery.
Precision	System precision ≤ 1.0 %	0.36 %; 0.20 %	☒	☐	
	Repeatability ≤ 2.0 %	0.62 %; 0.88 %	☒	☐	Experiments performed by two operators.
	Intermediate precision ≤ 3.0 %	0.83 %	☒	☐	
Limit of quantitation	LOQ ≤ 0.05 %[2)]	LOQ = 0.05 %	☒	☐	
	RSD_{LOQ} ≤ 10 %	RSD_{LOQ} = 6.9 %	☒	☐	Validated with MC, required for analysis of placebo batches.
	Mean recovery: 90.0 to 110.0 %	102.3 %	☒	☐	

1: 95 % confidence interval.
2: Corresponds to ICH reporting threshold.

Table 3-4 Validation protocol and summary of the test item 'Related Impurities' (degradation product DP1).

Validation characteristic	Acceptance criteria	Result			Remarks
		Test result	Complies	Does not comply	
Linearity	No deviation from linear response function (residual plot)	Random scatter of residuals	☒	☐	Unweighted linear regression $y = a + b \cdot x$.
	CI[1] of intercept a includes 0 (acceptable deviations to be justified)	a = –0.00091 CI: –0.0028 – 0.0010	☒	☐	
	Coefficient of correlation ≥ 0.99	r = 0.99990	☒	☐	
	Test according to Mandel (acceptable deviations to be justified)	Yes	☒	☐	No significant better fit by quadratic regression.
Accuracy	Mean recovery: 90.0 to 110.0 %	101.6 %	☒	☐	
Precision	System precision ≤ 1.0 %	0.87 %; 0.35 %	☒	☐	Experiments performed by two operators. Analyses conducted at 1 % concentration of DP1. Based on the experience gained it was possible to set identical acceptance criteria for MC and DP1.
	Repeatability ≤ 2.0 % Intermediate precision ≤ 3.0 %	1.12 %; 0.36 % 1.67 %	☒	☐	
Limit of quantitation	LOQ ≤ 0.05 %[2)]	0.05 %	☒	☐	
	RSD_{LOQ} ≤ 10 %	3.1 %; 2.1 %	☒	☐	
	Mean recovery: 90.0 to 110.0 %	99.1 %; 99.2 %	☒	☐	

1: 95 % confidence interval.
2: Corresponds to ICH reporting threshold.

3.3.1 Specificity

The specificity of the analytical procedure was demonstrated by a complete chromatographic separation of MC from three potential process-related impurities (SP1, SP2, SP3) and from the degradation product DP1. Furthermore, it was shown that the drug product matrix component P1 interferes neither with MC nor with the aforementioned process-related impurities and degradation products.

3.3.2 Linearity

The linearity of the test procedure was validated in the range 40 % – 130 % of the theoretical sample preparation for the active ingredient MC via graphical evaluation of the data and the evaluation of the calibration curve via linear regression. A linear response function as well as a negligible intercept was demonstrated, justifying a three-point calibration that includes the origin (i.e. forced through zero) in routine analyses. In addition the linearity of the test procedure was proven in the range 0.025 % – 1.3 % for the specified degradation product DP1. Routine analyses are carried out applying three-point calibrations with MC and the respective response factor of DP1 for calculating its amount.

3.3.3 Precision

The relative standard deviations of 0.36 % and 0.62 % for system precision and repeatability for the assay of MC in authentic lyophilisate batches are acceptable.

For the specified related impurity DP1 relative standard deviations of 0.87 % and 1.12 % for system precision and repeatability were found (each at 1 % of MC).

A second analyst could demonstrate adequate intermediate precision. The relative standard deviations of 0.20 % and 0.88 % for system precision and repeatability, respectively, for the determination of assay of MC are very close to those of analyst 1 and therefore acceptable.

For the specified degradation product DP1 relative standard deviations of 0.35 % and 0.36 % for system precision and repeatability were determined. These results also demonstrate good agreement between the two analysts data.

3.3.4 Accuracy

The accuracy of the analytical procedure for the determination of assay of MC was demonstrated by a mean recovery of 100.7 % for three spikings at three concentration levels, i.e., 80, 100, and 120 %.

The accuracy of the analytical procedure for the determination of related impurities was demonstrated by a mean recovery of 101.6 % for DP1 throughout a working range of approximately 0.025 % – 1.3 %.

3.3.5 Detection and Quantitation Limit

For the specified degradation product DP1 the detection and quantitation limits were determined. The results obtained support a detection limit of 0.01 % and a quantitation limit of 0.05 % of the working concentration of MC (as it is in case of MC itself).

3.3.6 Robustness

The robustness of the analytical procedure was investigated as described in Chapter 2.7.

3.3.7 Overall Evaluation

An adequate degree of linearity, accuracy and precision was demonstrated for MC within a range of 80 % – 120 % and for DP1 within a range of 0.05 % – 1.3 %. The results of this validation study confirm the suitability of the analytical procedure for the determination of identity, assay and related impurities of MC.

3.4 Validation Methodology

3.4.1 Specificity

A test solution comprising MC and 1 % of the potential process related impurities SP1, SP2 and SP3 and of the degradation product DP1 was prepared and analysed. The chromatogram of the test solution (Fig. 3-1, No. 3) confirms that all impurities are completely separated from MC. The retention times and the resolutions of the peaks are listed in Table 3-5.

The chromatogram of a degraded sample (Fig. 3-1, No.2) proves additionally that the degradation product DP1 does not interfere with the detection of MC. The chromatogram of the placebo solution (Fig. 3-1, No.1) demonstrates that the excipients do not interfere either with the detection of MC or the impurities.

The presented chromatograms and peak purity analyses of the MC peak by means of HPLC-MS (not detailed here) confirm that the analytical procedure is suitable to determine MC specifically in the presence of its relevant impurities and the placebo component P1, as well as the impurities without interference from each other.

Table 3-5 Retention times and resolution of the principal peaks in the specificity solution.

Compound	Origin of substance	Retention time		Resolution[1]
		Absolute [min]	Relative	
MC	Active pharmaceutical ingredient	5.86	1.00	14.27
DP1	Degradation product	8.41	1.44	2.47
SP1	Process-related impurity	8.87	1.51	17.47
SP2	Process-related impurity	12.25	2.09	7.66
SP3	Process-related impurity	14.16	2.42	–

1: ... between the respective peak and the following.

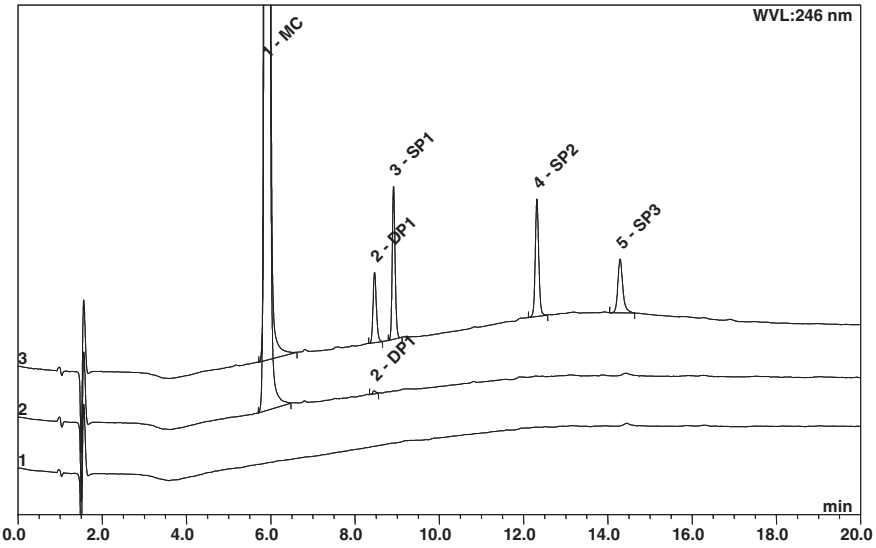


Figure 3-1 Chromatograms of a solution comprising the main component (MC), process-related impurities (SP1-3) and the degradation product (DP1) at 1 % of the working concentration (No. 3), of a degraded MC sample (No. 2), and of a placebo solution (No. 1).

3.4.2 Linearity

3.4.2.1 Linearity of MC (for test item assay)

Ten sample solutions of MC dissolved in acetonitrile 30 % were prepared in order to obtain a concentration range from 40 to 130 % of the test concentration 0.10 mg/ml. The results for the evaluation of the linearity are given in Table 3-6. The graphical presentations of the peak areas plot obtained for MC against the concentration of the test solution, as well as the residual plot of MC, are given in the Figures 3-2 and 3-3, respectively. In addition to the linear regression analysis and the graphical pre-

Table 3-6 Results for the evaluation of the linear relationship between the peak area of MC and DP1 and their concentrations. The linearity studies were performed using LC-system 1 and LC-system 5, respectively.

Sample no.	Active Pharmaceutical Ingredient (MC)		Degradation Product (DP1)	
	Concentration[1] [mg/ml] / (% label claim)	Peak area[2] [mAU · min]	Concentration[1] [µg/ml] / (% label claim)	Peak area[2] [mAU · min]
1	0.0411 / (40 %)	14.8523	0.02587 / (0.025 %)	0.00608
2	0.0503 / (50 %)	18.3688	0.05174 / (0.05 %)	0.01219
3	0.0601 / (60 %)	22.2653	0.07762 / (0.075 %)	0.01791
4	0.0709 / (70 %)	26.2577	0.10349 / (0.1 %)	0.02401
5	0.0816 / (80 %)	29.7511	0.25874 / (0.25 %)	0.05763
6	0.0930 / (90 %)	34.1243	0.51748 / (0.5 %)	0.11850
7	0.1006 / (100 %)	36.7359	0.77622 / (0.75 %)	0.18429
8	0.1110 / (110 %)	40.9953	1.03496 / (1.0 %)	0.24456
9	0.1202 / (120 %)	44.3249	1.13846 / (1.1 %)	0.26625
10	0.1301 / (130 %)	48.4543	1.24195 / (1.2 %)	0.29418
11			1.34545 / (1.3 %)	0.31536
Unweighted linear regression: y= a+b · x				
Slope	b= 372.83		b= 0.2360	
Intercept	a= – 0.41		a= -0.00091	
95 % Confidence interval	–0.98 to 0.16		–0.00282 to 0.00100	
Residual standard deviation	0.2462		0.00180	
Relative standard error of slope	0.77 %		1.28 %	
Coefficient of correlation	r= 0.99979		r= 0.99990	

1: ... of the test solution [mg/ml] / claim of the theoretical test sample concentration [%].
2: Mean of two injections.

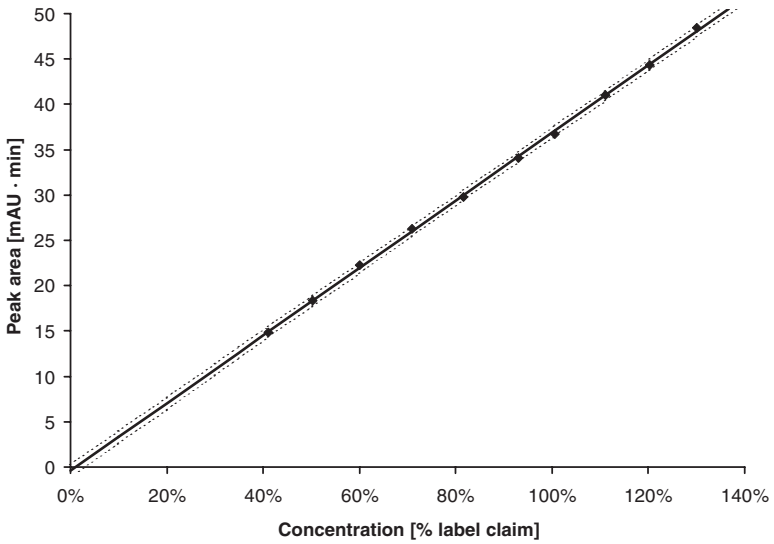


Figure 3-2 Peak area of MC as a function of its concentration. Besides the experimental data points and the unweighted linear regression line, the 95% prediction intervals (dotted line) are indicated.

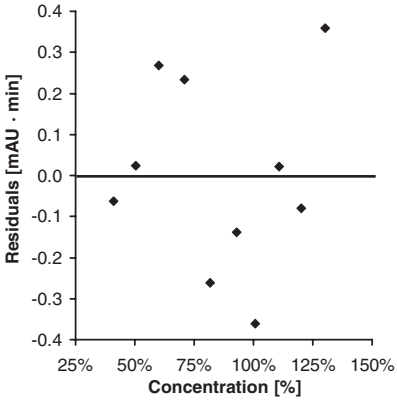


Figure 3-3 Residual plot for the linear regression analysis of MC. The scale of the y-axis corresponds to ±1.1% of the signal at 100% working concentration.

sentations, the Mandel test was performed, which is a statistical linearity test, revealing no significant better fit by quadratic regression. These results clearly proved a linear relationship between the MC concentration in the test solution and its corresponding peak area.

The confidence interval of the y-intercept includes zero. Routine analyses will be carried out by performing a three-point calibration, that includes the origin, to further minimize analytical uncertainties.

3.4.2.2 Linearity of DP1 (for test item Related Impurities)

The linearity was proven for DP1. Eleven sample solutions were prepared containing the drug product matrix component P1 in the same concentration as in drug product samples (0.15 mg/ml), MC at 0.10 mg/ml concentration. The samples were spiked with DP1 to obtain a concentration range from 0.025 % (LOQ estimated from previous validation studies) to 1.3 % related to the working concentration of MC, which corresponds to 3 % – 130 % related to the DP1 specification limit of 1.0 %.

The results for the evaluation of the linearity of the related impurity DP1 are given in Table 3-6. The graphical presentations of the plot of the peak areas obtained for DP1 against the concentration of the test solution as well as the residual plot of DP1 are shown in Figures 3-4 and 3-5, respectively. In addition to the linear regression analysis and the graphical presentations, the Mandel test was performed. This test revealed no significant better fit by quadratic regression. These results clearly demonstrate a linear relationship.

The confidence interval of the y-intercept includes zero. Therefore, the prerequisite for an area normalisation (100% standard) is fulfilled.

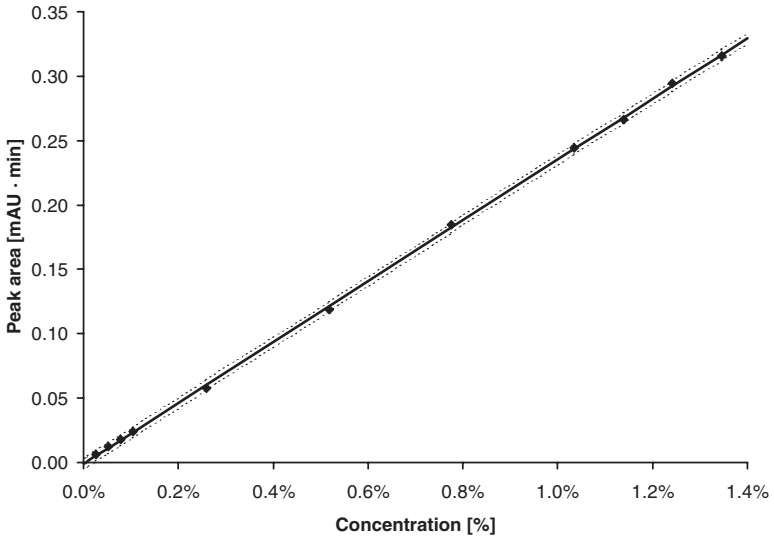


Figure 3-4 Peak area of DP1 as a function of its concentration. Besides the experimental data points and the unweighted linear regression line, the 95% prediction intervals (dotted line) are indicated.

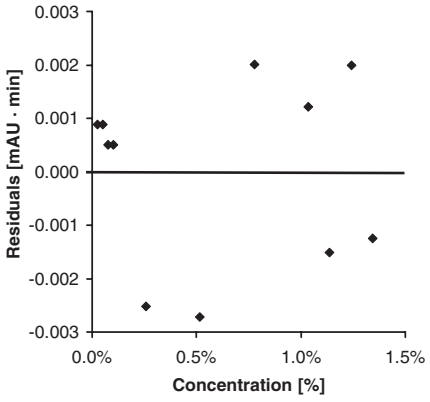


Figure 3-5 Residual plot for the linear regression analysis of DP1. The scale of the y-axis corresponds to ±1.3% of the signal at 100% working concentration.

3.4.3 Accuracy

3.4.3.1 Accuracy of MC (Assay)

For the determination of the validation parameter accuracy, an approach according to ICH and a calibration in accordance with the control test (three-point calibration with 0.075, 0.10, 0.125 mg/ml MC standard solutions) was chosen in this validation study. The test preparation containing the drug product matrix component P1 in the

Table 3-7: Results for the recovery of the MC from spiked placebo.

Sample no.	MC added [mg] and [% claim]	MC found [mg]	Recovery [%]
1	8.13 / 80	8.17	100.5
2	8.26 / 80	8.31	100.7
3	8.11 / 80	8.23	101.4
4	10.16 / 100	10.23	100.7
5	10.15 / 100	10.28	101.3
6	10.23 / 100	10.36	101.2
7	12.03 / 120	12.01	99.9
8	12.07 / 120	12.10	100.3
9	12.12 / 120	12.18	100.5
	Mean recovery [%]		100.7
	95 % Confidence interval		100.3 to 101.1
	RSD [%]		0.52

same concentration as in drug product samples (0.15 mg/ml) was spiked with accurate amounts of MC, corresponding to approximately 80, 100 and 120 % of label claim, three times each, i.e., at nine concentrations. The percentage recoveries (see Table 3-7) were calculated.

The mean recovery for all concentration levels was calculated to 100.7 % and the relative standard deviation to 0.52 %. The 95 % confidence interval ranges from 100.3 % to 101.1 %. Consequently, the theoretical value of 100 % is not included. However, the deviation from the theoretical recovery is small and the requirement for mean recovery in this validation study (see Table 3-3) is met. No practically relevant dependency of the recovery from the concentration level is observed (Fig. 3-6).

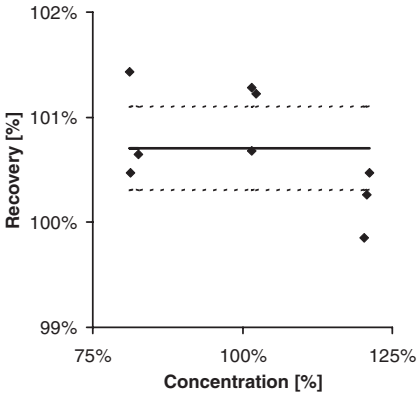


Figure 3-6 Recovery of MC from spiked placebo. The mean recovery and its 95% confidence limits are indicated by solid and dotted line(s), respectively.

3.4.3.2 Accuracy of DP1 (Related Impurities)

The procedure described below is based on peak area normalization (100 % standard) taking the response factor of DP1 into consideration.

To evaluate the accuracy, eleven sample solutions were prepared. The test preparation containing the drug product matrix component P1 in the same concentration as in drug product samples (0.15 mg/ml) and MC at 0.10 mg/ml concentration, was spiked with DP1 to obtain a concentration range from 0.025 % (LOQ estimated from previous validation studies) to 1.3 % related to the working concentration of MC, corresponding to 3 % – 130 % related to the DP1 specification limit of 1.0 %.

The percentage recoveries for DP1 were calculated and are summarized in Table 3-8. The mean recovery for all concentration levels was calculated to 101.6 % and the relative standard deviation to 2.9%. The 95 % confidence interval ranges from 99.6 % to 103.6 %. Consequently, the theoretical value of 100 % is included. No sig-

Table 3-8 Results for the recovery of DP1 from spiked placebo and MC.

Sample no.	DP1 added [%]	DP1 found [%]	Recovery [%]
1	0.0259	0.0279	107.7
2	0.0517	0.0550	106.4
3	0.0776	0.0790	101.8
4	0.1035	0.1050	101.4
5	0.2587	0.2534	98.0
6	0.5175	0.5132	99.2
7	0.7762	0.7879	101.5
8	1.0350	1.0387	100.4
9	1.1385	1.1361	99.8
10	1.2420	1.2543	101.0
11	1.3455	1.3471	100.1
	Mean recovery [%]		101.6
	95 % Confidence interval		99.6 to 103.6
	RSD [%]		2.9

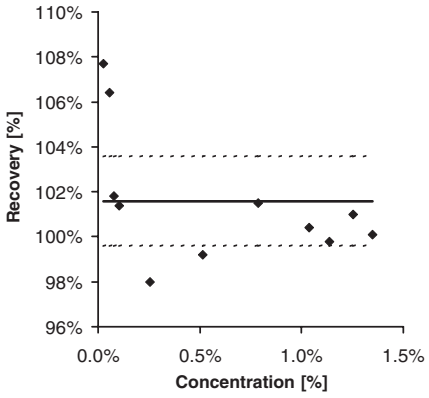


Figure 3-7 Recovery of DP1 from spiked placebo and MC. The mean recovery and its 95% confidence limits are indicated by solid and dotted line(s), respectively.

nificant dependency on the recovery from the concentration is observed (Fig. 3-7). The response factor was calculated to 1.3 using the slopes obtained from the quantitation limit studies on DP1 and the active ingredient MC (see Table 3-12).

3.4.4 Precision

The precision of the method was confirmed by investigations of the system precision, repeatability and intermediate precision.

3.4.4.1 System Precision

The system precision of the method was proved by seven injections of one sample solution of drug product. Furthermore, a sample solution was prepared containing the related impurities DP1, SP1, SP2 and SP3 at 1 % of the MC working concentration. This solution that also contained the drug product matrix component P1 in the same concentration as in the drug product samples (0.15 mg/ml) and MC at 0.10 mg/ml concentration, was injected seven times. A second analyst also performed the same analyses. The results are summarized in Table 3-9. The relative standard deviations below 1 % for all components confirm an acceptable degree of system precision and comply with the requirement defined in the validation plan for the parameter system precision.

Table 3-9 Results for the determination of system precision of MC and DP1, SP1, SP2 and SP3 at 0.10 mg/ml and at 0.001 mg/ml each, respectively.

	Peak area [mAU · min]				
Sample no.	**MC**	**DP1**	**SP1**	**SP2**	**SP3**
1	32.55459	0.27471	0.57982	0.51694	0.34444
2	32.64139	0.26991	0.57465	0.51275	0.33865
3	32.62365	0.27584	0.57604	0.52068	0.33929
4	32.74303	0.27484	0.57422	0.51878	0.34493
5	32.81275	0.27005	0.56733	0.50974	0.34016
6	32.61518	0.27198	0.57470	0.51723	0.34286
7	32.87590	0.27353	0.56669	0.50992	0.33773
Mean Value	32.700	0.27298	0.57335	0.51515	0.34115
RSD [%]	0.36	0.87	0.82	0.84	0.85
RSD [%] Analyst 2	0.20	0.35	0.44	0.66	0.71

3.4.4.2 Repeatability

The repeatability of the method (with regard to MC) was investigated by analysing seven samples each at 100 % of the test concentration. In addition to that a drug product sample spiked with 1 % DP1 was analysed seven times to evaluate the repeatability of the determination of the DP1 at its specification limit of 1 %. The results obtained by two analysts are summarized in Table 3-10. The relative standard

Table 3-10 Results for the determination of repeatability and intermediate precision.

Sample no.	MC, content [mg] / vial		DP1, content [%]	
	Analyst 1	Analyst 2	Analyst 1	Analyst 2
1	180.749	181.251	1.06004	1.06727
2	179.244	181.058	1.04758	1.06540
3	177.457	177.162	1.05813	1.06838
4	178.181	181.027	1.02484	1.07031
5	179.494	180.462	1.04675	1.07437
6	179.251	178.394	1.04058	1.06262
7	177.981	180.803	1.04609	1.07082
Mean	178.9	180.0	1.046	1.068
95 % Confidence Interval	177.9 to 179.9	178.5 to 181.5	1.035 to 1.057	1.065 to 1.072
RSD [%]	0.62	0.88	1.12	0.36
Overall mean	179.5		1.057	
95 % Confidence interval	178.6 to 180.3		1.047 to 1.068	
Overall repeatability [%]	0.77		0.82	
Intermediate precision [%]	0.83		1.67	

deviations of 0.88 % and 0.62 % confirm an acceptable degree of repeatability for the determination of assay of MC lyophilisate.

The relative standard deviations of 1.12 % and 0.36 % confirm an acceptable degree of repeatability for the determination of the related impurity DP1. All results meet the acceptance criterion defined in the validation protocol.

3.4.4.3 Intermediate Precision

The intermediate precision was proved by investigations with variation of time, analysts and equipment (including columns, reagents, etc.). Therefore, a second analyst carried out all experiments described in Section 3.4.4.2 as well (Table 3-10). The overall repeatability of 0.77 % and the intermediate precision of 0. 83 % as well as their good agreement, confirm an acceptable degree of precision for the determination of MC.

The overall repeatability below 1.0 % and the intermediate precisions below 2.0 % and their good agreement (Table 3-10) confirm adequate precision for the determination of the degradation product DP1 at its specification limit (1 % of the MC working concentration).

All results reported in this section for the validation characteristic intermediate precision fulfil the acceptance criterion RSD ≤ 3.0 % defined in the validation plan.

3.4.5 Range

The range for the determination of MC and DP1 is defined from linearity, accuracy and precision of the analytical procedure.

The analytical procedure provides an acceptable degree of linearity, accuracy and precision for MC and DP1 in the range of 80 – 120 % and 0.05 – 1.3 % of the nominal MC concentration (see also Section 3.4.6).

3.4.6 Detection Limit and Quantitation Limit

3.4.6.1 Detection Limit and Quantitation Limit of MC

For analysis of the MC lyophilisate placebo formulation it is mandatory to show that the placebo does not contain the active ingredient. For that reason the detection and the quantitation limit for MC need to be determined. Evaluation of both parameters was based on the regression line.

From MC, six test solutions were prepared by spiking certain aliquots of a reconstituted placebo formulation to obtain a concentration range from 0.01 % to 0.25 % related to the working concentration of MC (see Table 3-11). Based on the results of the calibration curve for MC and the residual standard deviation, a detection limit of 0.0039 µg/ml was calculated, corresponding to 0.004 % of the working concentration of MC (set to 0.01 % for practical reasons). A quantitation limit of 0.034 µg/ml (10% acceptable relative uncertainty) was calculated corresponding to 0.03 % of the working concentration of MC (see Table 3-12).

The limit of quantitation (LOQ) was verified by analysing one sample containing MC at LOQ concentration level. For practical reasons the ICH reporting level of 0.05 % was chosen. The test solution prepared was injected seven times and the mean recovery and RSD were calculated (see Table 3-13). This study revealed a mean recovery of 102 % and a RSD of 6.9 %. Both parameters meet the acceptance criteria defined in the validation plan.

Table 3-11 Linearity of active (MC) and degradation product (DP1) for determination of detection and quantitation limit.

	MC		DP1	
Sample No.	**Concentration [µg/ml]**	**Peak area [mAU · min]**	**Concentration [µg/ml]**	**Peak area [mAU · min]**
1	0.0100994	0.003108	0.025874	0.00608
2	0.0252486	0.008015	0.051748	0.01219
3	0.0504972	0.015540	0.077622	0.01791
4	0.0757458	0.022415	0.103496	0.02401
5	0.1009944	0.029510	0.258741	0.05763
6	0.2524860	0.074235		

Table 3-12 Determination of detection and quantitation limit for active (MC) and degradation product (DP1) from unweighted linear regression ($y = a + b \cdot x$).

Parameter / Calculation	MC	DP1
Slope:	b= 0.2914	b= 0.2205
Relative confidence interval (95 %):	+/– 2.11 %	+/– 2.76 %
Intercept:	a = 0.000513	a = 0.000744
Standard deviation:	0.000249	0.00252
Confidence interval (95 %):	–0.000278 to 0.000130	–0.000058 to 0.00155
Residual standard deviation:	0.000345	0.000349
Relative standard error of slope:	1.17 %	1.53 %
Coefficient of correlation:	0.99993	0.99989
Calculation from residual SD		
Detection limit:	0.003903	0.00523
Quantitation limit:	0.018281	0.01583
Calculation from the 95 % prediction interval		
Detection limit [µg/ml]	0.008	0.012
Quantitation limit [µg/ml]	0.012	0.018
Calculation according to DIN 32645		
Detection limit [µg/ml]	0.003	0.005
Quantitation limit [µg/ml]	0.010	0.018
Factor k (1/relative uncertainty)	3.00 (33.33 %)	3.00 (33.33 %)
Calculation from the relative uncertainty		
Detection limit [µg/ml] (ARU 50 %)	0.007	0.012
Quantitation limit [µg/ml] (ARU 33 %)	0.011	0.018
Quantitation limit [µg/ml] (ARU 10 %)	0.034	0.057

1: ARU = acceptable relative uncertainty

Table 3-13 Recovery of active (MC) and degradation product (DP1) at LOQ concentration level.

Sample no.	MC (0.051 µg/ml added)		DP1 (0.0506 µg/ml added)	
	Analyte found [µg/ml]	Recovery [%]	Analyte found [µg/ml]	Recovery [%]
1	0.051	100.6	0.048	94.9
2	0.055	108.5	0.052	102.8
3	0.056	110.5	0.049	96.8
4	0.045	88.8	0.051	100.8
5	0.052	102.6	0.049	96.8
6	0.051	100.6	0.050	98.8
7	0.053	104.5	0.052	102.8
Mean recovery [%]	102.3		99.1	
95 % Confidence interval	95.8 to 108.8		96.2 to 102.0	
RSD [%]	6.90		3.14	

3.4.7
Detection Limit and Quantitation Limit of DP1

The detection limit and quantitation limit for DP1 were determined based on the regression line. Five test solutions were prepared. The test preparation containing the drug product matrix component P1 in the same concentration as in drug product samples (0.15 mgl/ml) and also MC at 0.10 mg/ml concentration, was spiked with DP1 to obtain a concentration range from 0.025 % to 0.25 % related to the working concentration of MC (see Table 3-11).

Based on the results of the calibration curve for DP1 and the residual standard deviation, a detection limit of 0.0052 µg/ml was calculated corresponding to 0.005 % of the working concentration of MC (set to 0,01 % for practical reasons). A quantitation limit of 0.057 µg/ml (10 % acceptable relative uncertainty) was calculated corresponding to 0.06 % of the working concentration of MC (see Table 3-12).

The limit of quantitation was verified by analysing one sample containing DP1 at LOQ concentration level (for practical reasons at ICH reporting level 0.05 %) and P1 and MC at 0.15 mg/ml and 0.10 mg/ml, respectively. The test solution prepared was injected seven times and the mean recovery and RSD were calculated (see Table 3-13). The study revealed a mean recovery of 99 % and a RSD of 3.1 %. Both parameters meet the acceptance criteria defined in the validation plan.

3.4.8
Robustness

For guidance on performing robustness studies see Section 2.7, where detailed explanations supplemented by some examples are given.

3.5
Conclusion

The results of this validation study confirm the suitability of the analytical procedure for the determination of identity, assay and related impurities of MC (range for the determination of assay: 80 % – 120 %; range for the determination of DP1: 0.05 % – 1.3 %).

References Part I

[1] International Conference on the Harmonization of Technical Requirements for the Registration of Pharmaceuticals for Human Use (ICH), http://www.ich.org/UrlGrpServer.jser?@_ID=276&@_TEMPLATE=254
a) ICH Q2A: Validation of Analytical Methods (Definitions and Terminology), 1994
b) ICH Q2B Analytical Validation – Methodology, 1996
c) ICH Q3A(R), Impurities in New Drug Substances, 2002
d) ICH Q3B(R), Impurities in New Drug Products, 2003
e) ICH Q6A, Specifications: Test Procedures and Acceptance Criteria for New Drug Substances and New Drug Products, Chemical Substances, 1999
f) ICH Q7A: GMP Guide to Active Pharmaceutical Ingredients, 2000
g) ICH Q1A(R2): Stability Testing of New Drug Substances and Products, 2003.

[2] Guidelines for Submitting Samples and Analytical Data for Methods Validation, US Food and Drug Administration, Centre for Drugs and Biologics, Department of Health and Human Services, 1987.

[3] CDER Guideline on Validation of Chromatographic Methods, Reviewer Guidance of Chromatographic Methods, US Food and Drug Administration, Centre for Drugs and Biologics, Department of Health and Human Services, 1994.

[4] Draft Guidance Analytical Procedures and Methods Validation. US Food and Drug Administration, Centre for Drugs and Biologics, Department of Health and Human Services, 2000. http://www.fda.gov/cder/guidance/2396dft.htm#III

[5] United States Pharmacopoeia 24, National Formulary 19, Section <1225> "Validation of Compendial Methods", United States Pharmacopoeial Convention, Rockville, 2000.

[6] The Rules Governing Medicinal Products in the European Community, Volume 3, Addendum 1990.

[7] Acceptable Methods. Drug Directorate Guidelines, National Health and Welfare, Health Protection Branch, Health and Welfare Canada, 1992.

[8] EURACHEM: The fitness for purpose of analytical methods. A Laboratory Guide to Method Validation and Related Topics. http://www. eurachem.ul.pt, 1998.

[9] G.C. Hokanson: A life cycle approach to the validation of analytical methods during pharmaceutical product development, Part I. Pharm. Technol. 1994, *18 (10)*, 92–100; Part II. Pharm. Technol. 18 (1994) 118–130.

[10] J. Ermer and H.-J. Ploss: Validation in Pharmaceutical Analysis Part II: Central importance of precision to establish acceptance criteria and for verifying and improving the quality of analytical data. J. Pharm. Biomed. Anal. in press (2004).

Method Validation in Pharmaceutical Analysis. A Guide to Best Practice. Joachim Ermer, John H. McB. Miller (Eds.)

ISBN: 3-527-31255-2

[11] European Pharmacopoeia: Chapter III Analytical Validation. Pharmeuropa Technical Guide (1996) 28–40.
[12] M. Bakshi and S. Singh: Development of validated stability-indicating assay methods – critical review. J. Pharm. Biomed. Anal. 28 (2002) 1011–1040.
[13] T. Mirza, M.J. Lunn, F.J. Keeley, R.C. George, J.R. Bodenmiller: Cleaning level acceptance criteria and a high pressure liquid chromatography procedure for the assay of Meclizine Hydrochloride residue in swabs collected from pharmaceutical manufacturing equipment surfaces. J. Pharm. Biomed. Anal. 19 (1999) 747–756.
[14] A.G.J. Daas and J.H.M. Miller: Relationship between content limits, System Suitability for Precision and Acceptance/Rejection criteria for Assays Using Chromatographic Methods. Pharmeuropa 11 (1999) 571–577.
[15] European Pharmacopoeia 4th Ed: 2.2.46 Chromatographic Separation Techniques, Council of Europe, Strasbourg (2002).
[16] USP: <1010> Analytical Data – Interpretation and treatment. Pharmacopeial Forum 30 (2004) 236–263.
[17] EURACHEM: Guide on selection, use and interpretation of proficiency testing schemes, 2000. http://www.eurachem.ul.pt
[18] A.G.J. Daas and J.H.M. Miller: Content limits in the European Pharmacopoeia (Part 1). Pharmeuropa 9 (1997) 148–156.
[19] W. Horwitz: The variability of AOAC methods of analysis as used in analytical pharmaceutical chemistry. J. Assoc. of Anal. Chem. 60 (1977) 1355–1363.
[20] DIN ISO: 5725-2: Accuracy (trueness and precision) of measurement methods and results; a basic method for the determination of repeatability and reproducibility of a standard measurement method, 1990.
[21] EURACHEM/CITAC Guide: Quantifying uncertainty in analytical measurement 2nd edition, 2000. http://www.eurachem.ul.pt
[22] Analytical Methods Committee: Robust statistics: a method of coping with outliers. amc technical brief No. 6, April 2001. http://www.rsc.org/lap/rsccom/amc/amc_techbriefs.htm
[23] P.J. Huber: Robust statistics. Wiley & Sons, New York, 1981.
[24] Analytical Methods Committee: The bootstrap: A simple approach to estimating standard errors and confidence intervals when theory fails. amc technical brief No. 8, August 2001. http://www.rsc.org/lap/rsccom/amc/amc_techbriefs.htm
[25] Y. Hayashi and R. Matsuda: Deductive prediction of measurement precision from signal and noise in LC. Anal. Chem. 66 (1994) 2874–2881.
[26] T. Anglov, K. Byrialsen, J.K. Carstensen, et al.: Uncertainty budget for final assay of a pharmaceutical product based on RP-HPLC. Accred. Qual. Assur. 8 (2003) 225–230.
[27] G. Maldener: Requirements and Tests for HPLC Apparatus and methods in pharmaceutical quality control. Chromatographia 28 (1989) 85–88.
[28] Software MVA 2.0 – Method Validation in Analytics (2001) http://www.novia.de
[29] DIN ISO: 5725-3: Accuracy (trueness and precision) of measurement methods and results – Part 3: Intermediate measures on the precision of a test method (1991).
[30] S. Burke: Analysis of variances. LC-GC Europe Online Supplement statistics and data analysis (2001) 9–12. http://www.lcgceurope.com/lcgceurope/article/articleList.jsp?categoryId=935.
[31] R. Ficarra, P. Ficarra, S. Tommasini, S. Melardi, M.L. Calabro, S. Furlanetto, M. Semreen: Validation of a LC method for the analysis of zafirlukast in a pharmaceutical formulation. J. Pharm. Biomed. Anal. 23 (2000) 169–174.
[32] Ch. Ye, J. Liu, F. Ren, N. Okafo: Design of experiment and data analysis by JMP (SAS institute) in Analytical Method Validation. J. Pharm. Biomed. Anal. 23 (2002) 581–589.

[33] S. Gorog: Chemical and analytical characterization of related organic impurities in drugs. Anal. Bioanal. Chem. 377 (2003) 852–862.
[34] R. Ficarra, M.L. Calabro, P. Cutroneo, S. Tommasini, S. Melardi, M. Semreen, S. Furlanetto, P. Ficarra, G. Altavilla: Validation of a LC method for the analysis of oxaliplatin in a pharmaceutical formulation using an experimental design. J. Pharm. Biomed. Anal. 29 (2002) 1097–1103.
[35] J. De Beer, P. Baten, C. Nsengyumva, J. Smeyers-Verbeke: Measurement uncertainty from validation and duplicate analysis results in HPLC analysis of multivitamin preparations and nutrients with different galenic forms. J. Pharm. Biomed. Anal. 32 (2003) 767–811.
[36] S. Kojima: Evaluation of intermediate precision in the validation of analytical procedures for drugs. Pharm. Tech. Japan 18 (2002) 51–59.
[37] N. Kaniwa and Y. Ojima: Experimental Designs for Evaluating Accuracy/Trueness and Precision of Analytical Procedures for Drugs and Drug Products (1). Pharm. Tech. Japan 16 (2000) 171–179.
[38] W. Horwitz and R. Albert: Performance characteristics of methods of analysis used for regulatory purposes. I. Drug dosage forms. B. Gas chromatographic methods. J. Assoc. of Anal. Chem. 67 (1984) 648–652.
[39] B. Renger: System performance and variability of chromatographic techniques used in pharmaceutical quality control. J. Chromatogr. B 745 (2000) 167–176.
[40] W. Horwitz and R. Albert: Performance of methods of analysis used for regulatory purposes. I. Drug dosage forms. D. HPLC methods. J. Assoc. Off Anal. Chem. 68 (1985) 191–198.
[41] A.G.J. Daas and J.H.M. Miller: Content limits in the European Pharmacopoeia (Part 2). Pharmeuropa 10 (1998) 137–146.
[42] W. Horwitz and R. Albert: Performance of methods of analysis used for regulatory purposes. I. Drug dosage forms. E. Miscellaneous methods. J. Assoc. Off Anal. Chem. 68 (1985) 830–838.
[43] B. Law and P. Hough: The development and evaluation of a diluter – ultraviolet spectrophotometric analysis system for the determination of drugs in solutions and suspensions. J. Pharm. Biomed. Anal. 15 (1997) 587–592.
[44] S. Ebel, A.G.J. Daas, J.-P. Fournier, J. Hoogmartens, J.H.M. Miller, J.-L. Robert, F.J. Van de Vaart, J. Vessman: Inter-laboratory trials to assess a validation procedure for volumetric titrations. Pharmeuropa 12 (2000) 18–25.
[45] M. Margosis, W. Horwitz, R. Albert: Performance of methods of analysis used for regulatory purposes. I. Drug dosage forms. F Gravimetric and titrimetric methods. J. Assoc. Off Anal. Chem. 71 (1988) 619–635.
[46] K.D. Altria, N.W. Smith, C.H. Turnbull: A review of the current status of capillary electrochromatography technology and application. Chromatographia 46 (1997) 664 – 674.
[47] K.D. Altria: Improved performance in CE using internal standards. LC-GC Europe September (2002) 588 – 594.
[48] H. Wätzig, M. Degenhardt, A. Kunkel: Strategies for CE: Method development and validation for pharmaceutical and biological applications. Electrophoresis 19 (1998) 2695–2752.
[49] G.S. Wynia, G. Windhorst, P.C. Post, F.A. Maris: Development and validation of a capillary electrophoresis method within a pharmaceutical quality control environment and comparison with high - performance liquid chromatography. Journal of Chromatography A 773 (1997) 339–350.
[50] K.D. Altria, N.G. Clayton, R.C. Harden, J.V. Makwana, M.J. Portsmouth: Inter-company cross validation exercise on capillary electrophoresis. Quantitative determination of drug counter-ion level. Chromatographia 40 (1995) 47.

[51] G. Pajchel, K. Pawlowski, S. Tyski: CE versus LC for simultaneous determination of amoxicillin/clavulanic acid and ampicillin/sulbactam in pharmaceutical formulations for injections. J. Pharm. Biomed. Anal. 29 (2002) 75–81.

[52] J.H.M. Miller, H. Binder, R. De Richter, G. Decristoforo, A. Mayrhofer, E. Roets, S. Saarlo, C. van der Vlies: Collaborative study to assess the reproducibility of a reverse-phase LC method to determine the content and to estimate the impurities of benzathine benzylpenicillin. Pharmeuropa 12 (2000) 3–7.

[53] J.H.M. Miller, T. Burat, D. Demmer, K. Fischer, M.G. Kister, A. Klemann, F. Parmentier, E. Souli, L. Thomson: Collaborative study of a liquid chromatographic method for the assay of the content and for the test for related substances of oxacillins. Part III, Dicloxacillin sodium. Pharmeuropa 9 (1997) 129–134.

[54] J.H.M. Miller: System Suitability Criteria A Case Study: The Determination of Impurities in Dicloxacillin Sodium. Pharmeuropa 12 (2000) 8–17.

[55] G.F. Pauli: qNMR –a versatile concept for the validation of natural product reference compounds. Phytochem. Anal. 12 (2001) 28–42.

[56] G. Maniara, K. Rajamoorthi, S. Rajan, G.W. Stockton: Method performance and validation for quantitative analysis by 1H and 31P NMR spectroscopy. Applications to analytical standards and agricultural chemicals. Anal. Chem. 70 (1998) 4921–4928.

[57] E.J. Wojtowicz, J.P. Hanus, R.H. Johnson, A. Lazar, R.E. Olsen, J.M. Newton, G. Petzinger, R. Ristich, M.L. Robinette: Colorimetric determination of disulfiram in tablets: collaborative study. J. Assoc. Off Anal. Chem. 64 (1981) 554–556.

[58] A. Abarca, E. Canfranc, I. Sierra, M.L. Marina: A validated flame AAS method for determining magnesium in a multivitamin pharmaceutical preparation. J. Pharm. Biomed. Anal. 25 (2001) 941–945.

[59] T. Wang, S. Walden, R. Egan: Development and validation of a general non-digestive method for the determination of palladium in bulk pharmaceutical chemicals and their synthetic intermediates by graphite furnace atomic absorption spectroscopy. J. Pharm. Biomed. Anal. 15 (1997) 593–599.

[60] B. Olsson, J.M. Aiache, H. Bull, D. Ganderton, P. Haywood, B.J. Meakin, P.J. Schorn, P. Wright: The use of inertial impactors to measure the fine particle dose generated by inhalers. Pharmeuropa 8 (1996) 291–298.

[61] M. Laasonen, T. Harmia - Pulkkinen, C. Simard, M. Rasanen, H. Vuorela: Development and validation of a near-infrared method for the quantitation of caffeine in intact single tablets. Anal. Chem. 75 (2003) 754–760.

[62] K. Molt, F. Zeyen, E. Podpetschnig – Fopp: Quantitative Nahinfrarotsprektrometrie am Beispiel der Bestimmung des Wirkstoffgehaltes von Tolbutamid-Tabletten. Pharm. Ind. 58 (1996) 847–852.

[63] R. Albert and W. Horwitz: A heuristic derivation of the Horwitz curve. Anal. Chem. 69 (1997) 789–790.

[64] J. Ermer, C. Arth, P. De Raeve, D. Dill, H.D. Friedel, H. Höwer-Fritzen, G. Kleinschmidt, G. Köller, H. Köppel, M. Kramer, M. Maegerlein, U. Schepers, H. Wätzig: Precision from drug stability studies. Collaborative investigation of long-term repeatability and reproducibility of HPLC assay procedures. J. Chromatogr. A, submitted for publication (2004).

[65] U. Schepers, J. Ermer, L. Preu, H. Wätzig: Wide concentration range investigation of recovery, precision and error structure in HPLC. J. Chromatogr. B 810 (2004), 111-118.

[66] J.B. Crowther, M.I. Jimidar, N. Niemeijer, P. Salomons: Qualification of laboratory instrumentation, validation, and transfer of analytical methods.

In: Analytical Chemistry in a GMP Environment. A practical guide. Eds. J.M. Miller and J.B. Crowther, Wiley, New York 2000, 423–458.
[67] S. Küppers, B. Renger, V.R. Meyer: Autosamplers – A major uncertainty factor in HPLC analysis precision. LC-GC Europe (2000) 114–118.
[68] J. Vessman: Selectivity or specificity? Validation of analytical methods from the perspective of an analytical chemist in the pharmaceutical industry. J. Pharm. Biomed. Anal. 14 (1996) 867–869.
[69] H. Brückner and C. Keller-Hoehl: HPLC separation of DL-amino acids derivatized with N2(5-fluoro-2,3-dinitrophenyl)-L-amino acid amides. Chromatographia 30 (1990) 621–627.
[70] J. Ermer and M. Vogel: Applications of hyphenated LC-MS techniques in pharmaceutical analysis. Biomed. Chromatogr. 14 (2000) 373–383.
[71] D. Song and J. Wang: Modified resolution factor for asymmetrical peaks in chromatographic separation. J. Pharm. Biomed. Anal. 32 (2003) 1105–1112.
[72] C.M. Riley: Statistical parameters and analytical figures of merit. In: Development and Validation of Analytical Methods. Eds. C.M.Riley and T.W.Rosanske, Elsevier, Oxford 1996, 15–71.
[73] R.E. Kaiser: Gas Chromatographie, Geest & Portig, Leipzig 1960, 33.
[74] J. Ermer: Validierung in der pharmazeutischen Analytik. In: Handbuch Validierung in der Analytik. Eds. S.Kromidas, Wiley-VCH, Weinheim 2000, 320–359.
[75] V.R. Meyer: Quantitation of chromatographic peaks in the 0.1 to 1.0% Range. Chromatographia 40 (1995) 15–22.
[76] K.D. Altria and Y.K. Dave: Peak homogeneity determination and micro – preparative fraction collection by capillary electrophoresis for pharmaceutical analysis. J. Chromatogr. 633 (1993) 221–225.
[77] J.B. Castledine and A.F. Fell: Strategies for peak - purity assessment in liquid chromatography. J. Pharm. Biomed. Anal. 11 (1993) 1–13.
[78] D.K. Bryant, M.D. Kingswood, A. Belenguer: Determination of liquid chromatographic peak purity by electrospray ionization mass spectrometry. J. Chromatogr. A 721 (1996) 41–51.
[79] D. Grion: Thermal analysis of drugs and drug products. In: Encyclopedia of Pharmaceutical Technology. Eds. J. Swarbrick and J.C. Boylan, Marcel Dekker, 2002, 2766–2793.
[80] P.H. Zoutendam, D.L. Berry, D.W. Carkuff: RP HPLC of the cardiac glycoside LNF - 209 with refractive index detection. J. Chromatogr. 631 (1993) 221–226.
[81] A. Körner: Uncovering deficiencies in mass balance using HPLC with chemiluminescence nitrogen – specific detection. LCGC North America 20 (2002) 364–373.
[82] CPMP: NIR (Note for Guidance), 2002.
[83] Analytical Methods Committee: Fitting a linear functional relationship to data with error on both variables. amc technical brief No. 10 March, 2002. http://www.rsc.org/lap/rsccom/amc/amc_techbriefs.htm
[84] FDA: Cleaning Validation, 1993.
[85] WHO: Supplementary Guideline on GMP: Validation (Draft), 2003.
[86] PIC/S: Guide to good manufacturing practice for medicinal products September 2003. www.picscheme.org
[87] J. Lambropoulos, G.A. Spanos, N.V. Lazaridis: Development and validation of an HPLC assay for fentanyl, alfentanil, and sufentanil in swab samples. J. Pharm. Biomed. Anal. 23 (2000) 421–428.
[88] Z. Katona, L. Vincze, Z. Vegh, A. Trompler, K. Ferenczi - Fodor: Cleaning validation procedure eased by using overpressured layer chromatography. J. Pharm. Biomed. Anal. 22 (2000) 349–353.

[89] R.J. Carroll and D. Ruppert: Transformation and Weighting in Regression, Chapman and Hall, New York, 1988.
[90] W. Huber. Z. Anal. Chem. 319 (1984) 379–383.
[91] K. Baumann: Regression and calibration for analytical separation techniques. Part II: Validation, weighted and robust regression. Process Control and Quality 10 (1997) 75–112.
[92] Analytical Methods Committee: Is my calibration linear? amc technical brief No. 3, December 2000, http://www.rsc.org/lap/rsccom/amc/amc-tech-briefs.htm
[93] American Society for Testing and Materials: ASTM Designation E 1303 – 89, Philadelphia 1989.
[94] S. Burke: Regression and Calibration. LC – GC Europe Online Supplement statistics and data analysis (2001) 13–18. http://www.lcgceurope.com/lcgceurope/article/articleList.jsp?categoryId=935
[95] M.M. Kiser and J.W. Dolan: Selecting the Best Curve Fit. LC - GC Europe March (2004) 138–143.
[96] DIN 38402 Teil 51: Deutsche Einheitsverfahren zur Wasser-, Abwasser- und Schlammuntersuchung, Allgemeine Angaben (Gruppe A), Kalibrierung von Analysenverfahren, Auswertung von Analysenergebnissen und lineare Kalibrierfunktionen für die Bestimmung von Verfahrenskenngrößen (A51), Beuth Verlag GmbH, Berlin, 1995.
[97] H. Mark: Application of an improved procedure for testing the linearity of analytical methods to pharmaceutical analysis. J. Pharm. Biomed. Anal. 33 (2003) 7–20.
[98] J. Vial and A. Jardy: Taking into account both preparation and injection in HPLC linearity studies. J. Chromatogr. Sci. 38 (2000) 189–194.
[99] R.D. Cook and S. Weisberg: Diagnostics for heteroscedasticity in regression. Biometrika 70 (1893) 1–10.
[100] R.J. Carroll and D. Ruppert: Transformation and weighting in regression. Chapman and Hall, New York, 1988.
[101] A.C. Atkinson: Plots, transformations, and regression. Clarendon Press, Oxford 1985.
[102] A.F. Siegel: Robust regression using repeated medians. Biometrika 69 (1982) 242–244.
[103] W. Funk, V. Dammann, C. Vonderheid, G. Oehlmann: Statistische Methoden in der Wasseranalytik. Verlag Chemie, Weinheim 1985.
[104] J.E. Knoll: Estimation of the limit of detection in chromatography. J. Chromatogr. Sci. 23 (1985) 422–425.
[105] N. Kucharczyk: Estimation of the lower limit of quantitation, a method detection performance parameter for biomedical assays, from calibration curves. J. Chromatogr. 612 (1993) 71–76.
[106] I. Kuselman and A. Shenhar: Design of experiments for the determination of the detection limit in chemical analysis. Anal. Chim. Acta 305 (1995) 301–305.
[107] M.E. Zorn, R.D. Gibbons, W.C. Sonzogni: Weighted least - squares approach to calculating limits of detection and quantification by modeling variability as a function of concentration. Anal. Chem. 69 (1997) 3069–3075.
[108] DIN 32 645: Chemische Analytik: Nachweis-, Erfassungs- und Bestimmungsgrenze, Ermittlung unter Wiederholbedingungen. Begriff, Verfahren, Auswertung. Beuth Verlag GmbH, Berlin 1994.
[109] EURACHEM Guidance Document No. WDG 2: Accreditation for chemical laboratories: Guidance on the interpretation of the EN 45000 series of standards and ISO/IEC Guide 25, 1993.

[110] J. Vial and A. Jardy: Experimental comparison of the different approaches to estimate LOD and LOQ of an HPLC method. Anal. Chem. 71 (1999) 2672–2677.
[111] C. Burgess, D.G. Jones, R.D. McDowall, LC-GC International, December 1997, 791–795.
[112] N. Dyson: Chromatographic Integration Methods, 2nd Edition, The Royal Society of Chemistry 1998, ISBN 0-85404-510-4.
[113] A. Felinger: Data analysis and signal processing in chromatography. Elsevier, 1998, ISBN 0-444-82066-3.
[114] H. Lam in C.C. Chan, H. Lam, Y.C. Lee, X.M. Zhang (Eds): Analytical Method Validation and Instrument Performance Verification. Wiley, Interscience, 2004, ISBN 0-471-25953-5.
[115] E 685–93 (Reapproved 2000), Standard Practice for Testing Fixed-wavelength Photometric Detectors Used in Liquid Chromatography, ASTM.
[116] D.L. Massart, B.G.M. Vandeginste, L.M.C. Buydens, S. De Jong, P.J. Lewi and J. Smeyers - Verbeke: Handbook of Chemometrics and Qualimetrics, Part A, Elsevier, 1997, 97.
[117] European Pharmacopoeia 4th Ed, 2.6.21 Nucleic Acid Amplification Techniques, 2002.
[118] Validation of Analytical Procedures, General Information, The Japanese Pharmacopoeia, 14th edition, JP XIV, 2001.
[119] M. Mulholland: Ruggedness testing in analytical chemistry, TRAC, 7 (1988) 383–389.
[120] Youden, E. H. Steiner: Statistical Manual of the Association of Official Analytical Chemists; The Association of Official Analytical Chemists ed., Arlington, 1975.
[121] J.A. Van Leeuwen, L.M.C. Buydens, B.G.M. Vandeginste, G. Kateman, P.J. Schoenmakers, M. Mulholland: RES, an expert system for the set-up and interpretation of a ruggedness test in HPLC method validation, Part 1, The ruggedness test in HPLC method validation, Chemometrics and Intelligent Laboratory systems 10 (1991) 337–347.
[122] Y. Vander Heyden, F. Questier and D.L. Massart: Ruggedness testing of chromatographic methods, selection of factors and levels; Journal of Pharmaceutical and Biomedical Analysis 18 (1998) 43–56.
[123] F. Lottspeich, H. Zorbas: Bioanalytik, Spektrum Akademischer Verlag GmbH, Heidelberg, Berlin 1998.
[124] L. Sachs: Angewandte Statistik, Springer Verlag, Berlin, Heidelberg, New York, 1974.
[125] W. Funk, V. Dammann, G. Donnevert: Qualitätssicherung in der Analytischen Chemie; VCH Verlagsgesellschaft mbH, Weinheim, New York, Basel, Cambridge, 1992.
[126] S. Kromidas: Validierung in der Analytik, Wiley-VCH, Weinheim, New York, Chichester, Brisbane, Singapore, Toronto, 1999.
[127] U. Lernhardt, J. Kleiner: Statistische Qualitätssicherung in der Analytik (Statistik - programm PE - SQS V2.0), Bodenseewerk Perkin-Elmer GmbH, Überlingen, 1995.
[128] L. Huber: Validation and Qualification in Analytical Laboratories, Interpharm, Buffalo Grove, IL, USA, 1998.
[129] S. Bolton: Pharmaceutical Statistics, Marcel Dekker, Inc., New York, Basel, Hong Kong, 1997.
[130] J. Ermer in Handbuch Validierung in der Analytik, Ed. S. Kromidas, Wiley-VCH, Weinheim, New York, Chichester, Brisbane, Singapore, Toronto, 2000.

[131] Handbook of Pharmaceutical; Generic Development, Sect. 13.41, Chapter 13; http://www.locumusa.com/pdf/general/article01.pdf
[132] DryLab, LC Resources Inc., Walnut Creek, CA, USA; www.lcresources.com
[133] ChromSword, Dr. Galushko Software Entwicklung, Mühltal, Germany, www.chromsword-auto.com
[134] ACD, Advanced Chemistry Development Inc., Toronto, Canada, www.acdlabs.com
[135] MODDE, Umetrics, Umeå, Sweden; www.umetrics.com, www.modde.com
[136] MINITAB, Minitab Inc., State College, PA 16801–3008, USA; www.mintab.com
[137] STATGRAPHICS, Manugistics Inc., Rockville, USA; www.manugistics.com
[138] Cs. Horvath, ed., High Performance Liquid Chromatography. Advances and Perspectives, Academic Press (New York) Vol. 1, Ch. 4, 1980.
[139] L. R. Snyder, J.L. Glajch and J. J. Kirkland: Practical HPLC Method Development (2nd ed.), Wiley–Intersciences, New York, 1997.
[140] DryLab Chromatography Reference Guide, LC Resources Inc., Walnut Creek, CA, 2000.
[141] V.R. Meyer: Pitfalls and Errors of HPLC in Pictures, Hüthig Verlag Heidelberg, 1997.
[142] I. Molnar: Robuste HPLC-Methoden Teil 1, Der Validierungsprozess bei HPLC-Analysen, LaborPraxis Juli/August (1998) 56–60.
[143] I. Molnar: Robuste HPLC-Methoden Teil 2, Definition und Überprüfung der Robustheit, LaborPraxis September (1998) 20–23.
[144] I. Molnar: Robuste HPLC-Methoden Teil 3, Robuste isokratische und robuste Gradientenmethoden, LaborPraxis November (1998) 72–79.
[145] I. Molnar: Robuste HPLC-Methoden Teil 4, Zulässige Toleranzen der eingestellten Arbeitsparameter, LaborPraxis März (1999) 54–59.
[146] J.W. Dolan, L.R. Snyder, N.M. Djordjevic, D.W. Hill and T.J. Waeghe: Reversed – phase liquid chromatographic separation of complex samples by optimizing temperature and gradient time II. Two run assay procedures, J. Chromatogr. A 857 (1999) 21–39.
[147] R.G. Wolcott, J.W. Dolan and L.R. Snyder: Computer simulation for the convenient optimization of isocratic reversed - phase liquid chromatographic separations by varying temperature and mobile phase strength, J. Chromatogr. A 869 (2000) 3–25.
[148] J.W. Dolan and L.R. Snyder: Maintaining fixed band spacing when changing column dimensions in gradient elution, J. Chromatogr. A 799 (1998) 21–34.
[149] J.W. Dolan, L.R. Snyder, N.M. Djordjevic, D.W. Hill, D.L. Saunders, L. van Heukelem and T.J. Waeghe: Simultaneous variation of temperature and gradient steepness for reversed-phase high-performance liquid chromatography method development; I. Application to 14 different samples using computer simulation, J. Chromatogr. A 803 (1998) 1–31.
[150] J.W. Dolan, L.R. Snyder, D.L. Saunders and L. van Heukelem: Simultaneous variation of temperature and gradient steepness for reversed – phase high – performance liquid chromatography method development; II. The use of further changes in conditions, J. Chromatogr. A 803 (1998) 33–50.
[151] J.W. Dolan, L.R. Snyder, R.G. Wolcott, P. Haber, T. Baczek, R. Kaliszan and L.C. Sander: Reversed-phase liquid chromatographic separation of complex samples by optimising temperature and gradient time; III. Improving the accuracy of computer simulation, J. Chromatogr. A 857 (1999) 41–68.
[152] J.L. Glajch and L.R. Snyder (Eds.): Computer-asssisted Method Development for High Performance Liquid Chromatography, Elsevier, Amsterdam, 1990.
[153] A.C. Atkinson, A.N. Donev: Optimum Experimental Designs, Oxford University Press, Oxford, 1992.

[154] R.A. Fisher: The Design of Experiments, 8th ed., Oliver & Boyd, London, 1996.
[155] E. Scheffler: Statistische Versuchsplanung und -auswertung – Eine Einführung für Praktiker, 3., neu bearbeitete und erweiterte Auflage von "Einführung in die Praxis der statistischen Versuchsplanung", Deutscher Verlag für Grundstoffindustrie, Stuttgart, 1997.
[156] E. Spenhoff: Prozesssicherheit durch statistische Versuchsplanung in Forschung, Entwicklung und Produktion, gfmt, München, 1991.
[157] A. Orth: Modellgestützte Versuchsplanung und Optimierung, Intensivkurs, Umesoft GmbH, Eschborn, 2003.
[158] A. Orth, S. Soravia: Design of Experiments, Reprint, Ullmann's Encyclopedia of Industrial Chemistry, 6th edition, Wiley-VCH, Weinheim, New York, Chichester, Brisbane, Singapore, Toronto, 2002.
[159] Y. Vander Heyden, D.L. Massart: Review of the use of robustness and ruggedness in Analytical Chemistry, in A. Smilde, J. de Boer and M. Hendriks (Eds.): Robustness of analytical methods and pharmaceutical technological products, Elsevier, Amsterdam, 1996, 79–147.
[160] A. Orth, Umesoft GmbH, Eschborn, Germany, personal communication.
[161] E. Morgan: Chemometrics, Experimental Design, Analytical Chemistry by Open Learning, Wiley, Chichester, 1991, 118–188.
[162] G. Box, W. Hunter, J. Hunter: Statistics of Experiments, an introduction to Design, Data analysis and Model Building, Wiley, New York, 1978, 306–418.
[163] R.L. Plackett, J.P. Burman: The design of optimum multifactorial experiments. Biometrika 33 (1946) 305–325.
[164] Y. Vander Heyden, K. Luypaert, C. Hartmann and D.L. Massart; J. Hoogmartens, J. De Beer: Ruggedness tests on the HPLC assay of the United States Pharmacopoeia XXII for tetracycline hydrochloride. A comparison of experimental design and statistical interpretations. Analytica Chimica Acta 312 (1995) 245–262.
[165] Y. Vander Heyden, A. Bourgeois, D.L. Massart: Influence of the sequence of experiments in a ruggedness test when drift occurs. Analytica Chimica Acta 391 (1999) 187–202.
[166] J.C. Miller and J.N. Miller: Statistics for Analytical Chemistry, Ellis Horwood, New York, 1993.
[167] W. Kleppmann: Taschenbuch Versuchsplanung, 3. überarbeitete Auflage, Carl Hanser Verlag, München, Wien, 2003.
[168] a) MODDE 7, Software for Design of Experiments and Optimisation, User Guide and Tutorial, UMETRICS AB, Umeå, Sweden, 2003. b) L. Eriksson, E. Johansson, N. Kettaneh-Wold, C. Wikström and S. Wold: Design of experiments, principles and applications. Umetrics Academy Umeå, 2000.
[169] A. Orth, D. Wenzel: Vorlesungsskript "Modellgestützte Versuchsplanung und Optimierung", Fachhochschule Frankfurt am Main, University of Applied Sciences, 2004.
[170] Draper and Smith: Applied Regression Analysis. Second Edition, Wiley, New York 1981.
[171] J. A. Nelder, R. Mead: A Simplex Method for Function Minimization. Comput. J. 7 (1965) 308–313.
[172] S. Furlanetto, S. Orlandini, P. Mura, M. Sergent, S. Pinzauti: How experimental design can improve the validation process, Studies in pharmaceutical analysis. Anal. Bioanal. Chem. 377 (2003) 937–944.
[173] Shao - Wen Sun, Hsiu-Ting Su: Validated HPLC method for determination of sennosides A and B in senna tablets. J. Pharm.. Biomed. Anal. 29 (2002) 881–994.

[174] I. Garcia, M. Cruz Ortiz, L. Sarabia, C. Vilches, E. Gredilla, Advances in methodology for the validation of methods according to the International Organization for Standardization, Application to the determination of benzoic and sorbic acids in soft drinks by high-performance liquid chromatography; J. Chromatogr. A 992 (2003) 11–27.
[175] R. Ragonese, M. Macka, J. Hughes, P. Petocz: The use of the Box–Behnken experimental design in the optimisation and robustness testing of a capillary electrophoresis method for the analysis of ethambutol hydrochloride in a pharmaceutical formulation. J. Pharm. Biomed. Anal. 27 (2002) 995–1007.
[176] P. F. Vanbel: Development of flexible and efficient strategies for optimising chromatographic separations. J. Pharm. Biomed. Anal. 21 (1999) 603–610.
[177] Q. Li, H. T. Rasmussen: Strategy for developing and optimising liquid chromatography methods in pharmaceutical development using computer – assisted screening and Plackett–Burman experimental design. J. Chromatogr. A 1016 (2003) 165–180.
[178] European Pharmacopoeia 4th Ed, 2.2.24 Infrared Absorption Spectrometry, Council of Europe, Strasbourg 2002.
[179] European Pharmacopoeia 4th Ed, 2.5.32 Microdetermination of water, Council of Europe, Strasbourg 2002.
[180] 2.4.8 Heavy Metals, Pharmeuropa 15 (2003) 359–36.
[181] 2.2.23 Atomic Absorption Spectrophotometry, Pharmeuropa 15 (2003).447–448.
[182] European Pharmacopoeia 4th Ed, 4.2.2 Volumetric Solutions. Council of Europe, Strasbourg 2002.
[183] Y. Zhu, J. Augustinjs, E. Roets, J. Hoogmartens: Robustness test for the liquid chromatographic method for the assay of amoxicillin. Pharmeuropa 9 (1997) 323–327.
[184] Y. Zhu, J. Augustinjs, E. Roets, J. Hoogmartens: Robustness test for the liquid chromatographic method for the assay of ampicillin. Pharmeuropa 9 (1997) 742–74.
[185] U. Rose: *In situ* degradation: a new concept for system suitability in monographs of the European Pharmacopoeia. J. Pharm. Bio. Anal. 18 (1998) 1–14.
[186] W.B. Furman, J.G. Dorsey, L.R. Snyder: System suitability tests in regulatory liquid and gas chromatographic methods: Adjustments versus modifications. Pharm. Tech., June, (1998) 39–42.
[187] R. Cox, G. Menon: System Suitability. Pharmeuropa 10 (1998) 136.
[188] Reagent 411, European Pharmacopoeia 4th Ed., Council of Europe, Strasbourg (2002).
[189] http//wwwpheurorg
[190] K. Kimata, K. Lwaguchi, S.O. Nishi, K. Jinno, R. Eksteen, K. Hosoya, M. Araki, N. Tanaka: Chromatographic characterisation of Silica C-18 packing materials. Correlation between a preparation method and retention behaviour of stationary phases. J. Chromatogr. Sci. 27 (1989) 721–728.
[191] B. Walczak, L. Morin-Allory, M. Lafosse, M. Dreux, R. Chrétien: Factor analysis and experiment. J. Chromatogr. 395 (1987) 183–202.
[192] F. Delaney, A.N. Papas, M.J. Walters: Chemometric classification of reversed-phase high performance liquid chromatographic columns. J. Chromatogr. 410 (1987) 31–41.
[193] T. Welsch et al.: Chromatographia 19 (1984) 457.
[194] C. Chatsfield, A.J. Collins: Introduction to Multivariate Analysis. Chapman & Hall, London (1984).
[195] E. Cruz, M.R. Euerby, C.M. Johnson, C.A. Hackett: Chromatographic classification of commercially available reverse-phase columns. Chromatographia 44 (1997) 151–161.

[196] S.D. Rogers, J.G. Dorsey: Chromatographic solvent activity test procedures: a quest for a universal limit. J. Chromatogr. A 892 (2000) 57–65.

[197] C. Stella, S. Rudaz, J.-L. Veuthey, A. Tehapla: Silica and other materials as supports in liquid chromatography. Chromatographic tests and their importance for evaluating these supports; Parts I & II. Chromatographia 53 (2001) S113–S149.

[198] D. Visky, Y. Varder Heyden, T. Iványi, P. Baten, J. De Beer, B. Noszal, E. Roets, D.L. Massart, J. Hoogmartens: Characterisation of reversed-phase liquid chromatographic columns by chromatographic tests. Pharmeuropa (2002) 14288–297.

[199] M.E. Euerby, P. Petersen: Chromatographic classification and comparison of commercially available reverse-phase liquid chromatographic columns using principal component analysis. J. Chromatogr A 994 (2003) 13–36.

[200] D. Visky, Y. Varder Heyden, T. Iványi, P. Baten, J. De Beer, Z. Kovacs, B. Noszal, E. Roets, D.L. Massart, J. Hoogmartens: Characterisation of reverse-phase liquid chromatographic columns by chromatographic tests. Evaluation of 36 test parameters: repeatability, reproducibility and correlation. J. Chromatogr. A 977 (2002) 39–58.

[201] T. Iványi, Y. Varder Heyden, D. Visky, P. Baten, J. De Beer, J. Lázar, E. Roets, D.L. Massart, J. Hoogmartens: Minimal number of chromatographic test parameters for the characterisation of reversed-phase liquid chromatographic stationary phases. J. Chromatogr A 954 (2002) 99–114.

[202] D. Visky, Y. Varder Heyden, T. Iványi, P. Baten, J. De Beer, Z. Kovacs, B. Noszal, E. Roets, D.L. Massart, J. Hoogmartens: Characterisation of reversed-phase liquid chromatographic columns by chromatographic tests? Column classification by a minimal number of column test parameters. J. Chromatogr. A 1012 (2003) 11–29.

[203] P. Dehouck, D. Visky, Y. Varder Heyden, E. Adams, Z. Kovacs, B. Noszal, D.L. Massart, J. Hoogmartens: Characterisation of reversed-phase liquid chromatographic columns by chromatographic tests. Comparing column classification based on chromatographic parameters and column performance for the separation of acetylsalicylic acid and related compounds. J. Chromatogr. A 1025 (2004) 189–200.

[204] N.S. Wilson, M.D. Nelson, J.W. Dolan, L.R. Snyder, R.G. Wolcott, P.W. Barr: Column selectivity in reversed-phase liquid chromatography I. A general quantitative relationship. J. Chromatogr. A 961 (2002) 171–193.

[205] N.S. Wilson, M.D. Nelson, J.W. Dolan, L.R. Snyder, P.W. Barr: Column selectivity in reversed-phase liquid chromatography II. Effect of a change in conditions. J. Chromatogr. A 961 (2002) 195–215.

[206] N.S. Wilson, J.W. Dolan, L.R. Snyder, P.W. Barr, L.C. Sander: Column selectivity in reversed-phase liquid chromatography III. The physico-chemical basis of selectivity. J. Chromatogr. A 961 (2002) 217–236.

[207] J.J. Gilroy, J.W. Dolan, L.R. Snyder: Column selectivity in reversed-phase liquid chromatography IV. Type B alkylsilica columns. J. Chromatogr A 1000 (2003) 757–778.

[208] Technical Guide for Elaboration of Monograph 3rd Ed., Pharmeuropa Special Issue (1999).

[209] UIPAC Recommendations 2000 „The hold-up volume concept in column chromatography".

[210] G.E. Berendsen, P.J. Schoenmakers, L. de Galan: On the determination of the hold-up time in reversed phase liquid chromatography. J. Liq. Chrom. 3 (1980) 1669–1686.

[211] P.L. Zhu: Adsorption isotherms of irganic modifiers and the determination of the dead volume in RPLC, Chromatographia 20 (1985) 425–433.
[212] R.A. Djerki, R,J, Lamb: Solute retention in column chromatography. IX Comparison of methods of determination of the void volume in liquid liquid chromatography. J. Liq. Chromatogr. 10 (1987) 1749–1767.
[213] Y.V. Kazakevich, H.M. McNair: Thermodynamic definition of HPLC dead volume. J. Chromatogra Sci. 31 (1993) 317–322.
[214] R. Walter, W. Arz: Liquid chromatography, the problem of reproducing chromatographic separations with gradient systems (dwell volume and dwell time of HPLC systems). Pharmeuropa 9 (1997) 558–559.
[215] A.J. Grizodoub, M.G. Levin, D.A. Leontiev, V.P. Georgievski: Standardisation of the chromatographic analysis of drugs. I Metrological aspects of using HPLC. Pharmacom 7 (1995) 8–19.
[216] European Pharmacopoeia 4th Ed, 2.4.24 Identification and Control of Residual Solvents. Council of Europe, Strasbourg 2002.
[217] European Pharmacopoeia 4th Ed, 2.4.25 Ethylene oxide and Dioxan. Council of Europe, Strasbourg 2002.
[218] A. Marin, E. Garcia, A. Garcia, C. Barbas: Validation of a HPLC quantification of acetaminophen, phenylephrine and chlorpheniramine in pharmaceutical formulations: capsules and sachets. J. Pharm. Biomed. Anal. 29 (2002) 701–714.
[219] M.A. Raggi, G. Casamenti, R. Mandrioli, C. Sabbioni, V. Volterra: A rapid LC method for the identification and determination of CNS drugs in pharmaceutical formulations. J. Pharm. Biomed. Anal. 23 (2000) 161–167.
[220] A. Zotou and N. Miltiadou: Sensitive LC determination of ciprofloxacin in pharmaceutical preparations and biological fluids with fluorescence detection. J. Pharm. Biomed. Anal. 28 (2002) 559–568.
[221] J.H.M. Miller, T. Burat, D. Demmer, K. Fischer, M.G. Kister, A. Klemann, F. Parmentier, E. Souli, L. Thomson: Collaborative study of a liquid chromatographic method for the assay of the content and for the test for related substances of oxacillins. Part III, Dicloxacillin sodium. Pharmeuropa 9 (1997) 129–134.
[222] T. Radhakrishna and R.G. Om: Simultaneous determination of fexofenadine and its related compounds by HPLC. J. Pharm. Biomed. Anal. 29 (2002) 681–690.
[223] I. Ismail Salem, M.C. Bedmar, M.M. Medina, A. Cerezo: Insulin evaluation in pharmaceuticals: Variables in RP-HPLC and method validation. J. Liq. Chrom. 16 (1993) 1183–1194.
[224] R.R. Kenney, R.J. Forsyth, H. Jahansouz: Solid - phase extraction and liquid chromatographic quantitation of the antiarrhythmic drug L-768673 in a microemulsion formulation. J. Pharm. Biomed. Anal. 17 (1998) 679–687.
[225] A. Rafiq Khan, M. Jamil Akhtar, R. Mahmood, S. Muied Ahmed, S. Malook, M. Iqbal: LC assay method for oxfendazole and oxyclozanide in pharmaceutical preparation. J. Pharm. Biomed. Anal. 22 (2000) 111–114.
[226] R.W. Sparidans, J. Den Hartigh, W.M. Ramp-Koopmanschap, R.H. Langebroek, P. Vermeij: The determination of pamidronate in pharmaceutical preparations by ion-pair liquid chromatography after derivatization with phenylisothiocyanate. J. Pharm. Biomed. Anal. 16 (1997) 491–497.
[227] L. Suntornsuk: Direct determination of s-carboxymethyl-l-cysteine in syrups by reversed-phase high-performance liquid chromatography. J. Pharm. Biomed. Anal. 25 (2001) 165–170.
[228] J. Vial, I. Menier, A. Jardy, A. Jardy, P. Amger, A. Brun, L. Burbaud: How to better define the characteristics of dispersion of results in liquid chromatographic analyses trough an interlaboratory study. Example

of collaborative studies on ketoprofen and spiramycin. J. Chromatogr. B 708 (1998) 131–143.
[229] A.I. Gasco-Lopez, R. Izquirdo-Hornillos, A. Jiminez: Development and validation of an hplc method for the determination of cold relief ingredients in chewing gum. J. Chromatogr. A 775 (1997) 179–185.
[230] I.P. Nnane, L.A. Damani, A.J. Hutt: Development and validation of stability indicating HPLC assays for ketotifen in aqueous and silicon oil formulations. Chromatographia 48 (1998) 797–802.
[231] G. Parhizari, G. Delker, R.B. Miller, C. Chen: A stability-indicating HPLC method for the determination of benzalkonium chloride in 0.5% tramadol ophthalmic solution. Chromatographia 40 (1995) 155–158.
[232] L.K. Shao and D.C. Locke: Determination of paclitaxel and related taxanes in bulk drug and injectable dosage forms by RP - LC. Anal. Chem. 69 (1997) 2008–2016.
[233] R.N. Saha, C. Sajeev, P.R. Jadhav, S.P. Patil, N. Srinivasan: Determination of celecoxib in pharmaceutical formulations using UV spectrophotometry and liquid chromatography. J. Pharm. Biomed. Anal. 28 (2002) 741–751.
[234] M.A. Raggi, F. Bugamelli, V. Pucci: Determination of melatonin in galenic preparations by LC and voltammetry. J. Pharm. Biomed. Anal. 29 (2002) 283–289.
[235] M. Candela, A. Ruiz, F.J. Feo: Validation of an analysis method for 4-amino-3-hydroxybutyric acid by reversed-phase liquid chromatography. J. Chromatogr. A 890 (2000) 273–280.
[236] R.D. Marini, A. Pantella, M.A. Bimazubute, P. Chiap, Ph. Hubert, J. Crommen: Optimisation and validation of a generic method for the LC assay of six corticosteroids and salicylic acid in dermopharmaceutical forms. Chromatographia 55 (2002) 263 – 269.
[237] T. Radhakrishna, D.S. Rao, K. Vyas, G.O. Reddy: A validated method for the determination and purity evaluation of benazepril hydrochloride in bulk and in pharmaceutical dosage forms by liquid chromatography. J. Pharm. Biomed. Anal. 22 (2000) 941–650.
[238] A.F.M. El Walily: Analysis of nifedipine-acebutolol hydrochloride binary combination in tablets using UV-derivative spectroscopy, capillary gas chromatography and high performance liquid chromatography. J. Pharm. Biomed. Anal. 16 (1997) 21–30.
[239] C.S. Eskilsson, E. Bjorklund, L. Mathiasson, L. Karlsson, A. Torstensson: Microwave-assisted extraction of felodipine tablets. J. Chromatogr. A 840 (1999) 59–70.
[240] M.L. Qi, P. Wang, L.J. Wang, R.N. Fu: LC method for the determination of oxcarbazepine in pharmaceutical preparations. J. Pharm. Biomed. Anal. 31 (2003) 57–62.
[241] L.M. Morsch, C.F. Bittencourt, M.J. Souza, J. Milano: LC method for the analysis of cefetamet pivoxil hydrochloride in drug substance and powder for oral suspension. J. Pharm. Biomed. Anal. 30 (2002) 643–649.
[242] J. Milano, L.M. Morsch, S.G. Cardoso: LC method for the analysis of Oxiconazole in pharmaceutical formulations. J. Pharm. Biomed. Anal. 30 (2002) 175–180.
[243] T. Radhakrishna, J. Satyanarayana, A. Satyanarayana: LC determination of rosiglitazone in bulk and pharmaceutical formulation. J. Pharm. Biomed. Anal. 29 (2002) 873–880.
[244] A.K. Dash and A. Sawhney: A simple LC method with UV detection for the analysis of creatine and creatinine and its application to several creatine formulations. J. Pharm. Biomed. Anal. 29 (2002) 939–945.

[245] A.R. Khan, K.M. Khan, S. Perveen, N. Butt: Determination of nicotinamide and 4-aminobenzoic acid in pharmaceutical preparation by LC. J. Pharm. Biomed. Anal. 29 (2002) 723–727.
[246] F.J. Ruperez, H. Fernandez, C. Barbas: LC determination of loratadine and related impurities. J. Pharm. Biomed. Anal. 29 (2002) 35–41.
[247] S.N. Makhija and P.R. Vavia: Stability indicating LC method for the estimation of venlafaxine in pharmaceutical formulations. J. Pharm. Biomed. Anal. 28 (2002) 1055–1059.
[248] M.J. Souza, C.F. Bittencourt, L.M. Morsch: LC determination of enrofloxacin. J. Pharm. Biomed. Anal. 28 (2002) 1195–1199.
[249] R.M. Cardoza and P.D. Amin: A stability indicating LC method for felodipine. J. Pharm. Biomed. Anal. 27 (2002) 711–718.
[250] I. Pineros, P. Ballesteros, J.L. Lastres: Extraction and LC determination of lysine clonixinate salt in water/oil microemulsions. J. Pharm. Biomed. Anal. 27 (2002) 747–754.
[251] D. Castro, M.A. Moreno, J.L. Lastres: First-derivative spectrophotometric and LC determination of nifedipine in Brij(R) 96 based oil/water/oil multiple microemulsions on stability studies. J. Pharm. Biomed. Anal. 26 (2001) 563–572.
[252] T. Radhakrishna, D.S. Rao, G.O. Reddy: LC determination of rofecoxib in bulk and pharmaceutical formulations. J. Pharm. Biomed. Anal. 26 (2001) 617–628.
[253] D. Sreenivas Rao, S. Geetha, M.K. Srinivasu, G. Om Reddy: LC determination and purity evaluation of nefazodone HCl in bulk drug and pharmaceutical formulations. J. Pharm. Biomed. Anal. 26 (2001) 629–636.
[254] N. Erk: Comparison of spectrophotometric and an LC method for the determination perindopril and indapamide in pharmaceutical formulations. J. Pharm. Biomed. Anal. 26 (2001) 43–52.
[255] T. Ceyhan, M. Kartal, M.L. Altun, F. Tulemis, S. Cevheroglu: LC determination of atropine sulfate and scopolamine hydrobromide in pharmaceuticals. J. Pharm. Biomed. Anal. 25 (2001) 399–406.
[256] T. Radhakrishna, C. Lakshmi Narayana, D. Sreenivas Rao, K. Vyas, G. Om Reddy: LC method for the determination of assay and purity of sibutramine hydrochloride and its enantiomers by chiral chromatography. J. Pharm. Biomed. Anal. 22 (2000) 627–639.
[257] M.A. Moreno, M.P. Ballesteros, P. Frutos, J.L. Lastres, D. Castro: Comparison of UV spectrophotometric and LC methods for the determination of nortriptyline hydrochloride in polysorbate 80 based oil/water (o/w) microemulsions. J. Pharm. Biomed. Anal. 22 (2000) 287–294.
[258] E. Vega, V. Dabbene, M. Nassetta, N. Sola: Validation of a reversed-phase LC method for quantitative analysis of intravenous admixtures of ciprofloxacin and metronidazole. J. Pharm. Biomed. Anal. 21 (1999) 1003–1009.
[259] D. Castro, M.A. Moreno, S. Torrado, J.L. Lastres: Comparison of derivative spectrophotometric and liquid chromatographic methods for the determination of omeprazole in aqueous solutions during stability studies. J. Pharm. Biomed. Anal. 21 (1999) 291–298.
[260] A.M. Di Pietra, R. Gatti, V. Andrisano, V. Cavrini: Application of high – performance liquid chromatography with diode - array detection and on-line post-column photochemical derivatization to the determination of analgesics. J. Chromatogr. A 729 (1996) 355–361.
[261] X.-Z. Qin, J. DeMarco, D.P. Ip: Simultaneous determination of enalapril, felodipine and their degradation products in the dosage formulation by RP-HPLC using a Spherisorb C8 column. J. Chromatogr. A 707 (2003) 245–254.

Part II:
Life-cycle Approach to Analytical Validation

Method Validation in Pharmaceutical Analysis. A Guide to Best Practice. Joachim Ermer, John H. McB. Miller (Eds.)

ISBN: 3-527-31255-2

4
Qualification of Analytical Equipment

David Rudd

4.1
Introduction

Within any overall quality system where product suitability is ultimately determined by the output from analytical instrumentation, it is important to be able to demonstrate that such equipment is fit for its intended purpose and that it is calibrated and maintained in an appropriate state of readiness.

The verification of performance or qualification of analytical equipment may be achieved in a variety of different ways, depending on the type of equipment and its intended application, but, in general, there are a series of steps which need to be considered in order to ensure that such equipment is truly 'fit for purpose'. This chapter discusses the overall objectives of equipment qualification, recognising that different levels may apply at different stages of utilisation or application, and provides a systematic approach which can be adopted to satisfy current regulatory and laboratory accreditation requirements. Its contents are heavily derived from guidance developed by the Laboratory of the Government Chemist (http://www.lgc.co.uk) with assistance from the Eurachem-UK Instrumentation Working Group and which has been previously published in Accreditation and Quality Assurance (1996) 1: 265–274 (copyright Springer) by Peter Bedson and Mike Sargent [1] under the title 'The development and application of guidance on equipment qualification of analytical instrumentation'.

It must be recognised that, although a common philosophy for equipment qualification may be applied across different analytical techniques, inevitably there will be different specific approaches and requirements from one technique to another. For example, the qualification of UV-visible spectrometers will generally necessitate confirmation of wavelength accuracy using traceable standards, whereas calibration of a pH meter will depend on the use of certified buffer solutions. Both are concerned with the confirmation of accuracy, but the specific approach adopted, and the acceptance criteria used, are quite different.

Finally, even within a given analytical technique, the required level of equipment qualification will depend on the intended application. A liquid chromatography system used simply for product identification, based on co-incidence of retention time with a certified reference standard, may require substantially less qualification than

Method Validation in Pharmaceutical Analysis. A Guide to Best Practice. Joachim Ermer, John H. McB. Miller (Eds.)

ISBN: 3-527-31255-2

one used for accurate quantitative assessment of potentially toxic drug-related impurities, for example.

These last points lead to the conclusion that the qualification process itself must also be 'fit for purpose'. All guidance should be considered as a framework within which equipment qualification may be achieved in a systematic and justifiable way, rather than as a prescriptive set of procedures and practices, which must be adhered to under all circumstances. As always with any validation programme, it is the responsibility of the user to establish the level of qualification which will demonstrate the fitness for purpose of the particular piece of equipment for the intended application.

4.2 Terminology

The following list of definitions is provided in order to clarify a number of the terms used in this chapter. It should be noted that no universal set of such definitions currently seems to exist, so these are provided simply for clarity in the present work. Note, however, that they are consistent with those provided by Bedson and Sargent [1]:

Instrument: all types of measuring equipment ranging from simple stand-alone instruments through to complex multi-component instrument systems.

User: the organisation purchasing the instrument including its management and staff.

Supplier: the instrument manufacturer, vendor, lessor or approved agent.

Operational specification: the key performance characteristics of the instrument and ranges over which the instrument is required to operate and perform consistently, as agreed between the user and supplier.

Functional specification: the overall requirements of the instrument including the operational specification (see above) and other critical factors relating to its use (for example, level of training/expertise required by operators).

Equipment Qualification (EQ): the overall process of ensuring that an instrument is appropriate for its intended use and that it performs according to specifications agreed by the user and supplier. EQ is often broken down into Design, Installation, Operation and Performance Qualification.

Design Qualification (DQ): this covers all procedures prior to the installation of the system in the selected environment. DQ defines the functional and operational specifications of the instrument and details the conscious decisions in the selection of the supplier.

Installation Qualification (IQ): this covers all procedures relating to the installation of the instrument in the selected environment. IQ establishes that the instrument is received as designed and specified, that it is properly installed in the selected environment and that this environment is suitable for the operation and use of the instrument.

Operational Qualification (OQ): the process of demonstrating that an instrument will function according to its operational specification in the selected environment.

Performance Qualification (PQ): this is defined as the process of demonstrating that an instrument consistently performs according to a specification appropriate for its routine use.

Validation: the process of evaluating the performance of a specific measuring procedure and checking that the performance meets certain pre-set criteria. Validation establishes and provides documented evidence that the measuring procedure is fit for a particular purpose.

System Suitability Checking (SSC): a series of tests to check the performance of a measurement process. SSC may form part of the process of validation when applied to a particular measuring procedure. SSC establishes that the operational conditions required for a specific measurement process are being achieved.

Calibration: the set of operations which establish, under specified conditions, the relationship between values indicated by a measuring instrument or process and the corresponding known values of the measurand.

Traceability: the property of a result of a measurement whereby it can be related to appropriate standards, generally national or international standards, through an unbroken chain of comparisons.

4.3 An Overview of the Equipment Qualification Process

In keeping with the general validation principle of 'fitness for purpose', the overall process of Equipment Qualification (EQ) may be seen as the demonstration and documentation that an instrument is performing, and will continue to perform, in accordance with a pre-defined operational specification. In turn, this operational specification must ensure a level of performance which is appropriate for the intended use of the instrument.

Generally a four-part model for the EQ process is recognised, reflecting the various stages of the overall qualification procedure. These stages are usually referred to as: Design Qualification (DQ), Installation Qualification (IQ), Operational Qualification (OQ) and Performance Qualification (PQ) and are defined as shown in Figure 4-1. Each of these stages of EQ is described more fully later in this chapter.

DQ is the 'planning' part of the EQ process and is most often undertaken as part of the purchasing of a new instrument, although it may be appropriate to repeat aspects of DQ following a major change to the instrument or its use. While the responsibility for the qualification of the actual instrument design resides with the manufacturer of the instrument, the user of the instrument also has an important role in DQ by ensuring adoption of a user requirement specification (URS), which ensures suitability for the intended use.

IQ, OQ and PQ are the 'implementation' stages of the EQ process and provide an assurance that the instrument is installed properly, that it operates correctly and that

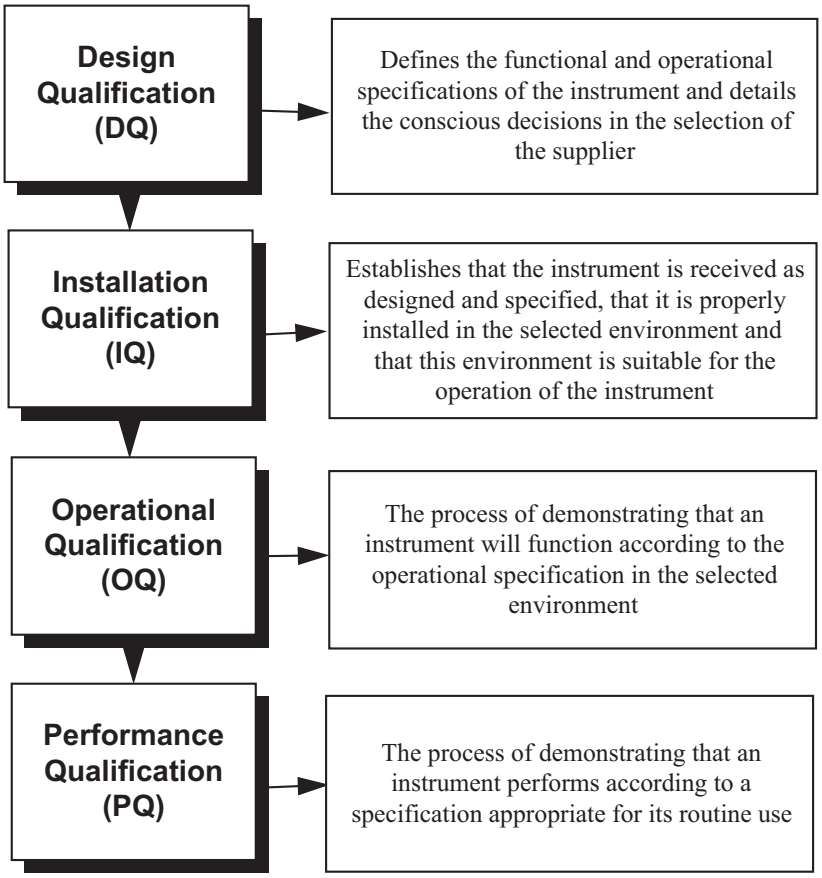


Figure 4-1 The equipment qualification process.

its ongoing performance remains within the limits required for its intended application. IQ covers the installation of the instrument up to and including its response to the initial application of power. OQ should be carried out after the initial installation of the instrument (IQ) and repeated following a major event (for example, re-location or maintenance) or periodically at defined intervals (for example, annually).

PQ is undertaken regularly during the routine use of the instrument. The role of PQ is to provide continued evidence that, even though the performance of the instrument may change due to factors such as wear or contamination, its performance remains within the limits required for its intended application. As such, much of the evidence needed for PQ is available from routine usage (for example, method validation, system suitability checking (SSC), routine calibration and analytical quality control).

Each stage of the qualification process involves the same general approach – that is, the preparation of a qualification plan defining the scope of qualification (for example, the tests to be performed and the acceptance criteria to be used), the execu-

tion of the plan (during which the results of the tests are recorded as the tests are performed) and the production of a report (and, if required, a certificate) in which the results of EQ are documented.

While this chapter describes a general approach to the EQ process, more specific guidance relating to individual analytical techniques is also available. For example, high performance liquid chromatography (HPLC) has been covered by Bedson and Rudd [2], while Holcombe and Boardman [3] provides information on the qualification of UV-visible spectrophotometers.

4.4 Documentation of the EQ Process

EQ must be documented. EQ documentation can be prepared and provided by the user, the supplier or both. Where it is provided by the supplier (for example, in a qualification protocol), it remains the responsibility of the user and should be written in such a way that it can be readily followed and understood by the user.

Documentation covering EQ should include the following:

a) The instrument and all modules and accessories must be uniquely identified, particularly Reports and Certificates, including:
   The supplier's name, instrument name, model and serial number;
   Any identifying number allocated by the user;
   The version and date of issue of any computer hardware, firmware and software
   It may also be useful to include a brief description of the instrument and its role in the measurement process.
b) A clear statement of the intervals at which aspects of EQ and/or specific checks and tests should be performed, and the responsibility level of the operator required to perform the tests.
c) Details of each check and test to be performed, the specification and acceptance criteria to be used. This information should be concise enough to allow the operator to make an unambiguous judgement on the result of the test.
d) Sufficient information on the procedures and materials that are required to perform each check and test. This should also advise on the need to achieve traceability to national or international standards and how this can be achieved.
e) Where qualification of one part of the instrument is dependent on the correct functioning of another part, any relevant assumptions made must be recorded.
f) The date on which qualification was performed and the result of qualification and each check or test.
g) The reason for performing qualification (for example, following installation of a new instrument, following routine service or following instrument malfunction).
h) Clear information on the action to be taken in the event of test or qualification failure.

i) The circumstances which may or will necessitate re-qualification of the instrument (for example, following service or re-calibration).
j) The name(s) and signature(s) of the person(s) who actually performed qualification and/or each individual check and test. In addition, the documentation should contain the name and signature of the user who is authorising completion of qualification.

It is strongly recommended that log-books are kept for all instruments. Many Quality Standards place a heavy emphasis on keeping records of instrument history. Maintaining an up-to-date log-book of the overall history of an instrument provides a convenient mechanism for recording information and can provide the basis for satisfying the requirements of many laboratory accreditation systems.

Instrument log-books should identify the individual modules and accessories which constitute the instrument and may be used to record the overall history of the instrument (for example, the date of purchase, the initial qualification and entry into service; the dates when subsequent maintenance, calibration and qualification have been performed and when these are next due). In some circumstances, it may be appropriate for all relevant information to be recorded in, or appended to, the instrument log-book (for example, operating instructions and Standard Operating Procedures (SOPs), maintenance and calibration records, and qualification and qualification protocols and reports). In others, it may be more appropriate to use the log-book as a summary record of key information which references where more detailed procedures, reports and certificates can be accessed.

Following qualification, the instrument log-book must be updated with the results of qualification. The instrument itself should also be labelled to provide a clear indication of when the next qualification, calibration or performance test is due.

4.5 Phases of Equipment Qualification

4.5.1 Design Qualification (DQ)

Design Qualification is concerned with what the instrument is required to do and links directly to fitness for purpose. DQ provides an opportunity for the user to demonstrate that the instrument's fitness for purpose has been considered at an early stage and built into the procurement process.

DQ should, where possible, establish the intended or likely use of the instrument and should define appropriate operational and functional specifications. This may be a compromise between the ideal and the practicalities of what is actually available. While it is the responsibility of the user to ensure that specifications exist, and that these specifications are appropriate for the intended application, they may be prepared by the user, the supplier(s) or by discussion between the two.

The operational specification should define the key performance characteristics of the instrument and the ranges over which the instrument is required to operate and to perform consistently.

The functional specification should consider the overall requirements of the instrument, including the operational specification (see above) and other critical factors relating to its use, for example:

a) the overall business requirement;
b) documentation relating to the use of the instrument (for example, clear, easy-to-use operating manuals, identified by version and date; protocols for IQ, OQ and PQ; model SOPs, etc.);
c) the level of skill required to operate the instrument and details of any training necessary and courses provided by the supplier;
d) sample throughput, presentation and introduction needs;
e) data acquisition, processing and presentation needs;
f) requirements for, and expected consumption of, services, utilities, and consumables (for example, electricity, special gases);
g) environmental conditions within which, or the range over which, the instrument must work;
h) suggested contents of, intervals between and procedures for maintenance and calibration of the instrument, including the cost and availability of any service contracts;
i) the period for which support (qualification, maintenance, parts, etc.) for the instrument can be guaranteed;
j) information on health and safety and environmental issues and/or requirements.

In undertaking DQ, information and knowledge of existing equipment should be taken into account. If an instrument is mature in design and has a proven track record, this may provide a basic level of confidence and evidence about its suitability for use. For new techniques or instruments, DQ may require more extensive effort.

The selection of the supplier and instrument is entirely at the discretion of the user. However, in selecting the supplier and instrument, the user should bear in mind that regulatory agencies are likely to require evidence of the use of rigorous design and specification methods; fully-documented quality control and quality assurance procedures; the use, at all times, of suitably qualified and experienced personnel; comprehensive, planned testing of the system; and the application of stringent change control, error reporting and corrective procedures. A suitable questionnaire, third-party audit or independent certification of the supplier to an approved quality scheme may provide the user with evidence that regulatory requirements have been met. Where such evidence is not available, it is the responsibility of the user to carry out more extensive qualification in order to provide the necessary assurance of the instrument's fitness for use.

Where instruments are employed to make measurements supporting regulatory studies, the user may also need to seek confirmation that the manufacturer is pre-

pared, if required, to allow regulatory authorities access to detailed information and records relating to the instrument's manufacture and development (for example: source codes; instrument development records and procedures; calibration and qualification documentation; batch test records and reports; hardware and software qualification documentation and credentials of staff involved with the development of the instrument).

4.5.2 Installation Qualification (IQ)

It is often questionable as to what EQ aspects should be included under Installation Qualification and what should be included under Operational Qualification. Indeed, the judgement may be different for different manufacturers and/or different instruments. As an arbitrary, but pragmatic approach, it is recommended that IQ should generally cover the installation of the instrument up to and including its response to the initial application of power.

IQ involves formal checks to confirm that the instrument, its modules and accessories have been supplied as ordered (according to specifications agreed between the user and supplier), and that the instrument is properly installed in the selected environment. IQ must be formally documented (see previous Documentation section) and should confirm the following:

a) that the instrument (including all modules and accessories) has been delivered as ordered (delivery note, purchase order, agreed specifications) and that the instrument has been checked and verified as undamaged;
b) that all required documentation has been supplied and is of correct issue (for example, operating manuals – which should also include their issue number and date of issue, the supplier's specification, and details of all services and utilities required to operate the instrument);
c) that recommended service, maintenance, calibration and qualification intervals and schedules have been provided. Where maintenance can be carried out by the user, appropriate methods and instructions should be referenced along with contact points for service and spare parts;
d) that any required computer hardware, firmware and software has been supplied and is of correct issue;
e) that information on consumables required during the normal operation of the instrument system, and during the start-up or shut-down procedures, has been provided;
f) that the selected environment for the instrument system is suitable, with adequate room for installation, operation and servicing, and appropriate services and utilities (electricity, special gases, etc.) have been provided. (Note: significant time and effort can be saved if these basic requirements are checked prior to formal IQ of the instrument);
g) that health and safety and environmental information relating to the operation of the instrument has been provided. It is the responsibility of the sup-

plier to provide appropriate safety information, on which the user must act, and to document the acceptance of this guidance;

h) that the response of the instrument to the initial application of power is as expected or that any deviations are recorded. If the system is designed to perform any automatic diagnostic or start-up procedures, the response to these should also be observed and documented.

IQ may be carried out either by the supplier and/or the user. However, it should be noted that, in some cases, the complexity of the instrument alone may preclude the user performing IQ and, in others, the unpacking of the equipment by the user may invalidate the warranty.

IQ must be undertaken by a competent individual and in accordance with the supplier's instructions and procedures. The success or failure of each of the IQ checks performed should be formally recorded and, where these have been carried out by the supplier, the results of these tests must be communicated to the user.

4.5.3
Operational Qualification (OQ)

The purpose of Operational Qualification (OQ) is to demonstrate and provide documented evidence that the instrument will perform according to the operational specification in the selected environment.

OQ normally takes place after the IQ of a new instrument or after a significant change to the instrument or a component, such as repair or service.

OQ may be carried out either by the supplier or the user, but must remain under the control of the user. However, for complex instruments, it may only be possible for the supplier to undertake OQ.

OQ should be carried out in accordance with the supplier's instructions and procedures, using suitable materials and protocols, and should satisfy the general requirements set out in the previous section on Equipment Qualification. It is not possible at this stage to give further general guidance on OQ requirements as the checks and tests necessary to demonstrate an instrument's compliance with its operational specification are specific and vary depending on the type of instrument undergoing qualification. However, OQ must be formally documented in accordance with the general requirements set out in the previous section on Documentation.

4.5.4
Performance Qualification (PQ)

The purpose of PQ is to ensure that the instrument functions correctly and to a specification appropriate for its routine use. This specification may be the original operational specification or one more appropriate for its current use. PQ provides the continuing evidence of control and acceptable performance of the instrument during its routine use.

The frequency of, and need for, PQ should be specified in in-house operating manuals or in a Standard Operating Procedure (SOP) and should be based on need, type and previous performance of the instrument, including the time that the instrument calibration has been found, in practice, to remain within acceptable limits.

Where possible, all operational checks and tests should be performed using parameters as close as possible to those used during normal routine operation of the instrument. For most analytical instruments, there will be an indeterminate area between the optimum and unacceptable levels of performance. Wherever this is the case, the user must identify a threshold, below which the instrument's performance is deemed to be unacceptable and where it should not be used until its performance is improved.

Aspects of performance qualification are often built into analytical methods or procedures. This approach is often called System Suitability Checking (SSC) which demonstrates that the performance of the measuring procedure (including instrumental operating conditions) is appropriate for a particular application. SSC should be used before and during analysis to provide evidence of satisfactory operation or to highlight when performance is no longer acceptable.

When a complete measuring system is provided by the supplier, PQ can be performed by the supplier, but must remain under the control of the user. In some circumstances, PQ may also involve repeating many of the checks and tests carried out during OQ and, therefore, these can also be performed by the supplier. However, wherever PQ is performed by the supplier, it is likely that the user will also have to undertake more frequent checks and tests to confirm the continued satisfactory performance of the instrument during routine use.

PQ should be carried out in accordance with the general requirements set out in the previous section on Equipment Qualification. It is not possible at this stage to give further general guidance on PQ requirements as the checks and tests necessary to demonstrate an instrument's satisfactory performance are specific and dependent on both the instrument type and the analytical application. However, PQ must be formally documented in accordance with the general requirements set out in the previous section on Documentation.

4.6 Calibration and Traceability

It can be important, and necessary, to establish traceability to national and international standards to ensure the accuracy of the data produced during the measurement process. Where this is not relevant or possible, the basis for calibration or the approach taken to establish the accuracy of results must be documented.

Where instruments are used to determine absolute values of a parameter (for example, temperature or wavelength) the instrument should be calibrated using reference materials or standards traceable to national or international standards. Most analytical instruments are not used in this way. Instead, the instrument measurement (for example, mV) is compared with the value for a known quantity of the

determinand of interest, in a calibrant, in a way which obeys definable laws. Thus, the traceability of the actual parameter measured (mV) is unimportant so long as the standard used to calibrate the measurement is traceable and the instrument response, in relation to the concentration of the determinand, is predictable.

For many applications, the accuracy of the instrument's operating parameters (for example, mobile phase flow rates in HPLC systems) is not critical and hence the need for traceable calibration to national or international standards is less important. In such circumstances, the accuracy of the operating parameter is secondary, provided that it remains consistently reproducible during the analysis of both the sample and the standard, and the satisfactory performance of the measuring system can be demonstrated (for example, by System Suitability Checking).

However, in other circumstances, the accuracy of an instrument's operating parameters, and hence calibration traceable to national or international standards, will be more important (for example, where an analytical procedure developed in one laboratory is to be transferred for routine use in another laboratory or where the accuracy of the parameter may have a critical impact on the performance of the measurement process).

Traceability to national and international standards is usually, and often most efficiently, established through the use of certified reference materials or by standards which are themselves traceable in this way.

Users should avoid over-specifying calibration and/or traceability requirements (for example, for parameters which are not critical to the method) as independent reviewers will expect users to demonstrate that any tolerances specified in the procedures can reasonably be met.

4.7 Re-qualification

In general, an instrument will undergo a variety of changes during its operational life. These can vary from the routine replacement of a single consumable part, through to very significant changes affecting the entire instrument system.

Examples of such circumstances include:

- Movement or re-location of the instrument.
- Interruption to services or utilities.
- Routine maintenance and replacement of parts.
- Modification (for example, instrument upgrades or enhancements).
- Change of use.

Whenever such changes take place, it is essential to repeat relevant aspects of the original qualification process. This procedure is widely referred to as 're-qualification'.

The level of re-qualification required will depend on the extent to which change has occurred and its impact on the instrument system. In many cases, re-qualification can be performed using the same EQ protocols and checks and tests which were undertaken prior to the routine use of the instrument.

The nature of, and reason for, any change to the instrument system, along with the results of all re-qualification checks and tests performed, should be formally documented according to the requirements set out in the previous section on Documentation.

Re-qualification may not necessarily mean repeating the entire EQ process. However, it must clearly address the specific change and re-qualify those parts of the instrument system which are affected by the change. For example, the replacement of an HPLC detector source (for example, a deuterium lamp) would require the detector to be re-qualified using appropriate OQ/PQ procedures and protocols, but would be unlikely to require the individual re-qualification of other components of the HPLC instrument (for example, an injector or pump). However, because the change had affected the instrument as a whole, it would also be necessary to carry out PQ checks on the entire system to demonstrate its satisfactory performance following the change.

Similarly, for some 'modular' systems, it is often possible to interchange components depending on the application and intended use of the instrument. Changes to the instrument system configuration (for example, replacing one detector with another) may not necessarily require re-qualification of the individual modules, but would require re-qualification of the instrument system as a whole.

Significant changes to the instrument system (for example, major component or software upgrades, or enhancements which increase the instrument's capabilities) will normally require more extensive re-qualification. Indeed, for such substantial changes, it is often arguable as to what might be considered to be re-qualification and what constitutes qualification of a new component.

Upgrades to the instrument and/or its software should be fully documented and should describe the reasons for the change, including differences, new features and benefits of the change. Users should ascertain and seek documented evidence from suppliers that upgrades have been developed and manufactured to appropriate standards and formally validated during production. Software upgrades should, as far as possible, be compatible with previous versions and, where this is not possible, the supplier should offer a 'validated' transfer of existing data to the upgraded system.

Following installation of the upgrade, the instrument should be re-qualified using appropriate checks and tests. Where possible, the checks and tests used for re-qualification should be designed so that the results can be compared with those obtained using earlier versions. Any differences in the test results obtained from old and new versions should be identified, documented and resolved.

4.8 Accreditation and Certification

Although different laboratory accreditation systems will have different specific requirements, there are a number of basic principles which apply.

a) Accreditation is intended to provide users and their customers with confidence in the quality of the user's testing activities, and in the technical and commercial integrity of the user's operations. Users are normally assessed and accredited to perform specific tests in specific fields of measurement.
b) The basic requirement is that instruments must be fit for purpose and suitable for their intended use. There should be adequate capacity to meet the requirements of the studies and tests which will be carried out. Generally, assessors will be concerned with the instrument's fitness for purpose in the context of the test concerned and the accuracy required of results. In this respect, consideration must be given to the overall measurement uncertainty, which will include a contribution from the instrument.
c) Instruments must be protected, as far as possible, from deterioration and abuse, and must be kept in a state of maintenance and calibration consistent with their use. They must be capable of achieving the level of performance (for example, in terms of accuracy, precision, sensitivity, etc.) required, and to comply with any standard specifications relevant to the tests concerned. Records of maintenance and calibration must be kept.
d) Generally, instruments of established design must be used. Where other instruments are used, the user must demonstrate that they are suitable for their intended purpose. New equipment must be checked for compliance with appropriate specifications, commissioned and calibrated before use. All computer systems used to control instruments must themselves be subject to formal evaluation before use.
e) Instruments must only be operated by authorised and competent staff, and these must be named in the appropriate procedures. Adequate, up-to-date, written instrument operating instructions must be readily available for use by staff.

4.9 References

[1] P. Bedson, M. Sargent: The development and application of guidance on equipment qualification of analytical instruments. Accred. Qual. Assurance (1996) 1, 265–274.
[2] P. Bedson, D. Rudd: The development and application of guidance on equipment qualification of analytical instruments: High performance liquid chromatography. Accred. Qual. Assurance (1999) 4, 50–62.
[3] D. Holcombe, M. Boardman: Guidance on equipment qualification of analytical instruments: UV-visible spectro(photo)meters (UV-Vis), Accred. Qual. Assurance (2001) 6, 468–478.

5
Validation During Drug Product Development – Considerations as a Function of the Stage of Drug Development

Martin Bloch

5.1
Introduction

ICH Guidelines Q2A [1] and Q2B [2] provide guidance on the validation parameters to be covered, on the extent of validation required and the procedures to be followed for analytical method validation as part of the registration dossier. The first sentence from ICH 2QA [1] makes this clear: "*This document presents a discussion of the characteristics for consideration during the validation of the analytical procedures included as part of registration applications submitted within the EC, Japan and USA.*"

Thus, in the earlier phases, the ICH guidelines are not yet formally applicable. This leaves us with the question, what extent of method validation is needed at these stages of development. During all phases of drug development a wealth of analytical data has to be accumulated. It is commonly accepted that simplified procedures can be followed, reducing the amount of work and documentation needed for method validation.

"The objective of validation of an analytical procedure is to demonstrate that it is suitable for its intended purpose". The sentence is from the introduction to ICH Q2A [1]. These very few words contain the fundamental concept of method validation, the guiding thought behind it. In the view of the author, the sentence is so important, that it is worthwhile to spend some further thoughts on it.

Under the influence of GMP and all the numerous international and company-internal guidelines, regulations and standard operating procedures (SOPs) we – the analytical scientists – have increasingly adopted an attitude which is driven by the simple wish to 'comply' with the 'requirements'. More and more we perceive these as *external constraints*. Our constant question is: what am I expected to do? In our servile desire to comply, we have become all too willing to do almost anything – if some authority suggests it. While doing so, we risk to lose more and more of our personal competence, judgment and responsibility. Yet: do we not owe it to *our own, most intrinsic* professional pride and responsibility – irrespective of any external authority and guideline – to guarantee that the analytical results which we produce have been obtained by a methodology which is 'suitable for its intended purpose'? Would *we* accept the opposite, that the applied method was not suitable?

Method Validation in Pharmaceutical Analysis. A Guide to Best Practice. Joachim Ermer, John H. McB. Miller (Eds.)

ISBN: 3-527-31255-2

While the sentence is from ICH, it is so true that there can be no doubt: it must apply to all phases of development. Any analytical results – wherever and whenever reported – must be 'correct', that is, correct in the sense that conclusions derived from them are based on a sound ground – or else the analytical results are of no value.

Much of the existing literature on analytical method validation is dominated by the phrase 'you should'. By contrast, it is the intention of the present chapter's author to propose practical suggestions and approaches as to *how* we can achieve what 'we should' – or better, what we know, to be necessary. These were the leading thoughts when some of my colleagues at Novartis and myself sat together to design the proposals which are presented in this chapter:

- We commit to do serious analytical work, which can be relied upon.
- We want to make economic use of our available resources. We consider analytical method validation neither to be an art nor a science of its own.
- We are searching for solutions, which combine scientific seriousness with economic efficiency.

Or simply:

- How can we guarantee the trustworthiness of our results with a minimum effort for validation?

This leaves us with the question: what exactly is 'suitable', under particular conditions, in this or that respect? One statement has already been made above: we want to make sure our results are 'correct', that is, correct in the sense that solid conclusions and founded decisions can be derived from them. Thus, we do not need accuracy or precision or sensitivity just for the sake of them. We need sufficient (but no more) accuracy, precision and sensitivity in order to make sure that conclusions drawn from the analytical results can be relied upon and well-founded decisions can be based on them.

ICH guidelines on validation do not exist at the early stages of pharmaceutical development. Starting with the ICH guidelines as a benchmark and, knowing that adequate simplifications during early development can be justified, we asked: *What* do we have to do in order to make sure that a method is suitable for its intended purpose? We 'empowered' ourselves to rely on our own professional expertise and responsibility and to look for practical answers.

5.2 Validation During Early Drug Development

As described above, ICH guidelines Q2A [1] and Q2B [2] are not yet binding at this stage. It is commonly accepted, that simplified validation procedures are adequate. A common recommendation says: In early development, start with no or only a 'crude' validation, then refine and expand the validation step-by-step during product development until finally it fulfils ICH requirements at the time of registration.

This is a valid approach. Yet, upon further reflection, a number of drawbacks can be identified as follows.

- There is no practical way to add additional concentration levels to an existing linearity experiment in order to expand and refine it.
- It is likely that, during development, the analytical method itself undergoes certain improvements and changes. Strictly, the old validation results will no longer be valid for the modified method.
- Quite generally, in actual practice it turns out to be difficult and time consuming to draw together results from validation experiments that are spread over years and to write a neat, consistent validation report on that basis.
- The chances are, that one will end up with a rather messy set of raw data. When, during a pre-approval inspection, a government inspector wants to have a look at them, we will have to dig out raw data spreading over the complete development period and we will have to present them to the inspector; the lack of structure in the data set may provoke him or her to ask uncomfortable questions.
- For validation in full development it is expected to base the experiments on a specific validation protocol with pre-defined procedures and acceptance criteria. If much of the validation work has previously been performed, writing a protocol afterwards is of questionable value.

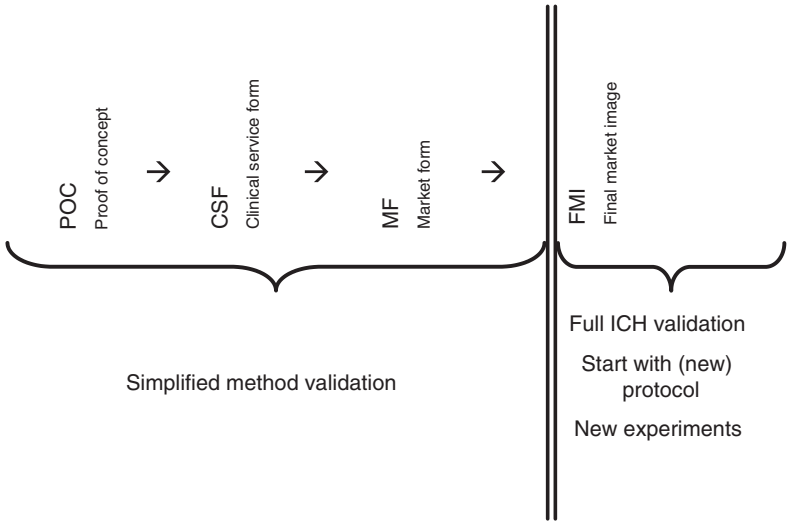


Figure 5-1 Method validation during drug development.

For these reasons, the approach in Figure 5-1 is proposed. It suggests the application of a simplified validation methodology from the very early phase of development up to and including the market form development. But then, at the time of defining the final market image, a new validation protocol is written and all relevant

analytical tests are validated from scratch. In this way, the validation protocol, the experiments and the associated raw data, together with the validation report, will form a clean and neat set.

5.2.1 Simplifications During Early Development

Table 5-1 summarizes the parameters which have to be considered when validating a method for full development. The table is an adapted and expanded version of similar tables found in ICH Q2A [1] and in USP. For example, it is self-explanatory that, for an identity test, specificity needs to be validated, while accuracy, linearity, precision, limit of detection and limit of quantification have no meaning in its context. Similarly, although the limits of detection and quantification are important quality parameters of an impurity test, they are, however irrelevant for the assay, and so on.

Table 5-1 Types of tests and parameters to be validated for full development.

Parameter	Type of test				
	Identity	**Assay/content uniformity/ dissolution**	**Impurity testing: semi-quantitative or limit test**	**Impurity testing: quantitative test**	**Physical tests**
Specificity	Yes	Yes	Yes	Yes	No
Linearity	No	Yes	No	Yes	No
Accuracy	No	Yes	*	Yes	No
Precision (repeatability)	No	Yes	*	Yes	Yes
Precision (intermediate precision)	No	Yes	No	Yes	*
Precision (reproducibility)	No	**	No	**	No
Range	No	Yes	*	Yes	No
Limit of detection	No	No	Yes	(No)	No
Limit of quantitation	No	No	No	Yes	*
Stability of the solutions	No	Yes	*	Yes	*
Robustness	*	Yes	*	Yes	*

* may be required, depending on the nature of the test
** in exceptional cases

Furthermore, at the later stage of development, it is commonly understood, that the validation experiments are based on a specific *validation protocol* presenting information on

- the testing instructions concerned, the product name, etc.
- the tests concerned
- the parameters to be validated for each test
- acceptance criteria

- details on the design of the validation experiments, such as type and number of solutions to be prepared and exactly how they are prepared
- batches, reference materials
- equipment
- responsibilities / signatures.

For an 'ICH-validation' the validation report refers to the validation protocol and it addresses the abovementioned items; in addition, it presents tables of results together with explanatory text and graphical representations as well as conclusions to be drawn from the results.

What are the simplifications that may be envisaged during the earlier stages of development? Every analytical scientist is aware of the innumerable sources of error that could possibly hamper a measurement and thus the quality and trustworthiness of his results. No analytical result can be meaningful if it is reported without some information on its reliability, that is, its sensitivity, accuracy and precision. For these reasons it is indispensable to perform certain reliability checks for any kind of analytical measurement. In this context, a sentence from the ICH Q7A guideline [3] on "Good manufacturing practice for active pharmaceutical ingredients" can be quoted: "*While analytical methods performed to evaluate a batch of API for clinical trials may not yet be validated, they should be scientifically sound*". Depending on the test under consideration, some information on linearity, accuracy, specificity, precision/repeatability, reporting level / limit of quantification and limit of detection, is an essential prerequisite for any analytical work, in particular also during early development.

It may be sobering to realize that this encompasses much of what is needed for a full validation according to ICH and the question remains: *is there any room left for permissible simplifications?* Where exactly can the effort be reduced? Here are a few proposals.

- A formal validation protocol is not yet mandatory. Instead, for instance, an SOP may summarize the generalized requirements.
- Formally documented intermediate precision experiments are not yet needed. (However, if during development different laboratories are involved in the analyses, the responsible analytical managers must decide on the extent of necessary work for method hand-over and training.)
- Formally documented robustness testing is not yet required. (However, it is strongly advisable to build ruggedness into the methods at the time when they are developed, for instance, with the help of software such as DryLab or ChromSword; for more details see Section 2.7)
- The extent of testing and the number of replications may be reduced.
- For precision testing it is acceptable to use mixtures of drug substance and placebo. (Note that in late development, for a full ICH validation, Guideline Q2B [2] specifies that for precision testing 'authentic samples', that is real tablets, capsules, etc., should be analysed).
- Wider acceptance criteria may be adequate.

- The validation report may be presented in a simplified form based mainly on tables of results and certain pertinent graphs, together with the conclusions (but only with a minimum of additional explanatory text).

Of course, these suggestions still leave a lot of room for interpretation. Deliberately, no attempt is made in this chapter to set up detailed and generally valid rules or 'recipes', which would be valid for all methods, tests and circumstances. Instead, a selection of pertinent specific examples will be discussed and it will be left to the reader to apply similar thoughts to his particular case.

5.2.2 Example 1: Assay or Content Uniformity of a Drug Product by HPLC During Early Drug Product Development: Proposal for a Validation Scheme

Here is a proposal for a simple scheme for the validation of HPLC methods for the determination of the assay and / or for content uniformity. As always, our design criteria are best summarized by the sentence: How can we guarantee the trustworthiness of our results with a minimum effort for validation?

5.2.2.1 Accuracy, Linearity, Precision and Range

We found that sufficient information on accuracy, linearity and precision can be derived from one set of injections from only seven solutions, see Table 5-2: The five solutions A–E are prepared by adding varying amounts of drug substance always to the same amount of excipients (100 % = the nominal content of excipients in the formulation). These mixtures are subjected to the sample preparation procedures described in the analytical method for the sample solution. The two solutions Ref1 and Ref2 are the reference solutions prepared according to the analytical method. The seven solutions A–E and the solutions Ref1 and Ref2 are injected twice each.

Table 5-2 Accuracy, linearity, precision and range combined for the validation of the assay or content uniformity by HPLC.

Solution	% Drug substance	% Excipients	% Recovery
A	50	100	RA
B	80	100	RB
C	100	100	RC
D	120	100	RD
E	130	100	RE
Ref1	100	–	–
Ref2	100	–	–

Evaluation: Five recoveries can be calculated for solutions A – E, as well as the averaged recovery which is reported as a quality parameter for the accuracy of the method. The standard deviation of the individual recoveries may be reported as a measure of precision. The responses of the solutions A – E are subjected to a linear

regression calculation. The linearity of the method is assessed from a graph (response versus injected amount) and from the residual standard deviation of the regression and its y-intercept. For completeness, the correlation coefficient is also reported (but should not be misinterpreted as a proof of linearity, see also Section 2.4.1.2). The residual standard deviation represents the scatter of the individual data points (y-values = detector response at the different, given concentrations) around the averaged regression line:

$$\text{Residual standard deviation (linear regression)} \quad s_y = \sqrt{\frac{(y_i - y_i^*)^2}{df}} \tag{5-1}$$

y_i = response at concentration x_i
y_i^* = calculated response from regression, at concentration x_i
df = no of degrees of freedom, for linear regression $df=n-2$

Note that the residual standard deviation for a linear regression has much the same form as the standard deviation of individual values around their mean:

$$\text{Standard deviation (around a mean)} \quad s = \sqrt{\frac{(x_i - \overline{x})^2}{df}} \tag{5-2}$$

x_i = individual data points
$\overline{x}$ = mean
df = no of degrees of freedom, for mean $df=n-1$

In the case of our experiment, the deviations of the individual responses from the calculated regression line include errors introduced during sample preparation as well as errors originating from the HPLC analysis. Thus the residual standard deviation can be taken as a valid measure for the precision of the analytical method. It represents much the same content of information as the standard deviation, which is calculated during an ICH validation when all samples have the same concentration of drug substance.

5.2.2.2 **Specificity / Selectivity**

Chromatograms of

- placebo
- DS
- DS + placebo
- DS + placebo stressed, for instance 8h at 80 °C (such that some, but not more than 10 % of the drug substance has degraded)

are recorded and visually compared.

5.2.2.3 **Stability of the Solutions**

Solutions C and Ref1 are re-injected after 24 hours – or better also after 36 hours – and the change of the absolute response is reported; (this simple approach is valid, if the system is left running during this period; normally within this timeframe the drift of the response can then be neglected). If the sample and reference solutions

prove to be stable over two days, this will normally be sufficient for analyses during pharmaceutical development, where sequences normally do not contain very large numbers of samples and hardly ever run longer than 24 hours. If it can be shown that they are also stable for three days, this will be of value in the case of problems one may later encounter during analyses: in the course of an investigation into the cause of the problem, we know that the solutions have not yet degraded and they can be re-injected on the following day.

A suggestion for acceptance criteria is found in Table 5-3. (Note: for practical reasons and in order to improve their analytical interpretation, the acceptance criteria for the residual standard deviation and the y-intercept of the regression line have been specified as percentages. An explanation will be given in Note 2 of Section 5.2.4)

Table 5-3 Early development. Acceptance criteria for the validation of the assay or content uniformity for a drug product.

Parameter	Acceptance criteria	
Accuracy	Average recovery	95 –105%
Precision	RSD[1)] of individual recoveries	≤ 2.5 %
	or	
	Residual standard deviation from linearity	≤ 2.5 %[2)]
Linearity	Residual standard deviation	≤ 2.5 %[2)]
	y-intercept	≤ 10% [2)]
	Correlation coefficient *R*	≥ 0.997
Stability of the solutions	Change of response over 24 (or preferentially 36 h), solutions C and Ref1	Each ≤ 2 %
Specificity	Visually compare chromatograms	No interference with drug substance peak

1 RSD = relative standard deviation
2 Relative to the response for 100% drug substance content (solution C)

5.2.3 Variation of Example 1: More than on Strength of Drug Product

If the analytical method comprises more than on strength of the product, for instance 0.5 mg, 1 mg and 3 mg, it is advantageous to design the analytical method such that the sample solutions are diluted to the same final concentrations. In our example, let us assume the following: 0.5 mg strength dissolved in 100 ml, no further dilution; 1 mg strength dissolved in 100 ml, then dilution by factor two; 3 mg strength dissolved in 100ml, then dilution by factor six. In this case the validation can be performed as described in Section 5.2.2 for the strength with the lowest drug substance / placebo ratio, normally the 0.5 mg strength. In addition to the solutions proposed in Section 5.2.2, in the sense of a bracketing approach, one additional solution F needs to be prepared for the 3 mg strength, with 100 % drug substance and

100 % placebo; from it, the recovery is calculated. Since, according to the analytical method, all sample solutions are diluted to the same concentration, the linearity calculation using solutions A – E is also valid for the other strengths; the response from solution F may be included in the regression.

5.2.4 Example 2: Degradation Products from a Drug Product by HPLC During Early Drug Product Development: Proposal for a Validation Scheme

Consider the following points, which are specific to degradation products.

- During early drug product development, degradation products may not yet have been elucidated and even then, reference standards are normally not yet available. It is likely, though, that the degradation products exhibit chemical and physical similarity to the drug substance. For this reason, in the absence of a better alternative, it is a commonly accepted approach to employ the drug substance itself as a representative substitute to measure the validation parameters for degradation products.
- Linearity of the method should be demonstrated down to the reporting level. At this low concentration, a recovery experiment may be difficult to conduct during early pharmaceutical development. (Note that at this stage, not much experience with a new method may yet be available.)
- The linearity test for degradation products may be combined with the test for the limits of detection (LOD) and quantitation (LOQ).

Based on these considerations, the design of the following scheme deliberately differs from the one presented for the assay in Section 5.2.2. On purpose, the test for linearity is not combined with the one for accuracy and precision.

5.2.4.1 Accuracy and Precision

Perform at least five recovery experiments at the level of the specification for individual degradation products. Thus, if the specification limits degradation products at 0.5 %, spike 100 % placebo with 0.5 % drug substance – spiking with a solution is acceptable. Carry through all sample preparation steps given in the analytical method for the sample solution. In order to measure accuracy, calculate the recovery by comparing the response to the response from reference solutions prepared according to the analytical method. The standard deviation of the individual recoveries is taken as a measure for precision.

5.2.4.2 Linearity, Limit of Quantitation (LOQ) and Range

Starting from a stock solution of the drug substance, prepare at least five dilutions at concentrations from the reporting level up to 1.2 or 1.5 times the specification. Inject the solutions and evaluate linearity from a graph as well as by calculation of a linear regression. Calculate and report the y-intercept, the residual standard deviation, and for completeness also the correlation coefficient (which should not be misinterpreted as a proof of linearity, see also Section 2.4.1.2). From the same linear

regression LOQ can also be estimated by the formula given in ICH Q2B [2]. (Please note that unrealistically high and unfavourable values for LOQ are obtained, if points with concentrations very far away from LOQ are included in the regression; see Section 2.6.4 and Figure 2.6-11; a separate experiment to determine LOQ may be necessary in such a case).

$$\text{Limit of Quantitation} \quad LOQ = \frac{10\ \sigma}{b} \tag{5-3}$$

σ = residual standard deviation (or standard deviation of the y-intercept)
b = slope of the calculated regression line.

Similarly,

$$\text{Limit of Detection} \quad LOD = \frac{3.3\ \sigma}{b} \tag{5-4}$$

5.2.4.3 **Verification of the Reporting Level**

If LOQ is calculated as described above, no separate experiment for the verification of the reporting level is required. As an alternative or in order to further verify the reporting level, one may choose to inject a solution containing the drug substance at the concentration of the reporting level at least five times. The relative standard deviation of the response is then calculated and reported. The logic is the following: if at the reporting level the relative standard deviation from the responses of repeated injections is below, for example, 20 %, this means that the peak can be quantitated with sufficient precision. Put differently, the requirement that the reporting level must not be lower than LOQ is fulfilled. For instance, the following statement can be made:

- At the reporting level of 0.1 % corresponding to a concentration of 20 ng/ml the relative standard deviation of the response was found to be 5.3 %. Since this is lower than 20 %, the reporting level lies above LOQ.

(Such a 'verification of the reporting level' is not only useful during method validation, but the author also strongly recommends it as an SST parameter in routine analyses for degradation products or other impurities.)

5.2.4.4 **Specificity/Selectivity**

Same as for the assay (visual comparison of chromatograms).

5.2.4.5 **Stability of the Solutions**

The reference solution and the placebo solution spiked with drug substance at the level of the specification are re-injected after 24 or, preferably, also after 36 hours and the change of the response is reported.

A suggestion for acceptance criteria is found in Table 5-4.

Table 5-4 Early development: Acceptance criteria for the validation of degradation products of drug products.

Parameter	Acceptance criteria	
Accuracy	Average recovery	80 –120%
Precision	RSD[1)] of individual recoveries	≤ 15 %
	or	
	Residual standard deviation from linearity	≤ 15 %[2)]
Linearity	Residual standard deviation	≤ 15 %[2)]
	y-intercept	≤ 25%[2)]
	Correlation coefficient R	≥ 0.98
LOQ	10*σ/slope[3)]	≤ Reporting level
Verification of the reporting level	RSD[1,4)]	≤ 20%
Stability of the solutions	Change of response[5)] over 24 h (or preferentially: 36h)	Each ≤ 20 %
Specificity	Visually compare chromatograms	No interference with DS peak

1 RSD = relative standard deviation
2 relative to the response at the concentration of the specification
3 from linear regression / linearity; σ = residual standard deviation or standard deviation of the y-intercept
4 solution containing drug substance at a concentration corresponding to the reporting level
5 reference solution and placebo solution spiked with drug substance at the level of the specification

Note 1: *Combined or separate linear regressions for low and upper range?*

As explained above, in order to validate the degradation product method in the absence of standards for degradation products, it is common practice to employ dilute solutions of the drug substance in lieu of the degradation products. Thus, in this case, the linearity has to be checked for the drug substance in the low range (reporting level to, for instance, 1 %) and also – for the validation of the assay – in the high range 80 % – 120 % of the declared drug substance content. If the degradation products are evaluated as area percentages, or if they are calculated with respect to a 100 % reference solution, it has also been suggested by some authorities, that linearity should be assessed using *one* regression calculation spanning the complete range from LOQ up to 120 %. The reasoning is as follows: if the degradation products are evaluated against the 100 % drug substance peak, the result is only correct if the method is linear down to the reporting level. In the view of the author, the suggested approach should not be adopted for the following reasons. It does not make analytical sense to space the individual concentrations equally over the complete range. Instead, one will probably choose five or six concentrations closely spaced between 0.05 and 1 % and another five or six between, for instance, 50 and 130 %. In such a situation, the injections between 50 and 130 % will have a levering effect on the y-intercept. (See also Section 2.4.1.4 and Figure 2.4-8). One may either be led to the erroneous conclusion that the y-intercept is unacceptable for degradation prod-

uct evaluation, or else the opposite may be disguised: that the y-intercept is satisfactory, when in fact it is not.

Instead of the combined linear regression over both concentration ranges, it is the better alternative to perform two separate linearity experiments and to calculate two separate linear regressions for the two concentration ranges.

If

- within relatively wide acceptance limits the slope for the lower range does not differ too much from the slope for the upper range, and
- the y-intercept for the regression in the lower range is acceptably small,

then only negligible errors are introduced when the degradation products are calculated by the rule of proportion with respect to a 100 % drug substance reference solution. However, note that it is the much better alternative to design the analytical method such that degradation products are calibrated with a dilute drug substance solution, for instance 1 % or 0.5 %, instead of 100 %; this will avoid the extrapolation error.

Note 2: *Interpretation of the results from the linear regression and acceptance criteria.*

Consider the data of Table 5-5 and Figure 5-2 from the validation of linearity for a degradation product analysis.

Table 5-5 Interpretation of the results from the linear regression and acceptance criteria.

Solution no	Concentration (mg in 100 ml)	Injected amount (ng)	Injected amount (% of nominal)	Area found (average)	Area calculated	Area difference
1	0.302	60.315	0.60	3246 552	3240 415	6137
2	0.241	48.252	0.48	2566 749	2587 911	–21 162
3	0.201	40.210	0.40	2186 501	2152 909	33 592
4	0.101	20.105	0.20	1026 678	1065 403	–38 725
5	0.040	8.042	0.08	412 340	412 899	–559
6	0.020	4.021	0.04	216116	195 398	20 718

	Estimate	s	95% Confidence interval
Slope	54 091.3	582.4	[52 474.2 ; 55 708.4]
y-intercept	–22 103.3	21 357.5	[–81 401.4 ; 37 194.8]
Residual standard deviation		29 762.0	
Correlation coefficient		0.999 768 203	
|y-intercept in % of area for 40 ng (tolerated amount)			1.0%
Residual standard deviation in % of area for 40 ng (tolerated amount)			1.4%

The residual standard deviation is 29762.0 and the y-intercept is -22103.3. What do these numbers tell us? Are the values excellent, or acceptable, or unacceptably bad? It is obvious, that we have no 'feeling' for the absolute numbers. If we want to interpret them, they must be put into perspective, relative to – well, to what? Similarly, in the somewhat analogous case of individual data (such as results of an HPLC

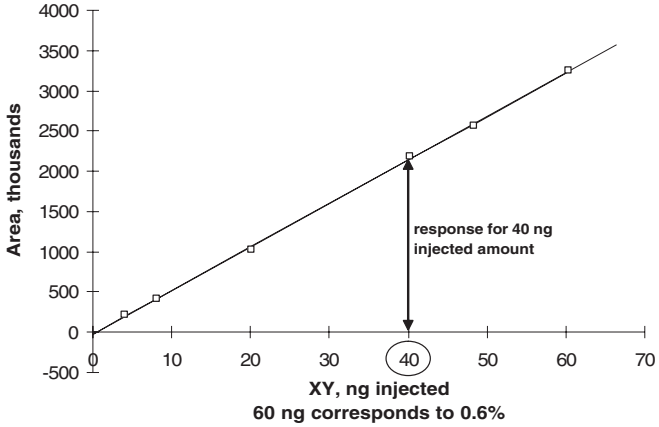


Figure 5-2 Interpretation of linear regression.

assay) scattering around their mean, only knowing that the standard deviation is, say 22654, is of little value unless we know the mean. In this case calculation of the relative standard deviation relates the value to the mean:

$$\text{Relative standard deviation} \quad RSD = \frac{100\ s}{\overline{x}}\ \% \tag{5-5}$$

s = standard deviation
$\overline{x}$ = mean.

In the example, assuming $\overline{x}$ = 9479829, RSD is calculated as 0.24 %. If the data represent the HPLC assay of the drug substance content in a drug product, we all have a good feeling for the number and we know that 0.24 % is fine. If we now return to the quality parameters of the linear regression, let us note the following.

- Upon inspection of the definition of the residual standard deviation (Eq. 5-1) it becomes clear that it is measured in the units of the y-scale, that is the response, or the peak area.
- The y-intercept is obviously also measured in units of the response (peak area).
- A meaningful mean response or mean peak area, to which the experimentally determined residual standard deviation and the y-intercept could be related, does not exist.

For this reason the author proposes to relate the residual standard deviation and the y-intercept to the responses (peak areas) which one obtains for prominent concentrations, as follows.

- In the case of an assay of a drug substance in a drug product: to the response (peak area) obtained for the amount corresponding to the declared amount in the drug product.

- In the case of a method for a degradation product (or of an impurity): to the response (peak area) obtained for the tolerated amount of the degradation product according to the product specification.

For the example of the data in Table 5-5 and Figure 5-2 let us assume the degradation product XY has been limited to 0.4 % according to its product specification and 0.4 % corresponds to an injected amount of 40 ng. The response for this amount is about 2100 000 and the residual standard deviation, 29 762.0, is 1.4 % thereof. Similarly, the absolute value of the y-intercept, 22 103.3, is just 1.0 % of 2100 000. Thus, for the present example, the following statements can be made.

- The scatter of the individual data points around the regression line is represented by the residual standard deviation which is 1.4 % of the peak area obtained when 0.4 ng are injected; 0.4 ng corresponds to the tolerated amount of 0.4 %.
- The y-intercept of the regression line is 1.0 % of the peak area obtained when 0.4 ng are injected, corresponding to the tolerated amount of 0.4 %.

These calculations have been detailed in the lower part of Table 5-5. Using similar considerations, *meaningful and easy-to-interpret acceptance criteria* can be specified for the residual standard deviation and the y-intercept, as follows.

For a degradation product method, for instance:

- residual standard deviation ≤ 15 % of the response for the tolerated amount;
- |y-intercept| ≤ 25 % of the response for the tolerated amount.

For an assay, for instance:

- residual standard deviation ≤ 2.5 % of the response for the declared amount;
- |y-intercept| ≤ 10 % of the response for the declared amount.

By the way, it has often been requested that acceptance criteria, for instance for the y-intercept, should be based on statistical significance tests. One would then request that the confidence interval for the y-intercept should include zero. In the view of the author, this is not really relevant in the context of method validation. The true question is not whether, based on statistical significance, the line passes through zero; what really is of interest is the question: if the y-intercept has a certain magnitude, what influence will it have on the results of the analyses when they are performed and calibrated according to the analytical method (see also Section 1.4.2)? If one does find that the intercept systematically differs from zero, depending on its magnitude one may

- conclude that it does not – to an analytically meaningful extent – affect the outcome of a (degradation product) analysis; or
- adapt the analytical method such that, instead of a one-point calibration a dilution series of calibration solutions is specified and calibration is accomplished via linear regression.

For these reasons the author strongly favours the 'practical' approach of specifying acceptance criteria with respect to the response (peak area) at a prominent concentration, as explained above.

Note 3: *Relative detector response factors*

It has been pointed out above that usually, early in the development of degradation products, reference substances of sufficient and known purity are not available. However, during development such materials may become available and the need will arise to establish the relative detector response factors. The relative response factors are best calculated from the slopes of the regression lines calculated for dilutions containing the respective substances in the relevant concentration range.

$$\text{Relative detector response factors} \quad Z_i = \frac{Slope\ \ Main\ \ Peak}{Slope\ \ Substance\ \ i} \tag{5-6}$$

For the example, presented in Figure 5-3, the response factors calculated according to this formula are ZA = 197 482 / 232 696 = 0.85 and ZB = 197 482 / 154 695 = 1.28.

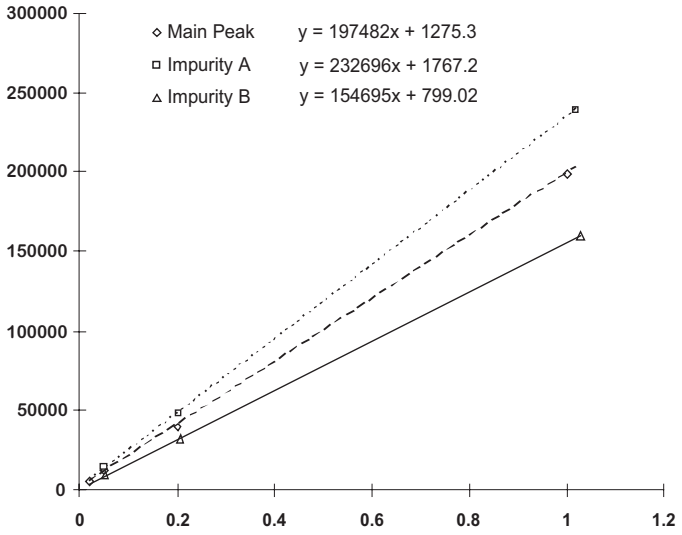


Figure 5-3 Relative detector response factors.

5.2.5

Example 3: Residual Solvents of a Drug Product by GC During Early Drug Product Development: Proposal for a Validation Scheme

Residual solvents can be treated much in the same way as degradation products, with the following exception: in the combined accuracy / precision experiment, obviously the residual solvent is spiked to a mixture of drug substance and placebo.

5.2.6 Example 4: Analytical 'Method Verification' for GLP Toxicology Study

Early in the development of a drug product, 'toxicology batches' are produced. These are mixtures of drug substance in a vehicle which is suitable for application to the animals in toxicological studies. Such batches must be analysed for release and monitored for their stability during the time of the study. The release and stability requirements for such batches are wide. Usually it is sufficient to show that the content of drug substance remains within 90–110 % during the study and that no substantial decomposition takes place. Also, usually not much time is available for analytical method development and validation. An elaborate method validation would certainly be a waste of resources. Nevertheless, false results are unacceptable in any case and it must be demonstrated that, within the above framework, the analytical results are reliably 'correct'. The challenge thus consists in designing the simplest possible *'method verification'* experiments, which are just sufficient to guarantee the requirements and which involve the least amount of analytical effort. We use the term *'method verification'* for the simplified approach that follows, in order to make clear that additional *method validation* effort will later be necessary in case the same method is subsequently also used for other types of analyses in early development. Method verification can be accomplished simultaneously with the analysis. Only six solutions and eight injections are required for method verification:

A	1 injection	vehicle
B	2 injections	100 % drug substance without vehicle
C	2 injections	100 % drug substance in vehicle
D	1 injection	85 % drug substance without vehicle
E	1 injection	115 % drug substance without vehicle
F	1 injection	0.5 % drug substance without vehicle

Solution B also serves as reference for the analysis.

Method reliability is then documented as follows:

- Recovery is measured from the response of solution C against solution B.
- Linearity is measured from the responses of solutions B, D and E. [Note: It must be stressed that this crude linearity experiment with only three levels (and with the wide acceptance criteria of Table 5-6) can certainly not be recommended in other cases; based on only three levels all values calculated from the regression will suffer from large uncertainties. The approach is only accepted in the present, simple situation where it is only required to show that the toxicology batch contains an amount of drug substance in the range 90–110 %. Alternatively, in order to avoid the regression with three levels, one may formulate an acceptance criterion like this: responses (area/concentration) of solutions D and E relative to C should be within 95–105 %.]
- The peak from solution F demonstrates adequate sensitivity of the method: if present, substantial degradation of the drug substance would be detected.
- Selectivity is visually assessed from the chromatograms.

Acceptance criteria may be set as suggested by Table 5-6.

Table 5-6 Release of a toxicology batch: Suggested acceptance criteria for method verification.

Parameter		Acceptance criteria
Accuracy	Recovery	95–105%
Linearity	Residual standard deviation	≤ 4 %[1)]
	y-intercept	≤ 25%[1)]
Stability of the solutions	Change of response over 24 h, solutions C and Ref1	Each ≤ 2 %
Specificity	Visually compare chromatograms	No interference of vehicle with drug substance peak
Sensitivity	Visually	0.5% drug substance peak visible

1 Relative to the response for 100% drug substance content (solution B)

Since the method verification experiments are performed together with the analysis, a separate system suitability test is not necessary and the analytical work can be kept to an absolute minimum while adequate quality standards are maintained.

5.2.7 Example 5: Dissolution Rate of a Drug Product During Early Drug Product Development: Proposal for Validation Schemes

Considerations similar to the ones in the above examples can also be applied to the validation of a dissolution-rate method during the early phases of development. Special attention must be paid to the fact that the range of expected, 'normal' analytical results in dissolution analyses may be wide and the validation range of the analytical method must be chosen accordingly. Two special cases are described in detail below.

5.2.7.1 Specifications for Dissolution (Q-value) in the Range 70–100 %

For this case, a possible validation scheme is shown in Table 5-7. Solutions A, B, C, D, E all contain 100 % placebo, plus varying amounts of drug substance. In addition, a number of reference solutions are required as well as a placebo solution. The last solution, Dc is for an optional 'filter check' (see below).

The evaluation is as follows – for proposed acceptance criteria see Table 5-8.:

- *Linearity*. The responses from solutions A, B, C, D, and E are evaluated in a linear regression and an x/y graph (response versus amount) is displayed for visual examination.
- *Accuracy*. From the same five solutions A, B, C, D, and E the recoveries are calculated with respect to the mean of reference solutions Ref1 and Ref2.
- *Precision*. A separate experiment for precision is not needed. Instead, the residual standard deviation from the linearity evaluation can be taken as a measure for precision or, alternatively, the standard deviation of the five recoveries may be reported.

Table 5-7 Dissolution rate. Suggested procedure for *Q*-values 70–**100%.**

Solution	DS %	Placebo %	Day of preparation	Recovery	Day of analysis
A	50	100	1	R1	1
B	80	100	1	R2	1
C	90	100	1	R3	1
D	100	100	1	R4	1,2,3,4
E	120	100	1	R5	1
Ref1	100	0	1	–	1,2,3,4
Ref2	100	0	1	–	1
Ref3[1)]	100	0	2	–	2
Ref4[1)]	100	0	3	–	3
Ref5[1)]	100	0	4	–	4
Placebo	0	100	1	–	1
Dc[2)]	100	100	1	R6	1

1 only for analysis by HPLC – not needed for UV analysis
2 From the same vessel as solution D, however, in the last step the sample is centrifuged instead of filtered

- *Selectivity / Specificity.* If the evaluation is by simple UV-absorbance measurement, it is necessary to show that the absorption of the placebo is below a certain, preset value, e.g., 2.5 % of the absorption of the '100 %' solution D. Should this value be exceeded, a correction may be applied to the dissolution rate method in the following way. If λ is the measurement wavelength for the absorption of the drug substance, a (higher) 'correction wavelength' λ_c is defined such that the absorbance of the placebo at the two wavelengths λ and λ_c is approximately the same, whereas the absorption of the drug substance $A(\lambda) \gg A(\lambda_c)$. In such a case, instead of $A(\lambda)$ the difference $A(\lambda) - A(\lambda_c)$ can be defined in the method for the evaluation of the dissolution solutions. Then, of course, the same corrections must be applied in the validation experiments. (In case a suitable λ_c cannot be found, the placebo interference may have to be eliminated by calculation of $A_T - A_{PL}$ where A_T is the absorption of the test solution and A_{PL} is the absorption of a placebo which has been subjected to the same dissolution experiment as the test sample.)

  If the analytical method specifies an HPLC measurement, the chromatograms of solutions D, Ref1 and from the placebo, must be visually checked for the absence of interferences from the placebo in the neighbourhood of the drug substance peak.

- *Stability of the solutions.* Dissolution experiments often last for a long time and it must be shown that the solutions remain stable during the duration of the complete dissolution experiment, for instance, over three days. For this reason, the responses for solutions D and Ref1 should again be measured

daily over the next four days. In the case of a simple UV absorbance measurement, the absorbance values may be directly compared. If an HPLC measurement is used, the stability of the system over three days or more may not be assumed and it is advisable to evaluate the measurement on days two, three and four against a freshly prepared additional reference solution (Ref3, Ref4, Ref5).

- *Filter check*. This test is an enhancement of the test for accuracy and, during the early development phase, it may be considered to be optional: Solution Dc is from the same vessel as solution D. However, in the last step of the sample preparation, the solution is centrifuged instead of filtered. If the absorbance values of the two final solutions D and Dc are the same, losses due to adsorption of the drug substance to the filter can be excluded.

Table 5-8 Dissolution rate, acceptance criteria for *Q*-value specifications 70–**100%.**

Parameter	Acceptance criteria	
Accuracy	Average recovery R1 – R5	90 –110%
Precision	RSD[1)] of individual recoveries R1 – R5	≤ 4 %
	or	
	Residual standard deviation from linearity	≤ 4 %[2)]
Linearity	Residual standard deviation	≤ 4 %[2)]
	y-intercept	≤ 10%[2)]
	Correlation coefficient *R*	≥ 0.985
Stability of the solutions	Change of response days 2, 3 and 4 relative to day 1, solutions D and Ref1[3)]	≤ 3 %[4)]
Specificity	Measurement by UV-absorption:	
	absorbance of placebo	≤ 2.5 %[2)]
	Measurement by HPLC: visually compare chromatograms from solutions D, Ref1 and placebo	No interference with drug substance peak
Filter check (optional)	Filter check: difference between response of solution D and Dc	≤ 4 %

1 RSD = relative standard deviation
2 Relative to the response for 100% drug substance content (solution D)
3 In the case of a simple UV absorbance measurement, the responses on day 3 are directly compared to the responses on day 1
In the case of HPLC analysis, on day 2, 3 and 4 the responses are measured against a freshly prepared reference solution (Ref3, Ref4, Ref5)
4 If the requirement is met for day 2 (or 3), but no longer for day 4 (or 3), the validation is still valid, but a note should be added to the analytical method, stating the limited stability of the solutions.

5.2.7.2 Specifications for Dissolution (*Q*-value) 10 % (Level 1) and 25 % (Level 2)

A validation scheme for this case is shown in Table 5-9.

Table 5-9 Dissolution rate: Suggested procedure for $Q = 10\%$ (level 1) and $Q = 25\%$ (level 2) (USP).

Solution	DS %	Placebo %	Day of preparation	Recovery	Day of analysis
A	2	100	1	–	1
B	5	100	1	R2	1
C	10	100	1	R3	1
D	25	100	1	R4	1,2,3,4
E	45	100	1	R5	1
Ref1	10	0	1	–	1,2,3,4
Ref2	10	0	1	–	1
Ref3[1)]	10	0	2	–	2
Ref4[1)]	10	0	3	–	3
Ref5[1)]	10	0	4	–	4
Placebo	0	100	1	–	1
Cc[2)]	10	100	1	R6	1

1 only for analysis by HPLC – not needed for UV analysis
2 From the same vessel as solution C, however, in the last step the sample is centrifuged instead of filtered

The evaluation is the same as for the case above, with the following exceptions.

- *Accuracy*. The recoveries are calculated only for solutions B, C, D and E, not A (since its concentrations may be near or even below LOQ). (Together with solutions B, C, D and E, solution A is used in the linear regression for the evaluation of linearity).
- *Limit of Quantitation (LOQ)*. For these measurements, in dilute solutions, the LOQ must be known. It can easily be calculated from the linear regression, with the definition from ICH 2QB [2] according to Eq.(5-3). (Please note that unrealistically high and unfavourable values for LOQ are obtained, if points with concentrations very far away from LOQ are included in the regression; see Section 2.6.4 and Figure 2.6-11; a separate experiment to determine LOQ may be necessary in such a case).
- *Acceptance criteria*. See Table 5-10.

5.2.7.3 Other Specifications for Dissolution (*Q*-value)

In other cases not covered by the two examples above, similar approaches may be chosen, whereby the drug substance concentrations of the solutions should span +/– 20 % of the specifications (*Q*-values) of all strengths.

Table 5-10: Dissolution rate, acceptance criteria for *Q*-value specifications 10% (level 1) and *Q* = 25% (level 2) (USP).

Parameter	Acceptance criteria	
Accuracy	Average recovery R2 – R5	90 –110%
Precision	RSD[1)] of individual recoveries R2 – R5	≤ 4 %
	or	
	Residual standard deviation from linearity	≤ 4 %[2)]
Linearity	Residual standard deviation	≤ 4 %[2)]
	y-intercept	≤ 10%[2)]
	Correlation coefficient *R*	≥ 0.985
Stability of the solutions	Change of response days 2, 3 and 4 relative to day 1, solutions D and Ref1[3)]	≤ 3 %[5)]
Specificity	Measurement by UV-absorption: absorbance of placebo	≤ 25 %[4)]
	Measurement by HPLC: visually compare chromatograms from solutions D, Ref1 and placebo	No interference with drug substance peak
Limit of quantitation	From linear regression (linearity test): 10 σ / slope	≤ 5% (=concentration of solution B)
Filter check (optional)	Filter check: difference between response of solution C and Cc	≤ 4 %

1 RSD = relative standard deviation
2 Relative to the response for 45% drug substance content (solution E)
3 In the case of a simple UV absorbance measurement, the responses on day 3 are directly compared to the responses on day 1
In the case of HPLC analysis, on day 2, 3 and 4 the responses are measured against a freshly prepared reference solution (Ref3, Ref4, Ref5)
4 Relative to the absorbance of solution C (10% drug substance)
5 If the requirement is met for day 2 (or 3), but no longer for day 4 (or 3), the validation is still valid, but a note should be added to the analytical method, stating the limited stability of the solutions.

5.2.8
Validation of Other Tests (Early Development)

Table 5-11 lists some additional analytical tests which are relevant for drug product development and which have not been covered in the above text. For these tests, the table proposes validation parameters and acceptance criteria.

Table 5-11 Proposed validation parameters and acceptance criteria for some other tests in early development.

Quality characteristics	Parameter to be validated	Acceptance Criteria
Appearance, disintegration time, density, mass, pH value, sulfated ash, bulk/tamp volume	Not to be validated	
Friability Crushing strength Viscosity Refractive index	Not to be validated	
Loss on drying	Precision / repeatability	Level 0.1–0.2% RSD ≤30%, $n \geq 5$ Level 0.2–0.5% RSD ≤20%, $n \geq 5$ Level 0.5%– 5% RSD ≤10%, $n \geq 6$ Level ≥5% RSD ≤ 5%, $n \geq 6$
Identity (HPLC, TLC, GC)	Selectivity / specificity	Peaks/spots separated
Identity (IR/UV)	Selectivity / specificity	Substance is clearly distinguished from similar products
Water (Karl Fischer)	Precision/repeatability	For assay, RSD ≤ 5 %, $n \geq 5$
	Influence on reaction time: absolute difference of water content measured at time according to analytical method and 1.3 – 1.5 times this reaction time	$\Delta \leq 10\%$
Other quality characteristics	If applicable: project specific	If applicable: project specific

1 RSD = relative standard deviation

Acknowledgements

In several discussions Thomas Gengenbacher, Hans-Peter Hertli and Martin Mueller-Zsigmondy, all at Novartis Pharma in Basel, Switzerland, contributed valuable input and support ; their help is gratefully acknowledged.

5.3 References

[1] CPMP/ICH/381/95 (Q2A), Note for Guidance on Validation of Analytical Methods: Definitions and Terminology (Step 5 - adopted Nov. 94), see http://www.ich.org/

[2] CPMP/ICH/281/95 (Q2B) Note for Guidance on Validation of Analytical Procedures: Methodology (Step 4 -adopted December 96), see http://www.ich.org/

[3] CPMP/ICH/4106/00 (Q7A) Note for Guidance Good Manufacturing Practice for Active Pharmaceutical Ingredients (Step 5 – released for consultation July 2000), see http://www.ich.org/

6
Acceptance Criteria and Analytical Variability

Hermann Watzig

6.1
Introduction

Establishing specifications, i.e., a list of tests, references to analytical procedures, and appropriate acceptance criteria [1] is one of the most important aspects during pharmaceutical development. Conformance to the defined criteria and acceptance limits will verify – as one part of the total control strategy – that a drug substance or drug product is suitable for its intended use. The corresponding ICH Guideline Q6A [1] describes the general concepts in developing specifications and provides recommendations for some tests and acceptance criteria. Of course, product attributes which are critical to ensure safety and efficacy need to be addressed with primary importance. This is reflected in more detailed guidance, for example, on setting impurity and dissolution acceptance limits [1–8].

The objective of the analytical testing is to evaluate the quality of analytes (drug substances or products, intermediates, etc.). However, the analytical result will always also include the variability of the measurement process. Ideally, this analytical variability should be negligible compared with the variability of the tested product, but this is often not realistic. Therefore, both the analytical and the manufacturing variability need to be taken into consideration in the process of establishing acceptance criteria. Apart from general concepts of process capability [2] or for drug substances [3, 9], no specific guidance is available on how to achieve an appropriate consideration of the analytical variability in assay procedures. Therefore, a thorough discussion process was started by the Working Group Drug Quality Control / Pharmaceutical Analytics of the German Pharmaceutical Society (DPhG) with a workshop on analytical uncertainty and rational specification setting in Frankfurt, January 31, 2002. As a conclusion of the presentations and discussion, a consensus paper was prepared and accepted at the annual meeting of the Working Group in October 2002 in Berlin [10].

Method Validation in Pharmaceutical Analysis. A Guide to Best Practice. Joachim Ermer, John H. McB. Miller (Eds.)

ISBN: 3-527-31255-2

6.2 Analytical Variability

6.2.1 Uncertainty of the Uncertainty

With respect to content determination of active ingredients, the analytical variability often consumes a significant part of the overall specification range. In some cases, for example, the assay of synthetic drug substances, the analytical variability is actually dominating.

Usually, it is expressed as a standard deviation $\hat{\sigma}$. Often it is normalised with respect to the mean and reported as a (percentage) relative standard deviation RSD (or RSD%). If not otherwise specified, a standard deviation describes the distribution of single analytical results.

Assuming a normal distribution, about 67% of the whole data population can be expected within one standard deviation around the mean, two standard deviations will include about 95%, and $\pm 3\,\hat{\sigma}$ will include 99.7% of all data (Fig. 2.1-2).

However, these estimations are only valid for normal distributions with known $\hat{\sigma}$. This value is rarely known, thus the above given limits are only roughly valid for higher numbers of samples ($n \geq 20$). The uncertainty in the determination of $\hat{\sigma}$ is regarded using the Student *t*-distribution, for example, estimating the confidence limits of means cnf($\bar{x}$) (Eq. 6-1).

$$\mathrm{cnf}(\bar{x}) = \bar{x} \pm t_{n-1,\alpha/2}\ \hat{\sigma}\ \sqrt{\frac{1}{n}} \tag{6-1}$$

The location of the confidence limit is defined by $\bar{x}$, the uncertainty is regarded considering $\hat{\sigma}$, the number of measurements n and the selected error probability α. The true mean μ can now be estimated from a random sample; with the chosen error probability it is found within cnf($\bar{x}$).

The membership of a population of future values or measurements is more relevant, for example, if a sample is the same as previously analysed ones. The corresponding questions are answered by prediction intervals (Eq. 6-2):

$$\mathrm{prd}(x) = \bar{x} \pm t_{n-1,\alpha/2}\ \hat{\sigma}\ \sqrt{\frac{1}{n} + \frac{1}{m}} \tag{6-2}$$

The formula is identical to Eq. (6-1), except for the square-root term. Here two different numbers of measurements are considered. The value n denotes the number that was used to determine $\bar{x}$ and $\hat{\sigma}$, m is the number of measurements which were used to calculate the future value (mean); m often equals 1, then the prediction interval corresponds to a single value. Therefore the square-root term is typically much bigger compared with Eq. (6-1).

The width of both the confidence and the prediction interval strongly depends on n. For small data numbers, the *t*-values become very large. Hence, for $n = 2$ the prediction interval is about 15 times wider than just considering the normal distribution, for $n = 3$ it is still three times wider and for $n = 4$ it is still twice as wide.

The strong influence of uncertainty can already be understood from this comparison. How about the uncertainty of the uncertainty itself? Standard deviations are examined for statistical difference by the simple Eq. (6-3):

$$T_F = \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} \tag{6-3}$$

The test is simply carried out by dividing the corresponding variances (the squares of the standard deviations). Here the larger is always divided by the smaller one. The resulting test value T_F is compared with the corresponding value in a tabled *F*-distribution. Again the uncertainty is especially high for low numbers of data. Considering two series with three measurements each (i.e. n-1=2 degrees of freedom) and $\alpha = 0.1$, $F_{2,2,0.1} = 9.000$ is obtained. This means that the standard deviations become statistically different only when they are more than three-fold different! Standard deviations are much more uncertain than mean values. If the high uncertainty in the determination of standard deviations is not properly considered, this can lead to very problematic premature evaluation of measurement uncertainty.

Example: Pseudo-optimisation

When capillary electrophoresis was first introduced, which of the parameters would influence the precision of a method, was only poorly understood. In order to optimise the precision, several parameters were empirically varied and their relationship to the precision was noted. Besides other factors, the dependence on the buffer concentration was examined (Table 6-1, Fig. 6-1).

At first sight there seems to be an optimum for a buffer concentration between 42 and 46 mmol/L – however, the determined standard deviations are not different in terms of statistical significance. If the same buffer were simply used for all seven series, a similar result for the distribution of the standard deviations would be obtained (meanwhile it is well understood, that the precision in CE hardly depends on the buffer concentration).

It is easy to comprehend that it cannot be right just to select the best of seven and use the value obtained for further considerations. However, the possibility of pseudo-optimisations often appears in disguise. In order to improve the precision of a method, it is the correct strategy to vary parameters. If a lower standard deviation is thereby obtained, every analyst hopes to have improved the method. In order to avoid pseudo-optimisation, data from preliminary experiments with typically low numbers of data have to be confirmed with higher data numbers. In the given example, a RSD% of 1.1% would have been obtained for every seventh experiment, on average, in routine analysis.

Table 6-1 Pseudo optimisation. Standard deviations of peak areas in CE at various buffer concentrations ($n = 6$ measurements each).

c (mmol/L)	36	38	40	42	44	46	48	50
$\hat{\sigma}$	2.711	1.575	2.442	1.31	1.324	1.110	2.006	4.006

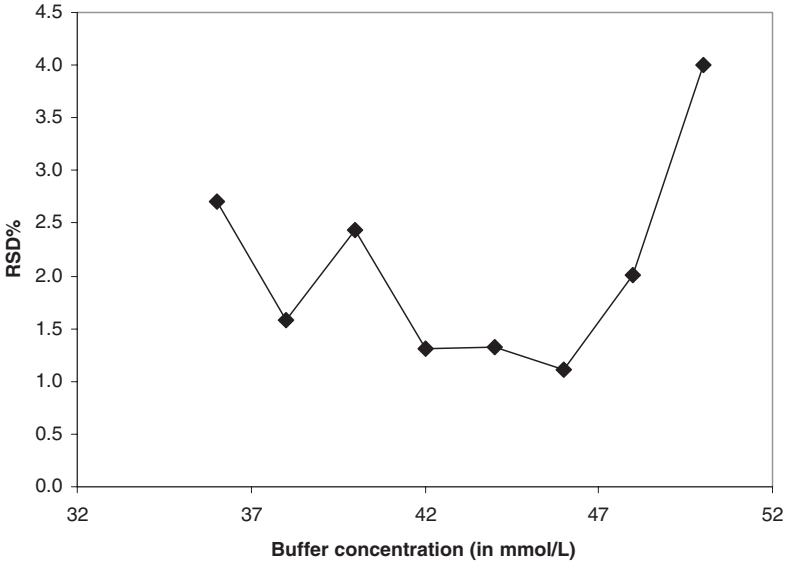


Figure 6-1 Pseudo relationship between buffer concentration and RSD% (see Table 6-1).

The confidence interval for the true standard deviation σ can also be given in an explicit equation (Eq. 2.1-7). Figure 2.1-4B shows the alarmingly huge confidence intervals for the true σ for low numbers of data. Table 6-2 shows which numbers are needed for satisfactory results. In order to avoid pseudo-optimisation, we suggest a minimum of 20 measurements to assess the decrease or increase in the standard deviation during method development. In critical cases, we often took $n = 60$ [11].

Table 6-2 Required degrees of freedom df (here: n–1), in order to guarantee a sufficient safety distance between a standard deviation $\hat{\sigma}$, which was estimated during method development, and a required limit value of 2% in this example.

$df = n - 1$	4	6	10	20	50	100	500
$\hat{\sigma}$ (%)[a]	1.43	1.5	1.58	1.67	1.77	1.83	1.9

a: calculation: $\hat{\sigma}(\%) = 2\frac{1}{\sqrt{F_{0.1,df,1000}}}$; the second degree of freedom is set to the very high value of $df = 1000$, the obtained value is very close to a comparison to an infinite population

Moreover, it is not only the statistical uncertainty of the determination of the standard deviations which has to be considered. If the standard deviation is used to define specification acceptance limits, not only the actual experimental results, but also all future assays have to conform. Therefore, the relevant standard deviation is the intermediate precision or reproducibility ([12]; see next section and Section 2.1.2.).

6.2.2 Estimating the Analytical Uncertainty

The previous section made clear how important it is to estimate uncertainty with a sufficient number of data. During method validation, the intermediate precision is determined for each analytical procedure. This value can be used as an estimate of the reproducibility and for the establishment or verification of acceptance limits, when all factors relevant for the future applications of the analytical procedures were properly taken into account. Often the question arises as to whether all relevant factors have been addressed and thus, how predictive precision data from method validation can be, for future long-term applications.

Obviously at the time of establishing the acceptance limits, only a limited amount of data are available, with limited reliability concerning the long-term variability. This lack of knowledge has to be considered in the process of establishing (or verifying) acceptance limits. Then again, there are two major options to estimate long-term precision, in order to confirm or falsify the results from intermediate precision studies: the use of the law of error propagation and the attempt to use experience from earlier long-term studies of similar analytical tasks.

6.2.2.1 Law of Error Propagation

This approach uses the Gaussian law of error propagation. According to this law, the total variance is just the sum of all individual variances, where applicable, weighted with the respective partial derivatives.

The simplest example for the use of the law of error propagation is the error calculation when using external standards. Here the error comes from the analysis of the analyte and the analysis of the standard. The total variance $\hat{\sigma}^2_{tot}$ is (Eq. 6-4):

$$\hat{\sigma}^2_{tot} = \hat{\sigma}^2_{ana} + \hat{\sigma}^2_{std} \tag{6-4}$$

Considering that the error is the same for the analyte and reference standard sample, it follows that the total variance is twice the analysis variance and hence

$$\hat{\sigma}_{tot} = \sqrt{2}\,\hat{\sigma}_{ana} \tag{6-5}$$

Therefore, the analytical error using an external standard is about 1.4 times higher than the standard deviation, from repeatedly analysing the same sample (repeatability).

The next example is slightly more complex. When preparing a standard solution, the concentration depends on the mass of the weighted amount m, the mass fraction w of the standard substance and the volume V of the measuring flask used (Eq. 6-6):

$$c = \frac{m\,w}{V} \tag{6-6}$$

All errors in these parameters affect the total error of the concentration. According to the general law of error propagation, variances behave additively, weighted with the partial derivatives of the respective error components. The variance of a parameter y, which is dependent on the parameters x_1 to x_n, is thus calculated using Eq. (6-7):

$$\hat{\sigma}_y^2 = \sum \hat{\sigma}_i^2 \left(\frac{\partial y}{\partial x_i}\right)^2 \tag{6-7}$$

This equation is a simplified description of the law, in certain cases additional covariance terms have to be considered. For quotients q, this general description of the law of error propagation is distinctly further simplified, because the partial derivatives are always the absolute values of the considered quotients, divided by the parameter of differentiation (Eq. 6-8):

$$\left|\frac{\partial q}{\partial x}\right| = \frac{q}{x} \tag{6-8}$$

For example, the partial derivative of the concentration to the weight is calculated using Eq. (6-9):

$$\frac{\partial c}{\partial m} = \frac{c}{m} = \frac{w}{V} \tag{6-9}$$

The general law of error propagation for the error in concentrations is thus simplified, after factoring out, to Eq. (6-10):

$$\hat{\sigma}_{tot}^2 = c^2 \left(\frac{\hat{\sigma}_m^2}{m^2} + \frac{\hat{\sigma}_w^2}{w^2} + \frac{\hat{\sigma}_V^2}{V^2}\right) \tag{6-10}$$

Inserting the values $m = 1.342$ g, $w = 0.98$, $V = 10$ mL and the respective standard deviations $\hat{\sigma}_m = 4.315 \times 10^{-3}$ g, $\hat{\sigma}_w = 1.155 \times 10^{-2}$ and $\hat{\sigma}_V = 2.462 \times 10^{-2}$ mL into Eq. (6-10) and taking the square-root results in a total error of $\hat{\sigma}_{tot} = 1.639$ mg/L.

The same approach can be used to estimate the influence of further error components on the total error of an analytical result. In an example considering the GC/ECD determination of a herbicide in urine, the main error source was the measuring process, the preparation of the reference standards was a negligible error source [13].

For a HPLC quantitation of the vitamins A and E in a paediatric pharmaceutical, the main reasons for the observed analytical uncertainty were random influences, derived from the total standard deviation after multiple determinations, and variations in recovery, determined by comparison with certified reference material.

Further significant sources of error were two steps of sample pre-treatment, namely a liquid–liquid extraction step and the consecutive evaporation in the rotavapor. All other contributions to the overall variance, such as weighting, variations in the mobile phase, etc., were only minor [14]. In a long-range concentration study in HPLC, sample preparation was identified as the major error source for higher sample concentrations, whereas integration was found to be the dominating factor at lower concentrations. The injection error was only the third most important [15].

More examples for the application of the law of error propagation in analytical science can be found on the Worldwide Web (see [13]). The estimation of analytical uncertainty from variance components can be laborious [16–18]. However, the effort can often be partly reduced by using suitable software packages [19–20].

Typically the contribution of smaller variance components such as weighting or dilution errors are well known, but unfortunately, these are only of minor importance. Because variances (and not standard deviations) behave additively, larger squared error components affect the total error much more strongly than do smaller ones. If the volume error in the above example is no longer considered, the overall error hardly changes at all (from 1.639 mg/L to 1.607 mg/L). However, the magnitudes of the critical variance components, e.g., from sampling or sample pre-treatment, are unknown. Sometimes an attempt is made to estimate these components, but this makes the conclusions so vague that one might as well simply estimate the total error itself.

6.2.2.2 Uncertainty Estimations from General Experience

There is another possibility of estimating uncertainty: the short-term analytical error (repeatability) can be individually determined. Then it can be extended from these shorter to longer terms, i.e., to intermediate precision and reproducibility. Generally four levels of precision are distinguished: system precision, repeatability, intermediate precision and reproducibility (see Section 2.1.2.). Over months and years the intermediate precision converges to the reproducibility, because over a long time the personnel and instrumentation within the same laboratory will also change. Therefore, in the following the term reproducibility ($\hat{\sigma}_R$, $RSD_R(\%)$) is used as general term, and also includes the intermediate precision.

Information about the system precision can easily be found, for example, in the corresponding instrument manual provided by the manufacturer. A compilation of this parameter for various techniques is given in Table 6-3 [21]. The European Pharmacopoeia does not give a fixed value for the system precision of chromatographic methods in general, but instead gives a range from 0.2 to 1.3% [22]. In a study about the precision of HPLC autosamplers, an average system precision of 0.8% was found, but in 5 out of 18 measurement series this parameter was above 1% [23]. Once more these ranges reflect the uncertainty of the uncertainty (Section 6.2.1.).

Additionally, a value for the system precision is not generally valid for one analytical technique such as HPLC, but also depends on the method. Side compounds in changing concentrations, which are not detected by themselves, can still contribute as 'chemical noise' to the analytical uncertainty. Further, even today inadequate inte-

Table 6-3 System performance (SST, corresponding to system precision), for various analytical techniques (modified from [21]).

Technique	Average RSD%	Highest and second highest RSD% observed	Number of measurements *n* used to calculate RSD%	Number of measurement series *N* investigated
HPLC, automated	0.3	0.5, 0.6	5	22
GC, direct inj.	0.7	0.9, 1.0	6	10
GC, headspace	1.1	1.8, 2.3	6	10
CE	0.7	1.1, 1.2	6	16
HPTLC[1)]	1.4–1.9	2.9, 2.9	8–16	20

1 Differences due to evaluation using peak areas or heights and due to the number of evaluated tracks per TLC plate

gration algorithms can cause substantially higher standard deviations, especially considering unfavourable baseline characteristics [24, 25].

System precision and repeatability are often described in original papers about quantitation, because the costs are limited to determining these values. However, typically, they depend on rather small numbers of data, hence they must be read with large confidence intervals (see Section 6.2.1 and Figure 2.1-4B). There is an elegant concept to estimate true repeatabilities σ_r as target standard deviations (TSDs) in inter-laboratory trials (Table 6-4) [26]. Considering the typical case with six repeated measurements, an acceptable repeatability must be less than twice the corresponding TSD (see Figure 2.1-4B, upper limit of the confidence interval for n=6). The concept of TSDs thus allows for quick estimations of analytical uncertainties. Therefore it can give direction for validations and acceptance criteria (see Section 6.3) [10, 27].

However, TSDs for many classes of methods have still to be established or verified. For example, it is not plausible that the TSD of UV spectrometry should be larger than the value of the combination LC within the same laboratory UV

Table 6-4 Target standard deviations (TSDs) from inter-laboratory trials [26]. The given TSDs are the geometric averages of repeatabilities.

Methods	Examples	TSD
Titrations		
– aqueous with alkali, color-indicator	Salicylic acid	0.2 %
– aqueous with alkali, potentiometric	Salicylic acid	0.3 %
– non-aqueous with alkali, color-indicator	Ephedrine hydrochloride, racemic	0.4 %
– non-aqueous with alkali, potentiometric	Ephedrine hydrochloride, racemic	0.4 %
UV spectrometry	Prednisolone acetate, etc.[1)]	0.6 %–1.3 %
Liquid chromatography (HPLC)	Cloxacillin sodium	0.6 %

1 Cinnarizine, dienestrol, albendazole and methylprednisolone hemisuccinate

(Table 6-4), although there is even an additional analytical error to be expected from sample injection and peak integration. It is still not clear whether classes of methods generally have the same, or at least a similar, TSD. At the present time TSDs are not available for all analytical scenarios. Their reliable determination requires a considerable experimental effort [26]. Due to the high costs of the necessary inter-laboratory trials, it is advisable to collaborate with partners who are interested in TSDs of the same classification.

There are only few publications concerning the long-term variability (intermediate precision/reproducibility) in pharmaceutical analyses. Therefore, a reliable estimation of the ranges of analytical variability or a generalisation is still difficult. There is just one investigation on HPLC precision data from stability studies, which nicely indicates typical ranges to be expected for repeatability and reproducibility (see Section 2.1.3.2.) [28, 29].

Thus it is often not easy to estimate reliably the order of magnitude of reproducibility in advance. Mainly the easier-to-obtain intermediate precision is hence used as an estimate. However, typically, the reproducibility is higher than the intermediate precision [21]. Especially in the early stage of a project it is nearly impossible to achieve information about the reproducibility of a method. Therefore there have been many attempts to roughly predict this parameter. It has been suggested that the standard deviation increases by a factor of 1.5 per precision level. If the system precision were 1% RSD%, the repeatability would be 1.5%, the intermediate precision 2.25% and the reproducibility 3.375% [21]. This rule of thumb is supported by some experience, but there is not enough supportive data material to generalise these factors. The conversion from system precision to repeatability by this factor is, however, very plausible. The law of error propagation suggests a factor of about 1.41 ($\sqrt{2}$, see first example in Section 6.2.1., Eq. 6-5), the factor of 1.5 can be caused by a small additional error during sample preparation (such as dilution). Other authors presume a factor of 1.7 between intermediate precision and reproducibility [30], or a factor of just 'approximately 2' was given between repeatability and reproducibility [31, 32].

Stability studies offer an interesting source which can be used to extract data about both repeatability and reproducibility (see Section 2.1.4.1) [28]. In a project which included 156 stability data sets of 44 different drug products from seven companies, the average repeatability ranged from 0.54 to 0.95%. The average reproducibility was between 0.99 and 1.58%; the upper values of these parameters were about 2.0% and 2.6%, respectively; (see Table 2.1-5). The data support an average factor of about 1.5 or less between these parameters. Considering the worst case, which would be the underestimation of the reproducibility obtained in the future, the upper limit of this factor was determined. It is approximately 2.5 for LC assays of formulations and 3.0 for drug substances. The level is probably higher for substances, because here the error contribution of the sample pre-treatment is only minor. Thus the variability of the reference standard over time, which is a significant error contribution to reproducibility, becomes more important.

As emphasised, the classification and generalisation of target standard deviations, error intervals and level factors, requires a substantial data base, because the uncer-

tainty of the measurement errors itself has always to be considered (see Section 6.2.1.). The level factors could well also depend on the analytical technique or on the class of methods. Considering these limitations, however, the level factors discussed differ surprisingly little. The estimation of 1.7 [30] per level corresponds to the upper expected factor for 2 levels ($1.7^2 \approx 3$) [28]. 'Approximately 2' for two levels [31, 32] is close to the average value found in [28]; the difference to a factor of 1.5 per level is insignificant ($1.5^2 = 2.25$) [21].

6.3 Acceptance Criteria

6.3.1 Assay of Drug Substances

The Working Group has basically proposed the concept of Daas and Miller [3], with some adjustments [10, 27]. In this concept, the mean (of usually three determinations) is the reportable result to be compared with the acceptance limits. If impurities are present and not included in the assay, asymmetrical limits are required:

$$\mathrm{LAL} = 100\% - \%\mathrm{TSI} - 3\,\mathrm{TSD} \quad \mathrm{UAL} = 100\% + 3\,\mathrm{TSD} \tag{6-11}$$

LAL and UAL:	lower and upper acceptance limit, respectively.
%TSI:	Total sum of impurities (for a selective assay)
TSD:	Target standard deviation from collaborative trials (as an estimate for the 'true' repeatability standard deviation). As an approximation, the pooled repeatability from several series can be used.

The terms (100%–%TSI) and (100%) correspond to the lower and upper basic limits of the synthesis process of a drug substance. The three-fold TSD describes the variability range as well as the long-term variability of the analytical procedure.

Alternatively, Eq.(6-12) can be used. The lower basic limit BL then corresponds to %TSI and the analytically required range is calculated from the specific prediction interval of the control test, instead of the general estimation with the three-fold target standard deviation.

6.3.2 Assay of Active Ingredients in Drug Products

For European submissions it is standard practice that the active ingredient in drug products should range between 95 and 105% of the declared content (release limits). These acceptance limits do not require an additional justification. This standard practice is reasonable and suitable in most cases. However, it is also clear that there

are some cases that require wider limits, with appropriate justification. Possible reasons for a higher variability not permitting the standard limits from 95 to 105% are:

- unavoidable high batch variability caused by the manufacturing process;
- very small analytical concentrations;
- complex matrix effects;
- unavoidable high variability of the analytical procedure.

For such cases, the Consensus Paper [10, 27] recommends the following approach.

The manufacturing variability is represented by basic limits (BL), the analytical variability is described as a prediction interval of the mean. This is a refinement of the original concept of Van de Vaart [9], who proposed using confidence intervals.

$$AL = 100\% \pm BL \pm \frac{t_{df,95\%}\ RSD_R(\%)}{\sqrt{n_{assay}}} \tag{6-12}$$

AL: Acceptance limits of the active ingredient (in percent of the label claim)

BL: Basic limits, maximum variation of the manufacturing process (in %). In case of shelf life limits, the lower basic limit will additionally include the maximum acceptable decrease in the content.

$RSD_R(\%)$: Reproducibility precision (relative standard deviation)

n_{assay}: Number of repeated, independent determinations in routine analyses (e.g., different initial weight, sample preparations, etc.), insofar as the mean is the reportable result, i.e., is compared to the acceptance limits. If each individual determination is defined as the reportable result, n=1 has to be used.

t_{df}: the t-factor for the degrees of freedom during determination of the reproducibility, correction factor for the reliability of the standard deviation.

This calculation has the advantage to include the reliability of the experimental analytical variability as well as the specific design of the control test. It also clearly demonstrates the interdependencies between acceptance limits, analytical variability, and number of determinations. The larger analytical variability can be counterbalanced by increasing the number of determinations, however, as no safety risk is involved, 'testing the statistics' is not justified and the Consensus Paper suggests not going beyond three determinations.

The reproducibility can be estimated from inter-laboratory trials (see Section 2.1.2.3), but often this data is already available from repeated series within one company over longer periods of time, for example, from stability testing (see Section 2.1.4.1, Table 2.1-6). In a recent study using this stability data approach with HPLC, the upper limit for the repeatability was estimated to 2.0% RSD, the reproducibility upper limit corresponded to approximately 2.6% RSD [28, 29]. This considerable, but not major, increase indicates that the most important error contributions are

already included in the short-term variability. According to a recent investigation, these major contributions are peak integration and sample pre-treatment [15].

At the time of submission, the basic limits are often not exactly known. However, the assumption that half of the acceptance range is consumed by the manufacturing process, should be realistic for standard processes. Consequently, the standard limits can be met with a relative (target, average) reproducibility of 1% using single determinations, or with a value of 1.7% with triplicates. If either the manufacturing or the analytical variability is much larger, Eq. (6-12) allows an estimation of suitable individual acceptance criteria. For example, it is well known that the relative standard deviation increases with decreasing concentration [15, 29, 33]. Due to the increasing analytical variability, wider acceptance criteria are required.

6.3.3 Dissolution Testing

Procedures are described in the pharmacopoeias, but statistically-derived criteria are not covered so far. However, the same equation as for formulations can be used (Eq. 6-12). Here the basic limit BL includes deviations in dosage, but also the amount which is not dissolved. The RSD_R (%) value includes the spread from the measurement and from the dissolution process. The latter is usually dominating; numbers of at least 10% have been given [34–36]. Therefore it should be sufficient to consider this dominating error contribution (see Section 6.2.1).

6.3.4 Stability Testing

The determination of acceptance criteria for stability testing is just another special case of the criteria determination for assays in general. Thus, Eq.(6-12) is also applicable here, with some minor modifications (Eq. 6-13):

$$AL = 100\% \pm BL - D \pm \frac{t_{df,95\%}\ RSD_R(\%)}{\sqrt{n_{assay}}} \qquad (6\text{-}13)$$

A number for the maximum acceptable decomposition/degradation D is included into this formula. Further, RSD_R(%) maybe slightly higher here compared with the value estimated for assay testing. An increase in chemical noise may be expected due to degradation products in some cases.

From this acceptance limit, the shelf time can be estimated by extrapolation of consecutively performed stability tests. The general approach is outlined in [37]. It is obvious, that the lower the analytical variability and the higher the number of data, the smaller the confidence interval and therefore the longer the estimated shelf time will be [38, 39]. Note that stability data measured at later times have a higher influence on the shelf time than measurements from the beginning of the stability test. In order to obtain higher numbers, it has often been suggested to pool stability data [37, 40]. This is very reasonable, sometimes even if the slopes of degradation are statistically significantly different. If the precision of the measurements is very good

and therefore the variability is very small, even a marginal difference between the slopes will become significant. It is sensible to define a relevant difference in slope and then just test whether the difference in slope complies. This suggestion is very similar to other equivalence tests (see Section 1.4.1).

The statistics of accelerated testing has not been frequently addressed so far [41, 42]. Sophisticated calculations may be rather unnecessary, because accelerated tests allow only rough estimations of the remaining content (± 3–4%) in any case[38].

6.3.5 Impurities

The recommended approach is based on the ICH guideline Q6A [1]. However, instead of using confidence intervals, which would reward usage of as few data as possible and penalise applicants with a large number of batches, the analytical variability is described by the standard deviation and an uncertainty factor.

$$\mathrm{AL} = \bar{x} + 3\,\hat{\sigma} \tag{6-14}$$

The obtained values should be rounded to one decimal place. The mean and standard deviation should be determined for at least five representative and, if possible, subsequent batches from clinical phases II and III [10, 27]. Of course, the limit thus obtained must be qualified toxicologically.

6.4 Conclusions

Provided that safety and efficacy requirements are fulfilled, 'a reasonable range of expected analytical and manufacturing variability should be considered' in the process of establishing acceptance limits in a drug substance or product specification [1]. The analytical uncertainty can be readily estimated from variance components, if all contributing components are known. Regrettably this is usually not the case. In particular, critical components are often unknown. Experience about the total uncertainty allows for rough estimations of the reproducibility that can be expected in the future. Factors to calculate between precision levels given by different authors agree surprisingly well. Thus they should be suitable to estimate the worst-case intermediate precision or reproducibility.

After a thorough discussion, the Working Group Drug Quality Control / Pharmaceutical Analytics of the German Pharmaceutical Society (DPhG) has published a consensus paper with specific proposals on how to take the analytical variability into account in the process of establishing acceptance criteria for assays of drug substances / drug products and for impurity determinations.

For assays, release limits of 95–105% can be applied as the standard approach, but their compatibility should be verified. It is recommended to calculate (statistically) acceptance limits only in cases where a larger analytical (or manufacturing)

variability can be justified. The approach for calculating acceptance criteria for assays can be generalised to dissolution and stability testing.

For the assay of drug substances, basically the concept of Daas and Miller [3] is proposed with some adjustments [10, 27]. Concerning impurity determinations, the ICH approach [1] was substantiated and then applied with some minor modifications.

Further efforts in this area are still desirable – there is still a lot of uncertainty about the uncertainty. Therefore additional insight into acceptance criteria can be expected.

Acknowledgements

I am grateful to Phillip Hasemann for his critical reading of the manuscript.

6.5 References

[1] International Conference on Harmonisation, Notes for Guidance, Q6A. http://www.ich.org/pdfICH/q6a.pdf or http://www.fda.gov/OHRMS/DOCKETS/98fr/122900d.pdf

[2] D. Shainin, P. D. Shainin, Statistical process control. In: Jurans's Quality Control Handbook. Eds. L. M. Juran, F. M. Gryna, McGraw-Hill, New York, 4th ed., 1988, sec. 24.

[3] A. G. J. Daas, J. H. McB. Miller: Relationship between content limits, system suitability for precision and acceptance/rejection criteria for assays using chromatographic methods. Pharmeuropa 11(4) (1999) 571–577.

[4] European Pharmacopoeia (Ph.Eur.), 4th Edition, Council of Europe, Strasbourg, Grundwerk (base volume), Deutscher Apotheker Verlag, Stuttgart (Ger), 2002, Ch. 2.9.3/2.9.4, pp. 240–245.

[5] E. Nürnberg, in: K. Hartke, H. Hartke, E. Mutschler, G. Rücker, M. Wichtl: Comments on the European Pharmacopoeia. Wiss. Verl.ges. Stuttgart (Ger), 10th suppl. 1999.

[6] United States Pharmacopoeia 26, The United States Pharmacopeial Convention, Rockville, MD, 2003, ⟨701⟩ Dissolution, pp. 2155.

[7] Japanese Pharmacopoeia XIV, The Japanese Ministry of Health, Labour and Welfare, 2001; 15. Dissolution Test, pp. 33. http://jpdb.nihs.go.jp/jp14e

[8] Guidance for Industry. Dissolution Testing of Immediate Release Solid Oral Dosage Forms. U.S. Department of Health and Human Services, FDA (CDER), August 1997; http://www.fda.gov/cder/guidance.htm

[9] F. J. van de Vaart. Content limits – setting and using, Pharmeuropa **9**(1) (1997) 139–143.

[10] H. Wätzig, J. Ermer, Positionspapier der Fachgruppe Arzneimittelkontrolle/Pharmazeutische Analytik der Deutschen Pharmazeutischen Gesellschaft zum Thema Festlegung von Akzeptanzkriterien. http://www.pharmchem.tu-bs.de/dphg_pospapier.pdf

[11] A. Kunkel, M. Degenhardt, B. Schirm, H. Wätzig: Quantitative capillary electrophoresis –performance of instruments, aspects of methodology and validation: an update. J. Chromatogr. A 768 (1997) 17–27.

[12] International Conference on Harmonisation, Notes for Guidance, Q2A. http://www.ich.org/pdfICH/q2a.pdf

[13] G. O'Donnell: The Estimation of Uncertainty for the Analysis of 2,4-D Herbicide in Urine by GC/ECD, www.measurementuncertainty.org/mu/examples/index.html; menu item: GC, file name: 24D%20in%20Urine.pdf
[14] Valid Analytical Measurements, VAM Project 3.2.1: Development and Harmonisation of Measurement Uncertainty Principles, Part (d): Protocol for uncertainty evaluation from validation data, January 2000, www.vam.org.uk/publications/publications_item.asp?intPublicationID=315; File name: LGC/VAM/1998/088
[15] U. Schepers, J. Ermer, L. Preu, H. Wätzig: Wide concentration range investigation of recovery, precision and error structure in high performance liquid chromatography. J. Chromatogr. B, 810 (2004) 111–118.
[16] Eurachem Guide: Quantifying Uncertainty in Analytical Measurement, 2nd Edition (2000), www.eurachem.ul.pt/index.htm
[17] www.measurementuncertainty.org
[18] American Association of Laboratory Accreditation (A2LA), www.a2la2.net
[19] M. Weber, Eidgenössische Materialprüfungs- und Forschungsanstalt (EMPA), St. Gallen, Unsicherheit über die Unsicherheit, Beitrag zum Workshop der Fachgruppe Arzneimittelkontrolle/Pharmazeutische Analytik der DPhG am 31.01.02 in Frankfurt/Main.
[20] www.uncertaintymanager.com
[21] B. Renger: System performance and variability of chromatographic techniques used in pharmaceutical quality control. J. Chromatogr. B 745 (2000) 167.
[22] Ph. Eur./DAB (German Pharmacopoeia) Official Edition. Deutscher Apotheker Verlag, Stuttgart, Suppl. 2001; Methods: 2.2.46 Chromatographische Trennmethoden.
[23] S. Küppers, B. Renger, V. R. Meyer: Autosamplers – a Major Uncertainty Factor in HPLC Analysis Precision. LC-GC Eur. 13 (2000) 114–118.
[24] B. Schirm, H. Wätzig: Peak recognition imitating the human judgement. Chromatographia 48 (1998) 331–346.
[25] Bernhard Schirm: Neue Möglichkeiten der Datenauswertung in der Kapillarelektrophorese durch Einsatz chemometrischer Verfahren. PhD thesis Würzburg (Ger) 2000.
[26] A.G.J. Daas, J.H. McB. Miller: Relationship Between Content Limits, System Suitability for Precision and Acceptance/Rejection Criteria for Assays Using Chromatographic Methods. Pharmeuropa 10(1) (1998) 137–146.
[27] H. Wätzig, J. Ermer, Positionspapier der Deutschen Pharmazeutischen Gesellschaft zum Thema Spezifikationssetzung. Pharmazie in unserer Zeit, 32(3) (2003) 254–256.
[28] J. Ermer, H. Wätzig, C. Arth, P. De Raeve, D. Dill, H.D. Friedel, H. Höwer-Fritzen, G. Kleinschmidt, G. Köller, M. Maegerlein: Precision from drug stability studies. Collaborative investigation of long-term repeatability and reproducibility of HPLC assay procedures. J. Chromatogr. A (in press)
[29] U. Schepers, J. Ermer, L. Preu, H. Wätzig: Präzision in der HPLC – Akzeptanzkriterien, Abhängigkeit von der Konzentration, Varianzkomponenten. http://www.pharmchem.tu-bs.de/dphg_waetzig.pdf
[30] T. Layloff: Do You Believe in Statistics? American Genomic Proteomic Technology, Jan/Feb 2002, 14– 7.
[31] W. Horwitz: The variability of AOAC methods of Analysis as used in analytical pharmaceutical chemistry. JAOAC 60, (1977) 1355–63.
[32] Klinkner, R.: Wie genau ist ein Analysenergebnis? Nachr. Chem. 50 (2002) 1302–04.
[33] R. Albert, W. Horwitz: A Heuristic Derivation of the Horwitz Curve. Anal. Chem. 69 (1997) 789–790.

[34] S.A. Qureshi, I.J. McGilveray: Typical variability in drug dissolution testing: study with USP and FDA calibrator tablets and a marketed drug (glibenclamide) product, European J. Pharmaceut. Sci. 7 (3) (1999) 249–258.
[35] M. Siewert, L. Weinandy, D. Whiteman, C. Judkins: Typical variability and evaluation of sources of variability in drug dissolution testing. Eur. J. of Pharmaceutics and Biopharmaceutics 53 (2002) 9–14.
[36] S. Furlanetto, F. Maestrelli, S. Orlandini, S. Pinzauti, P. Mura: Optimization of dissolution test precision for a ketoprofen oral extended-release product. Journal of Pharmaceutical and Biomedical Analysis 32 (2003) 159–165.
[37] International Conference on Harmonisation, Notes for Guidance, Q1A; http://www.fda.gov/cder/guidance/4282fnl.htm
[38] W. Grimm, G. Schepky: Stabilitätsprüfung in der Pharmazie, Editio Cantor, Aulendorf 1980.
[39] M. Holz, D. Dill, C. Kesselheim, T. Wember: Auswertung und Optimierung von Haltbarkeitsuntersuchungen bei pharmazeutischen Produkten, Pharm. Ind. 64 (12) (2002) 1279–1286.
[40] S. J. Ruberg, J. W. Stegeman: Pooling data for stability studies: testing the equality of batch degradation slopes. Biometrics 47 (1991) 1059–1069.
[41] C. R. Buncher, J.-Y. Tsay: Statistics in the Pharmaceutical Industry. Marcel Dekker, N. Y., 2nd ed. 1994.
[42] S.-Y. P. King, M.-S. Kung, H.-L. Fung: Statistical prediction of drug stability based on nonlinear parameter estimation. J. Pharm. Sci. 73 (1984) 657–662.

7
Transfer of Analytical Procedures

Mark Broughton and Joachim Ermer (Section 7.3)

7.1
Overview

Formal transfer of analytical technology became an issue for the pharmaceutical industry during the early 1990s. At that time several regulatory authorities including the US Food and Drug Administration (FDA) and the Medicines and Healthcare Products Regulatory Agency (MHRA) were concerned that the standards being applied to the transfer of methodology for the movement of new products from Research and Development to the receiving site, were inadequate. There was a perception that the industry was not carrying this out well, generally it was seen to be undertaken too hastily, with inadequate resources, at a point just before or probably coincident with qualification of the manufacturing process and the building of launch stocks. The industry responded to this demand, either due to the realisation that this was a real business issue or an awareness that this was becoming an area of focus for regulatory bodies in audits and inspections. Most laboratories responded by introducing 'collaborative' or 'cross-over' studies to support these transfers with acceptance criteria based on statistical tests such as 'f' and 't' tests. This approach to the transfer often created difficulties, such as failure to meet acceptance criteria, difficulties with the analytical method, inadequate training , lack of availability of the required materials or insufficient equipment in the receiving laboratory. Typical reports on analytical transfers carried out in this way are often filled with explanations of why pre-determined acceptance criteria were not met and explanations of why receiving laboratories should commence testing and releasing production material despite these failures. Over the past decade there has been recognition that the transfer of this knowledge, and that development of confidence in the technology is a key business process that is a foundation of the validation and long-term support of processes and justifies an appropriate level of investment. This has lead to an appreciation that a robust transfer of this knowledge requires a sound working relationship between the originating and receiving laboratories that allow issues to be highlighted and resolved before they can have an impact on business performance. This chapter describes a process that can be applied to transfers of the methodology for testing new products or to the transfer of existing technology to different laboratories, such as contract laboratories or receiving laboratories supporting second or alternative manufacturing sites.

Method Validation in Pharmaceutical Analysis. A Guide to Best Practice. Joachim Ermer, John H. McB. Miller (Eds.)

ISBN: 3-527-31255-2

7.1.1 Transfer Process

The process of transferring an analytical method has been broken down into five key steps that allow a thorough evaluation of the analytical procedures, current validation status and readiness of the receiving laboratory, before comparative studies takes place (Fig. 7-1). The issue of validation status should not normally be an issue in the transfer of modern analytical technology for new products from Research and Development to QC. It is the author's experience, however, that when mature products are transferred there can be significant gaps in the available validation documentation. It is likely that the process of transferring an existing, possibly well established, manufacturing process to a new site of manufacture will require that regulatory authorities are informed, this will probably prompt a review of the product file and there will be an expectation that the package will meet current standards. The point at which these methods are transferred seems to be an appropriate time to identify these issues and address them before regulatory questions arise.

The results of this process are improved confidence in the analytical methods and the data generated by them on completion of technology transfer.

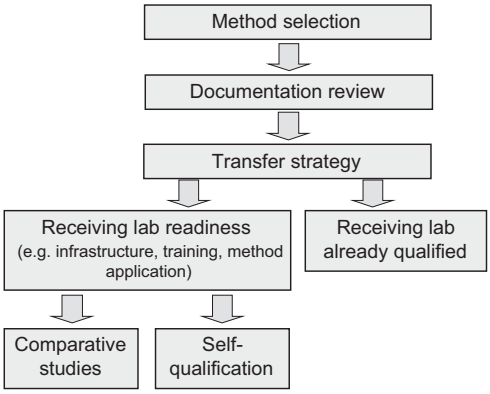


Figure 7-1 A five-step process for analytical transfer.

This process can be applied to the transfer of any analytical test procedure whether chemical or physical, from development to Quality Control, or between Quality Control laboratories at different sites, or to or from a contract laboratory. It is equally applicable to drug substances, formulated products and supporting methodology, such as that used for monitoring cleaning procedures or testing incoming materials and components.

The key steps allow the process to be flexible, modified to reflect specific needs, and well documented. Some of the documentation will be needed to support regulatory requirements. However, the authors believe that there is a business benefit in processes being well documented. This is because it is reasonable for the business to expect that the specification and methodology will support the control of the product over its life, which for a successful product might be ten, twenty or more years.

Quality or other manufacturing issues might be expected to arise during this time and these events will cast doubt or questions over the durability or capability of the analytical methodology, which might be related to manufacturing problems, customer complaints, process changes, raw material changes or genuine analytical issues. In addition, a well-organised laboratory will be undertaking periodic reviews of the methodology and updating technology where appropriate. This process is easier when there is a well-documented history. During this period the link to the methodology developed and validated by R and D and used to release the materials which demonstrate the safety and efficacy of the product, is of enormous value in demonstrating the consistency of the product or highlighting potential issues.

The timing and emphasis the receiving laboratory needs to apply to the various stages of this transfer process, for example, training or laboratory trials, will vary tremendously depending on factors such as the existing level of expertise in the laboratory or the level of support available from the originating laboratory. The amount of time and effort expended on each stage should be adjusted according to the need.

When transfer is being carried out from a development to a receiving laboratory, in order to support the transfer or establishment of the new product's manufacturing process it is important that any evaluation of the methods takes place early enough in the development process to allow any proposed changes to be implemented. Typically, methods and specifications are clearly defined and difficult to change after the Phase III clinical trials and stability studies have started. This may seem an early point in the development process to ask for involvement from the final receiving laboratory, because this could be 18 months before filing a regulatory submission. Usually, however, this is the last opportunity in the development process where significant method changes can be made without major changes to the regulatory submission, and attempts to change the methods after this time should be expected to meet with quite understandable resistance from the development function because of the impact this can have on these key studies.

7.2 Process Description

This section describes the main considerations for each step of the process, it is important to be aware that each of these stages is important in achieving a robust transfer. Each situation must, however, be considered on its own merits, the justification and extent of each stage should also be considered. There will be situations where some stages of a transfer might be extremely short or even absent and, where this is the case, the justification or explanation for this should be documented.

7.2.1 Method Selection

The success or failure of an analytical methods transfer and, for that matter, the continued success of the method in regular use in a Quality Control laboratory can

often depend on the way that the methods are selected and developed. It is important than the routine release tests and specifications that are applied, are carefully considered to give the required level of control. The specification acceptance criteria must obviously reflect any external requirements (compendia or regulatory requirements, for example) and those derived from local expertise or internal requirements. The outcome of these considerations should be a specification that includes the minimum number of tests to ensure a compliant, safe and effective product. Other tests are often introduced to provide additional information to support the development of a product, these must not be allowed to form part of the finished product specification. There are situations where these tests might be needed in QC to support process scale-up or validation, etc. In these situations, transfer clearly needs to take place but the temptation to include these in the product specification must be avoided. If such tests are allowed to become routine in a receiving laboratory, this has the effect of wasting resources and raises the possibility of delays and problems during transfer and during the life of the product in the receiving laboratory, due to the additional technical complexity.

The technique to be used for a particular test must also be chosen with care, and there should be an understanding of the technology available at the receiving laboratory as well as the capability to carry out the technique. This will simplify subsequent transfer and ensure that the expertise to maintain the methods is available throughout the life of the product. This should not prevent the introduction of new technology when there is a clear business benefit in terms of productivity, reliability, sensitivity, etc., but introduction of new techniques should be done pro-actively with an accompanying programme that establishes this knowledge in receiving sites with the appropriate skill-base.

The capability of the technique should also be considered, and it is important that the technique is capable of achieving a level of precision that is sufficiently small when compared with the specification width (see Chapter 6). This is essential to reduce the probability that out-of-specification results will occur, due to normal random variation, to an acceptably low level. This might mean that methods that have been considered suitable in the past are no longer appropriate. If we consider, for example, a typical bulk pharmaceutical chemical with a specification of 98–102% with a target of 100%, it is unlikely that a method with a standard deviation of 2% will be appropriate (see Section 1.4.1).

Generally, most pharmaceutical QC laboratories have expertise in HPLC and some, though typically smaller (in terms of quantity), have expertise in GC. It is therefore normally desirable that the chromatographic technique of choice, in this situation, should be reverse- phase HPLC, preferably isocratic, with the simplest eluent consistent with achieving the desired separation. Whilst gas chromatography may offer some benefits in terms of reliability and elegance, it is unlikely that these will outweigh the long-term costs caused by the reduced pool of expertise in such a receiving laboratory.

Clearly there is a regulatory expectation that applicable pharmacopoeia tests will be used wherever appropriate and there are benefits in terms of transfer in using these. It is, however, dangerous to assume that these do not need validation or trans-

fer. Such methods are usually based on well-established technology and are in use in a large number of laboratories. This reduces some of the issues associated with technology transfer; however, these methods should still be verified in the laboratory against a defined protocol that ensures that they operate in a controlled manner with an adequate level of precision and accuracy.

7.2.2 Early Review of the Analytical Procedure

A common complaint from QC laboratories (and receiving laboratories in general) is that they are expected to accept analytical methods that they have had little or no opportunity to review and they often feel that methods are inappropriate for frequent long-term use. They also often believe that many of these problems would not arise if they had been given the opportunity to review them. The expectation that receiving laboratories might have an input into the development of a new method does seem reasonable and some companies try to make this happen, but in reality this is can be difficult to achieve. In a typical transfer from Rand D to QC for a new product, the receiving laboratory very often has its first experience of the methodology several weeks before product launch. A request to change a specification or test method at a time that is close to product launch is unlikely to be realistic for the reasons discussed above. In addition to this, the manufacturing strategy is usually evolving alongside the product and there are situations where the manufacturing facility is not selected until late in the development of the product. This can effectively deny the QC laboratory any meaningful impact on the methodology and specification. There are many occasions, however, where the target QC laboratory is clearly known throughout the development process and in this situation it seems reasonable to give the QC laboratory an opportunity to review and comment on the methods. This should be done approximately 18 months before the planned filing date. The R and D function is then in a position to respond, either by making changes when it is practicable and appropriate, or by justifying the status quo.

Such a review can take place in two ways. There are situations where technology is well established in a receiving laboratory. This might be the case if the product is a line extension or a variant of an existing or similar product using very similar methodology. It might then be appropriate to limit the review to a documentation package consisting of the methods and available validation data. It must be accepted that, at this point in a product's development, this documentation will not be up to the standards of that prepared for a regulatory submission. However, there is still a great deal of value in this review.

There will be other occasions where the technique, or some feature of it, might be new to the laboratory; there may be issues about the clarity of the written material or the perceived capability of the methodology. In these situations it may be appropriate to extend the scope of the validation programme to include a more robust determination of the analytical variability (see Section 2.1.2).

7.2.3 Transfer Strategy

The transfer strategy can be summarised in a Strategy Document and should include all analytical procedures to be transferred and all actions needed to complete their successful transfer.

7.2.3.1 Transfer Category

Often, transfer is used synonymously for comparative studies. However, the receiving laboratory must demonstrate the ability to perform *all* analytical procedures detailed in a product's specification reliably and accurately. Of course, not each procedure requires comparative studies. For example, if the receiving laboratory is already running pharmacopoeial tests on a routine basis, no further activities are required. In the case of pharmacopoeial monographs, the general analytical technique is known and the receiving site can verify its ability by self-qualification, for example, successful System Suitability testing described in the respective monograph. In other cases, the originating laboratory may be not able to participate in experimental studies and the receiving site has to measure its performance versus a reference, such as a certificate or validation results. However, this should clearly be an exception or last resort, because in such an approach, the transfer of the knowledge which the donor site has gained, which is a major objective in a transfer, will be lacking. Alternatively, a re-validation may be performed by the receiving site.

In order to ensure a complete review, a full list of all procedures to be transferred should be prepared, with classification of the respective transfer activities, i.e., no further activities required, self-qualification, revalidation, and comparative studies.

7.2.3.2 Training Requirements

A listing of planned training with a brief justification should be prepared. This might be an extensive training programme, if a new technology is included in the transfer, or a very simple training exercise where the technology is very similar to that already in place.

7.2.3.3 Application of the Analytical Procedures

In the case of a transfer with self-qualification or comparative study without prior formal training, an application of the analytical procedure by the receiving site, based on the available documentation, should be performed. This also provides an opportunity to evaluate the quality of the documentation and instructions that support the methodology.

7.2.3.4 Comparative Studies

This section should contain a brief explanation of the comparative studies that will be carried out to support the transfer.

The protocol is required for the comparative studies that should include a timetable and resource estimates for both receiving and originating laboratories. This is essential to allow the laboratories involved to ensure that the required resources are

available. This document should be prepared after discussions between the originating and receiving laboratories relating to the outcome of the Laboratory Readiness activities and should be jointly approved.

7.2.4 Receiving Laboratory Readiness

This step ensures that the receiving laboratory is either fully prepared to accept the new technology or that the measures required to achieve this state are clearly understood. On occasion comparative studies are delayed because the receiving laboratory is incapable, due to such factors as lack of expertise or the equipment necessary to carry out the analytical work. The author has experienced situations where the R and D laboratory is forced to prevent release of a new product after product launch until the correct level of capability is achieved. This wastes valuable Rand D resource, delays the transfer, is likely to become an issue during a regulatory inspection and complicates the release processes in the quality organisation. This situation can be avoided if a thorough evaluation of the receiving site's capability takes place in time to allow any remedial action to be taken. This might include purchase of new equipment or recruitment and training of staff; and could take between 12 and 18 months if budgetary provision needs to be made.

In most organisations, a receiving laboratory is responsible for ensuring that it is prepared to accept new technology by either demonstrating that it has the equipment and expertise to do the work, or by purchasing the new equipment and recruiting the staff and expertise that are required. Whilst it might be reasonable to expect a laboratory that is part of the same organisation to do this well, the level of effort and support that is dedicated from the receiving laboratory should be adjusted in the light of knowledge and previous experience. Additional special care should be taken during transfers to laboratories where there is less experience in transfer, or in the technology that is the subject of transfer. Obviously this preparation cannot happen until there is a clear picture of the requirements. It is therefore important that an appropriate analytical documentation package (Analytical Package) is prepared and made available to the receiving laboratory. This needs to be done as soon as the methods and technology are stable. It must include an inventory of equipment, instrumentation and related expertise required to support the new methods (Technology Inventory). At this point the methods and specifications should be converted into the format of the receiving laboratory. These formatted methods should be reviewed and checked for consistency with the originals, and if this requires translation into a second language, this should include review by a suitable bilingual expert.

With this package it is now possible for the receiving laboratory to compare its available expertise and equipment against the Technology Inventory and carry out an analysis of any gaps and then develop a plan to address these. This should also include a review of the supporting systems in the laboratory, such as maintenance, calibration and training arrangements.

The following steps might occur during this evaluation.

7.2.4.1 Equipment Identification, Maintenance, and Laboratory Infrastructure

Equipment listed in the Technology Inventory should be identified in the receiving laboratory and records made of brand, model, status (de-commissioned/out of use/operational/in regular use etc.) and age. Any item that is missing or considered unsuitable for the application should be highlighted and a plan put in place to address this with a clear time-scale that should include installation, repair, qualification and training as required.

Equipment must be part of a maintenance and calibration programme that ensures that it performs consistently with the user requirement specification.

Verifying the laboratory infrastructure could include such items as water supplies, gas supplies, standards of chemist training, glassware washing and data handling systems to confirm suitability.

7.2.4.2 Consumables and Reference Standards

These must be available in the laboratory or a suitable source should be identified. In situations where the source is different from that used by the originating laboratory, the reliability and consistency of materials should be verified. This is particularly important if the receiving laboratory is in a different geographical region. There are still examples of what are often perceived as global suppliers of laboratory consumables, having similar descriptions for different items in different regions of the world.

7.2.4.3 Local Procedures

Any inconsistencies between procedures in place at the originating and receiving laboratories should be identified and reviewed for potential impact on the transfer and long-term performance of the methods. There can be interactions between these that are easier to manage if they are considered beforehand. For example, if the approach to be taken, in the event of an out-of-specification result being generated during a transfer process, is not clearly defined, it can become a significant issue, especially if the batches of product in question are also the subject of stability or clinical studies.

7.2.4.4 Receiving Laboratory Training

Ideally this can be achieved by an analyst or analysts from the originating laboratory visiting the receiving laboratory to carry out training. The extent and nature of this training should be adjusted to reflect the skill increase that is required. This function belongs to the Readiness Stage and is defined in the Strategy Document. For example, the level of training and preparation required to establish tablet-dissolution testing, in a laboratory that is not familiar with solid-dose testing, would be quite different from that required in a laboratory where similar products are already being analysed. In situations like this, where products and tests are very similar, the training requirement is usually very small. However, a decision to eliminate training completely should only be taken after very careful review.

There will be occasions where the originating and receiving laboratories are long distances apart, possibly in different countries or different continents. In these situ-

ations, consideration can be given to the use of other approaches to this, such as the use of video or videoconference, training by transfer of written training documentation, etc. In the author's experience, however, face-to-face training is by far the most reliable approach where there is a significant training need. If it is well planned and carried out according to a clear protocol with trainee performance criteria, it is usually the most cost- effective in the long term.

Training should be carried out against the pre-agreed training protocol with clear success criteria for the training; this should be documented according to normal company procedures.

There will also be occasions where it might be appropriate for an analyst from the originating laboratory to visit the receiving laboratory and observe the test being carried out. This also provides an opportunity to evaluate the quality of the documentation and instructions that support the methodology.

7.2.4.5 **Application of Analytical Procedures**

This is either an optional step in the process, which can be used to reduce the originating laboratory's input, or is done for the self-qualification of the receiving laboratory. It allows the precision of the methodology in the receiving lab. to be assessed and compared with validation data or other historical precision data, without requiring significant additional effort from the donor laboratory. It has been the author's experience that simple comparative studies often fail to generate confidence in the methodology. The intention of this phase of the transfer process is that the level of precision generated during development and validation of the method can be repeated in the new environment. Comparative studies often fail to do this because of the resources required in the originating laboratories and an understandable resistance to repeating work. This step should take place early in the transfer process to allow meaningful remedial action to take place and ensure successful comparative studies.

Again, the extent of these trials will vary depending on the level of expertise and experience within the receiving laboratory. It might also be dependent on experience with the method in the laboratories. The justification for the level of detail in this part of the process will again be found in the strategy document. These trials should be thoroughly planned, for example, by means of a protocol describing the objective of the trials, the methods to be performed, the experimental design for the study, the samples to be used and the responsibilities of those involved. It may also describe the acceptance criteria for success of the studies. This is an opportunity to investigate or verify that the effects of key sources of variation confirm the effectiveness of training. Review of the analytical package might have highlighted particular parameters that make a significant contribution to the overall variability of the methods; these parameters should be studied using an appropriate experimental design in the receiving laboratory to demonstrate that they are under appropriate control.

7.2.4.6 **Readiness Report**

At the end of this step, a brief report should be prepared and be issued before training takes place. This should describe the results of the assessments and identify any

outstanding issues and action plans to address them, and should also include any safety issues that have been apparent during the assessment.

7.2.5 Self-qualification

If the receiving laboratory is already experienced in general tests, in the case of pharmacopoeial monographs, or if the originating laboratory is not able to participate in experimental studies, then the receiving site can/must verify its ability to apply the analytical procedure by self-qualification. This is either combined with the application of the method as described in Section 7.2.4.5, or performed afterwards. In both cases, a formal protocol is required describing the design of the experimental investigation, the number of determinations, and especially, acceptance criteria. These may consists of System Suitability Test criteria of the respective pharmacopoeial monograph, or precision and accuracy criteria discussed in Section 7.3 in case of comparison to reference results (for example from certificates) or validation data.

7.2.6 Comparative Studies

This is the confirmation that the receiving laboratory is capable of performing the methods with a satisfactory level of accuracy and precision when compared with the originating laboratory.

The extent and level of detail in the comparative study will have been described in the Strategy Document. The detailed requirements must be defined, along with acceptance criteria, in an agreed protocol. Both the originating and receiving laboratories must approve this. The design of the comparative study, such as the number of batches of product used, the number of determinations, and the acceptance criteria, will vary depending on the type of control test and/or product (see Section 7.3).

On completion of these studies, the data should be jointly evaluated by the participating laboratories and assessed against the acceptance criteria in the protocol. If the capabilities of the methods and laboratories have been well understood and the acceptance criteria wisely chosen, this will normally result in a successful outcome. This can then be documented to allow the receiving laboratory to start testing in earnest; the results should be approved jointly by the receiving and originating laboratories. Situations where the resulting data do not satisfy acceptance criteria must be investigated carefully to reach a clear understanding of the reasons. Such a situation could be caused by issues related to the materials used for the comparative studies, capability of the receiving laboratory, or capability of the methodology. In such situations additional training or other corrective actions and a (partial) repeat of the comparative study might be required. It is also possible that acceptance criteria have been chosen inappropriately. However, this conclusion must only be reached after careful consideration of other potential causes as occurrences of this situation should be very rare.

7.3
Comparative Studies

7.3.1
General Design and Acceptance Criteria

In principle, the results of comparative studies can be evaluated by three general approaches:

- a simple comparison;
- statistical significance (or difference) tests;
- statistical equivalence tests.

In addition, any established system suitability test (see also Section 2.8) should be passed for all series performed [1].

The question of the suitable number of batches included in the comparative study is often debated. It ranges from general recommendations ("*as many batches as necessary ... to be considered representative...*" [2], or the 'magic' number of three batches [1] to one representative batch. However, we should have the objective of a transfer in mind, and that is the analytical procedure, and not the sample. This is a strong argument in favour of concentrating on one batch and instead increasing the number of determinations and thus the *result reliability*. Of course, the batch must be representative in order to make general conclusions. Therefore, in the case of impurity testing, it may be necessary to use several batches to cover the whole impurity profile, or (at least) all specified impurities. On the other hand, for assay, the specification range is usually rather small and there should be no problem in finding a representative batch (preferably at the process target). Any analysis at the limits of the working range of the analytical procedure must already be addressed in the validation and is therefore not the target in transfer studies.

In order to address all contributions properly, each of the series performed must be independent, i.e., should include the whole analytical procedure such as reference standard preparation and/or instrument calibration, sample preparation, etc.

In the following sections, the design and acceptance criteria are described or summarised for major types of analytical procedures. For further details on other types, such as identification, automated methods, cleaning verification, dose delivery, or particle size, the reader is referred to the ISPE-Guide [1].

7.3.1.1 Simple Comparison

Here, the results of the comparative investigations are compared with absolute acceptance limits, defined from experience [1, 4–6] or derived from statistical considerations (see Section 7.3.2). It was argued that, in this approach, neither α-, nor β-errors were controlled [2, 3]. (The former indicate the probability that an acceptable result will fail the acceptance criteria, the latter that an intrinsically unacceptable performance is not recognised.) However, this is only partly true. Although these risks cannot be defined numerically, they can be taken into consideration by the design of the study and of the acceptance criteria (see also Fig. 7-3). For example,

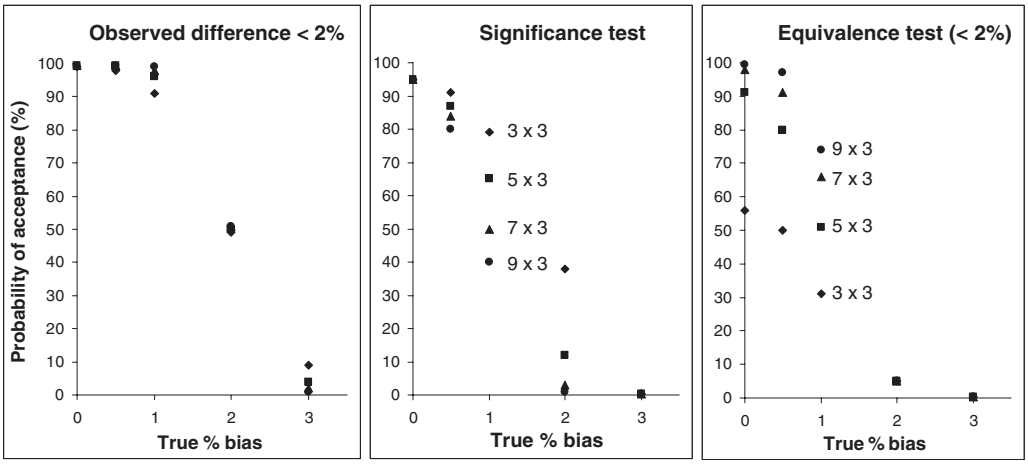


Figure 7-2 Probability of concluding that a relative bias is acceptable in dependence on the true bias and the number of determinations for the three comparison approaches. Three replicates each are performed on three (diamonds), five (squares), seven (triangles), and nine (circle) days. The simulations are based on a true repeatability and intermediate precision of 0.5 and 1.0%, respectively (data obtained from [3]).

the especially important ß-error (consumer's risk) can be taken into account by limiting both precision and accuracy of the data. The former is addressed by the repeatability (or intermediate precision) standard deviation, the latter by the difference between the means of the performed series. Limiting the precision by an absolute upper acceptance criterion will avoid the problem that an experimentally small difference in the mean results is only obtained by chance due to a large variability, whereas in fact the true difference is (unacceptably) large. At the same time, defining only a (practically relevant) upper limit avoids sensitivity to small experimental variability in some series, which is of no practical risk, but may lead to failing statistical significance tests, such as t-tests (see Section 7.3.2.2). In contrast to equivalence tests (see Section 7.3.2.3), where the variability of the analyses needs also to be included in the defined acceptable difference, in a simple comparison it is not explicitly taken into consideration, only as the (smaller) variability of the means. In the example shown in Figure 7-2, limiting the observed difference to < 1% instead of < 2% would result in similar failure rates (for a true bias of 2%) as in the figure, for a true bias of 3%. This evaluation corresponds to an acceptable difference of 2% in the equivalence test.

7.3.1.2 Statistical Significance or Difference Tests

This traditional approach of comparing precision and accuracy between two series of data by means of F-test and t-test, respectively, (see Section 2.3.1) or – in case of more than two series – with an analysis of variances (see Section 2.1.2.3) assumes that there is no difference between the two series, or to reference values [2]. Due to the tightening of confidence intervals with increasing number of determinations,

this approach results in scientifically illogical conclusions [3, 7], as shown in Figure 7-2. If there is a small, but acceptable bias, the chance of passing decreases with increasing sample size, from 79% with three series to 40% with nine series, i.e., with an increase in reliability (see Fig. 7-2, true bias of 1.0%)! Conversely, the smaller the number of determinations and the higher the variability (of both or all series), the higher is the chance of passing. Of course, this may be avoided by the design of the comparative study, i.e., by defining a suitable number of determinations and by limiting the variability. However, the unrealistic test hypothesis remains that there is no difference between the series (see Section 1.4.2).

Additionally, two-sample t-tests are sensitive to (abnormally) small variability in one of the series. This danger is avoided if the results of the receiving laboratory are compared with confidence intervals of 'reference values', obtained from a collaborative study, as proposed by Vial [2]. However, the other shortcomings still remain valid, not to mention the large effort (at least six participating laboratories in the collaborative study are proposed), as well as the problem that the initial collaborative study is a transfer itself.

7.3.1.3 Statistical Equivalence Tests

These tests (also known as two one-sided t-tests) reverse the objective to demonstrate *'significant sameness'* instead of 'significant difference' that may be without analytical relevance [1, 3, 7, 8] (see also Section 1.4.2). The consumer risk is strictly controlled, because it corresponds with the chosen confidence level (see Fig. 7-2, equivalence test, 2% true bias) and the test behaves in a scientifically 'logical' manner, because the power is increased with increasing sample size. However, in order to restrict the α-error, a sufficient number of determinations are required. Even in the case of no true bias between the laboratories, the acceptance rate is only 56% in the case of three days with three replicates each (Figure 7-2, equivalence test, true bias 0%). In the paper of Kringle [3], the acceptable difference is defined from calculated probability charts of passing the test, as a function of variability and bias for various nominal assay values. The number of determinations required to achieve a given probability is calculated as five – eight days per site for three replicates per day.

It was emphasised that such thorough studies provide a good database for establishing specification limits (see Chapter 6), but – of course – they require large resources and effort. For the purpose of a transfer, it is questionable whether such a tightly statistically controlled approach is of much added value. Often, the most important practical risk in a transfer is not a rather small bias, but misinterpretations or lack of sufficient detail in the control test description, which would be well recognised by less extensive approaches.

7.3.2 Assay

7.3.2.1 Equivalent Test Approach

Kringle et al. [3] provided two examples for assay of a drug product (specification limits 95.0–105.0%), with an acceptable difference of the means of ± 2.0% and inter-

mediate precision ≤ 2.5%, and for drug substance (specification limits 98.0–101.5%), with an acceptable bias between –1.0% and +0.5% and intermediate precision ≤ 1.0%. The required sample size was determined as five – eight days per site for drug product and drug substance, respectively, with three replicates per day. The number of batches is not explicitly mentioned, but it can be assumed that one batch was used.

7.3.2.2 ISPE Recommendations

In the ISPE Guide [1], it is recommended that at least two analysts in each laboratory should analyse three batches in triplicates. The means and the variability of the results are compared, either with defined absolute acceptance limits, or by applying an equivalence test with respect to an acceptable difference of 2% between the means. No specific recommendation is given for an acceptable precision.

The design proposed must be regarded as inappropriate, because comparing the batches separately will suffer from the large uncertainty connected with only three determinations (see Section 2.1.1.3). Because the equivalence test includes the variability, there is a high probability of failure. Even in the case of an experimental difference of zero between the means, the acceptable limit is exceeded if the experimental precision is larger than 0.9%. For a precision of 0.5%, the maximum difference between the means is 0.9% (Eq. 2.3-4). Pooling all batches is only justified if their content is the same, but even if this can be demonstrated statistically, it is known, a priori, that the different batches do not have exactly the same content, thus increasing the variability of the results. However, in this case the question must be, why use three batches at all?

7.3.2.3 Approaches Using Simple Comparison

Brutsche [4] proposed for LC, GC, and CE-assays of drug substances to perform four replicates by two analysts per laboratory and to limit the difference between the means to less than 1.0% and the intermediate precision to less than 2.0%. For drug products, six replicates are recommended and limits of 2.0% for the difference of the means as well as for intermediate precision. Other companies perform six independent sample preparations per laboratory and establish acceptance criteria for the difference between the means of ≤ 2% and ≤ 3% and for the precision (repeatability) of ≤ 1.0% and ≤ 2.0% for a drug substance and drug product, respectively [5]. The same acceptance limits as for a drug product are also proposed by Fischer [6].

Acceptance limits based on statistical considerations

In order to achieve a practical compromise between data reliability and effort, the authors recommend that one representative batch should be analysed by two analysts per laboratory with six replicates each. The results are evaluated by *simple comparison* to absolute acceptance limits. However, the acceptance limits are *derived from statistical considerations* based on the specification range and taking the manufacturing variability into account. The background corresponds to the concept of the process capability index (see Chapter 10, Eq. 10-4). In order to achieve compatibility between the overall variability and the specification range, a capability index of at

least unity is required. The overall variability consists of the variance contributions from the analytical procedure, and from the manufacturing process. Equation (7-1) can now be rearranged to obtain the maximum permitted analytical variability s_{max}, but this requires the manufacturing variability to be expressed as a multiple of the analytical variability (Eq. 7-2).

$$c_p \equiv 1 \leq \frac{SL_{upper} - SL_{lower}}{6\ s_{overall}} = \frac{SL_{upper} - SL_{lower}}{6\ \sqrt{s^2_{process} + s^2_{analyt}}} \tag{7-1}$$

$$s_{\max} = \frac{SL_{upper} - SL_{lower}}{6\ \sqrt{\nu + 1}} \text{ with } s^2_{process} = \nu\ s^2_{analyt} \tag{7-2}$$

The multiple ν can be estimated for different types of manufacturing processes, or obtained from historical data. For example, in the case of drug substance with only a small amount of impurities, the process variability can be neglected (with respect to the content of active), thus $\nu = 0$. For a standard drug product manufacturing process, it is reasonable to assume that the process variability is about the same as the analytical variability, i.e., $\nu = 1$.

Because for assay the analytical variability is usually an important, if not dominating contribution, a tighter control is required in the transfer. Therefore, two acceptance parameters are defined for precision and accuracy each:

A. Individual (analyst) repeatability (Eq. 2.1-2).
B. Overall (pooled) repeatability (Eq. 2.1-9).
C. Difference between the laboratory means.
D. Difference between the analyst mean and the grand mean.

The acceptance limits are calculated from the maximum permitted analytical variability s_{max} and factors for each acceptance parameter are calculated from the 95% confidence intervals. For the precision, the factors correspond to the upper limit of the 95% confidence interval (Eq. 2.1-7), i.e., for A and B with five and 20 degrees of freedom, respectively. In the case of A, the factor of 2.09 was tightened to 1.89 due to the result of simulation studies. For C and D, the acceptance limit corresponds to the 95% confidence interval of the respective difference (Eqs. 7-3 and 7-4).

$$CI_D = 2\ t_{2n-2,0.05}\ \sqrt{\frac{2}{n}}\ s_{\max} = 1.69\ s_{\max} \text{ with } n = 12 \tag{7-3}$$

$$CI_D = 2\ t_{m\ (n-1),0.05}\ \sqrt{\left(\frac{m-1}{m\ n}\right)}\ s_{\max} = 1.48\ s_{\max} \text{ with } n = 6 \text{ and } m = 4 \tag{7-4}$$

In Table 7-1, two examples are shown for assay of active in a drug product and a drug substance. The acceptance limits can be easily calculated from the constant factors (which depend only on the design of the comparative study, i.e., the number of series and determinations) and the maximum permitted analytical variability, obtained from the specification limits and an estimation of the contribution of the manufacturing variability. The latter two are the only variables that need to be considered, in the case of a drug substance these are usually just the specification limits.

Table 7-1 Acceptance criteria for the results of a comparative assay study involving two analysts at each laboratory analysing one representative batch six times each.

		Drug Product	**Drug Substance**
Specification Range		95 – 105 %	98 – 102 %
Variability Assumption		$s^2_{max} = s^2_{process}$	$s_{process} = 0$
s_{max}		1.18 %	0.67 %
Acceptance Parameters	Factors for s_{max}	Acceptance Limits	
Individual standard deviation	1.86	2.20 %	1.25 %
Overall standard deviation	1.36	1.60 %	0.90 %
Difference of the laboratory means	1.69	2.00 %	1.15 %
Difference of individual mean and grand mean	1.48	1.75 %	1.00 %

But what about the risks of such a simple comparison approach?

In contrast to a statistical equivalence test (see Section 7.3.1.3), they cannot be defined numerically, but the results of a simulation study demonstrate that both the α-risk (failure of an acceptable result) and the β-risk (passing of a non-acceptable result) are well controlled (Fig. 7-3). In the case of small (true) variability, up to the maximum acceptable standard deviation of 1.18% and no intrinsic bias, almost all experimental studies are acceptable. One unacceptably high (true) variability result would itself cause more than half of the studies to fail. A small bias is more likely to be tolerated in the case of small variabilities, but a bias of more than 2% will fail

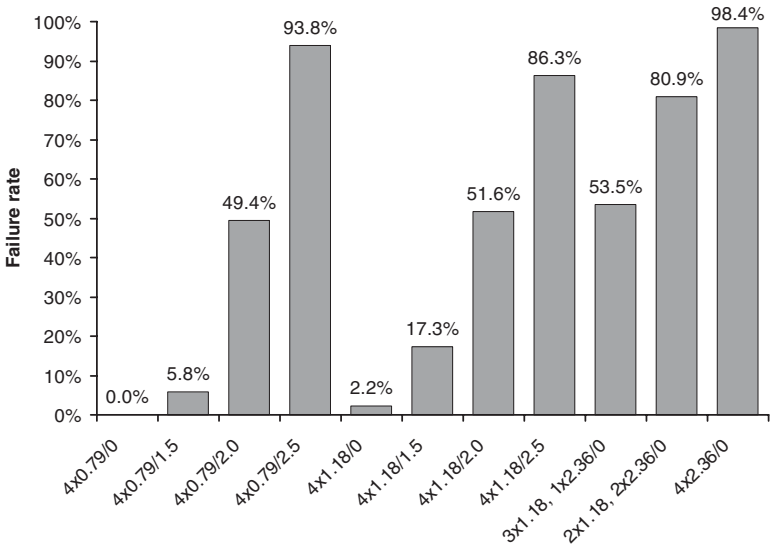


Figure 7-3 Failure rate for 166 665 simulated collaborative studies of a drug product assay with the acceptance criteria given in Table 7-1. The true relative standard deviation in the four series with six determinations each is indicated first, followed by the true percentage difference between the laboratory means (bias).

with high probability. (The reader must be aware that the bias and variabilities indicated are the true values. Using these values, normally distributed data sets are simulated; their individual means and standard deviations are distributed in a rather large range, as illustrated for the latter in Figure 2.1-3.)

7.3.3 Content Uniformity

No separate experimental investigations are required if the method is identical to assay. The ISPE Guide [1] recommends analysing one batch using two analysts in each laboratory for content uniformity, i.e., ten units each, and then comparing means and precision. The means of the receiving laboratory should be within ± 3% of the originating laboratory, or should conform to an equivalence test with an acceptable difference of 3%. No specific recommendation is given for an acceptable precision.

Peeters [5] describes the same design, but only one analyst per laboratory and acceptance criteria of ≤ 5.0% for the difference between the means and ≤ 6.0% for the precision. The latter corresponds to the USP criterion for content uniformity and is justified by the additional unit variability. The acceptable difference between the means seems to be rather large, because the unit variability is reduced in the mean.

The authors recommend that two analysts per laboratory perform the content uniformity test. The relative standard deviation of each analyst should be less than 6.0% and the difference between both the analyst's means within the laboratory and the laboratory means should be less than 3.0%.

However, the wider limits due to the influence of the unit variability may hide potential problems with the analytical procedure itself. Therefore, if a tighter control is required (for example, if there is no separate assay procedure, i.e., if the average of the content uniformity results is reported as the mean content), in an alternative design the unit variability can be cancelled out. This may be done by combining the ten test solutions prepared according to the control test and repeating them six times, i.e., 24 results (2 × 2 × 6) from 240 units. Another approach could be the normalisation of the unit content by the weight of the unit, insofar as the active is homogeneously distributed in the whole unit. In these cases, the same acceptance criteria as for assay can be applied.

7.3.4 Dissolution

A dissolution test with six units or a dissolution profile from 12 units is recommended [1], for immediate release and extended release or for less experience in the receiving laboratory, respectively. The data are either compared statistically, for example, by an F_2 test of the profiles [9], or based on an absolute difference of the means (5%). The same acceptance criteria are proposed for dissolution profiles of six units per laboratory [5] and of six or 12 units by two analysts each per laboratory.

7.3.5 Minor Components

The ISPE-Guide [1] recommends for impurities, degradation products, and residual solvents, that two analysts at each site should investigate three batches in duplicate on different days. Response factors and the limit of quantitation should be confirmed at the receiving laboratory and the chromatograms should be compared to ensure a similar impurity profile. Accuracy and precision should be evaluated at the specification limit, if spiked samples are used. For moderately high levels, an equivalence test with an acceptable difference of 10% is suggested, for lower levels an absolute difference of 25% relative or 0.05% absolute. No specific recommendation is given for an acceptable precision.

According to Peeters [5], six sample preparations per laboratory are performed, if required with spiked or stress stability samples. The acceptance criteria are dependent on the level of impurities (Table 7-2) or residual solvents (Table 7-3).

Table 7-2 Acceptance criteria for accuracy and precision of impurities [5].

Concentration level (with respect to the quantitation limit QL)	Relative difference between means (%)	Precision (%)
QL to < 2 ± QL	≤ 60	≤ 25
2 ± QL to 10 ± QL	≤ 40	≤ 15
10 ± QL to 20 ± QL	≤ 30	≤ 10
> 20 ± QL	≤ 20	≤ 5

Table 7-3 Acceptance criteria for accuracy and precision of residual solvents [5].

Concentration level	Absolute difference between means (ppm)	Precision (%)
< 200 ppm	≤ 20	≤ 20
200 to 1000 ppm	≤ 40	≤ 15
> 1000 ppm	≤ 60	≤ 10

In the approach recommended by the authors, two analysts per laboratory perform six determinations each. Preferably, one batch with a representative impurity profile or spiked samples should be used. If not representative, several batches may be used. All specified impurities have to be taken into account as well as the total sum. In order to get a consistent sum, all considered peaks, as well as a quantitation threshold, should be defined. The relative standard deviation for each analyst should be below an acceptance limit, which is defined on a case-by-case basis taking the concentration level of the actual impurity into account. For orientation, the same precisions as given in Table 7-2 are used, taken from [10]. The difference between the analyst means per laboratory and between laboratory means should be less than the acceptance limit. The values provided in Table 7-4 may be used for orientation, but in all cases appropriate scientific judgment must be used.

Table 7-4 Proposed acceptance criteria for the difference of means of impurities.

Level of impurities, degradants, residual solvent, or water[a] (individually specified and total sum) (%, relative to active)	Acceptance criterion
<0.1	appropriate limit, case-by case decision[b]
0.10 to 0.15	≤ 0.05 % absolute
> 0.15 to 0.30	≤ 0.10 % absolute
> 0.30 to 0.50	≤ 0.15 % absolute
> 0.50 to 0.80	≤ 0.20 % absolute
> 0.80 to 1.0	≤ 0.25 % absolute
> 1.0 to 5.0	≤ 20 % relative
> 5.0	≤ 10 % relative

a: actual amount in the samples investigated
b: for orientation, twice the value given in Table 7-2 for an acceptable precision may be used

7.4 Conclusion

The transfer process should be designed to ensure that well-selected and validated analytical methods are transferred into well-prepared laboratories. It will normally take several weeks to complete and, for a new product being transferred into a QC laboratory, this should start approximately eighteen months prior to the technology requiring use, in order to allow the receiving laboratory to give meaningful feedback and to complete possible investments. There will be occasions where transfer need to take place more quickly, and this can be achieved by shortening or omitting stages from the process. In the authors' experience this significantly increases the risk of transfer difficulties but this approach allows these risks to be assessed at the start of the process.

The main risks involved in analytical transfer are associated with 'knowledge-transfer', i.e., differences in handling and performing the analytical procedure due to 'cultural' or 'traditional' differences, misinterpretations, misunderstandings, lack of clarification, etc. In an appropriate risk-based approach, this should be taken into consideration in the design of comparative studies and acceptance criteria. The authors recommend the use of well-based absolute acceptance criteria and a sound compromise with respect to the number of determinations, to ensure the practical reliability of the results, and to avoid any 'testing into statistics'.

7.5 References

[1] ISPE: Good Practice Guide: Technology Transfer (2003) (www.ispe.org).

[2] J. Vial, A. Jardy, P. Anger, A. Brun, J.-M. Menet: Methodology for transfer of LC methods based on statistical considerations. J. Chromatogr. A 815 (1998) 173–182.

[3] R. Kringle, R. Khan-Malek, F. Snikeris, P. Munden, C. Agut, M. Bauer: A unified approach for design and analysis of transfer studies for analytical methods. Drug Information J. 35 (2001) 1271–1288.

[4] A. Brutsche: Transfer of analytical methods (Novartis). IAPT-Course 227: Transfer of analytical methods, Feb. 17–18 2004, Darmstadt (www.apv-mainz.de).

[5] L. Peeters: Transfer of analytical methods (Johnson & Johnson) IIR Lifesciences Course: Analytical Method Validation, Oct. 20–22 2003, London.

[6] M. Fischer: Transfer of analytical procedures (Lilly). European Compliance Academy Course: FDA-Compliance in Analytical Laboratories, April 28–30 2004, (www.concept-heidelberg.de).

[7] USP: <1010> Analytical Data – Interpretation and Treatment. Pharmacopeial Forum 30 (2004) 236–263.

[8] G. Limentani, M. Ringo, M. Bergquist, F. Ye, E. McSorley: A pragmatic approach to the design of scientifically rigorous method transfer and method equivalence studies using a statistical equivalence test. Proceedings of the Pharmaceutical and Biomedical Analysis Conference, Florence 2004, 14.

[9] FDA: Dissolution testing of immediate release solid oral dosage forms (Guidance for Industry) 2001.

[10] J.B. Crowther, M.I. Jimidar, N. Niemeijer, P. Salomons: Qualification of laboratory instrumentation, validation, and transfer of analytical methods. In: Analytical Chemistry in a GMP Environment. A Practical Guide. Eds. J.M. Miller and J.B. Crowther, Wiley, New York 2000, 423–458.

8
Validation of Pharmacopoeial Methods

John H. McB. Miller

8.1
Introduction

The purpose of the pharmacopoeia is to provide publicly recognised standards for use by health care professionals and others concerned with the quality and safety of medicines. Monographs of general methods published in the pharmacopoeias are employed by regulatory authorities in the licensing process for active substances and medicinal products in national medicine testing laboratories and by the medicines inspectors who audit pharmaceutical manufacturers. Manufacturers of active ingredients, excipients and medicinal products also apply tests of the pharmacopoeia to prepare their applications to the licensing authorities for approval to market their substances or products, and to control their quality after manufacture.

There are many pharmacopoeias published throughout the world but there are some which exert an international rather than a national influence or applicability. These include the European Pharmacopoeia (Ph.Eur) [1], which includes mandatory standards applicable in all countries signatory to the Convention [2], the British Pharmacopoeia (BP) [3] particularly its monographs on individual pharmaceutical formulations and the United States Pharmacopoeia (USP) [4].

Test procedures for the assessment of the quality of active pharmaceutical substances, excipients and medicinal products described in pharmacopoeias constitute legal standards in the countries or regions where they are applied. The European Pharmacopoeia is recognised as the official compendium in the directives [5] of the European Union and is enshrined in the medicines legislation of the other European countries adhering to the Convention. In the United States of America assays and specification in monographs of the USP are mandatory [6] and according to the regulation relating to Good Manufacturing Practice [7] the methods used for assessing compliance of pharmaceutical products to established specifications must meet proper standards of accuracy and reliability. It is essential, therefore, that all pharmacopoeial methods either new or revised are supported by a complete validation package to ensure their fitness for purpose.

In recent years there has been a massive effort to harmonise licensing requirements for approval of new medicinal substances and products involving the regulatory authorities and the pharmaceutical associations of the three major economic

Method Validation in Pharmaceutical Analysis. A Guide to Best Practice. Joachim Ermer, John H. McB. Miller (Eds.)

ISBN: 3-527-31255-2

regions of the world (United States of America, Japan and Europe). This process of harmonisation – the International Conference of Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use – affects all aspects of drug registration including efficacy, safety and quality.

Amongst the guideline documents concerning the quality aspects relating to new drug substances and products were two guides on analytical validation [8–9]. These guides document the basic elements expected to be present in a licensing application submitted by a manufacturer to show that the analytical methods proposed have been properly validated to demonstrate fitness-for-purpose. These guidelines have been adopted and applied by the Pharmaceutical Discussion Group (PDG), which consists of the United States Pharmacopoeia, the Japanese Pharmacopoeia (JP) and the European Pharmacopoeia, mirroring the stakeholders in the ICH process. The USP has published a chapter [10] in the General Information section whilst the Ph.Eur. have included the ICH guidelines as well as a supplementary text which is specific to the application of methods used in the pharmacopoeia in appendices to the Technical Guide for the Elaboration of Monographs [11]. The European Pharmacopoeia has also published a statement [12] in the introduction to the 4th Edition that "the procedures for the tests and assays published in the individual monographs have been validated, according to current practice at the time of their elaboration, for the purpose for which they are intended."

The analytical methodology presented in the monographs of the pharmacopoeia is appropriate to ensure the quality and safety (to the extent of limiting the presence or ensuring the absence of toxic impurities by application of analytical methods) of the drug substances and products. The quality is controlled by tests for identity, purity and assay of content. The tests are to be validated to demonstrate their fitness-for-use.

The extent of validation required or the emphasis given to particular validation characteristics depends on the purpose of the test. The characteristics to be evaluated are the same as those described for the submissions of regulatory approval (see Table 1-1).

The validation aspects to be considered for each type of analytical procedure included in a pharmacopoeial monograph as categorised in the USP is given in the Table 8-1 [10].

In many instances, the requirements of the monograph are described in part or in whole from the specification of the manufacturer(s). The manufacturer of a substance or product which is to be the subject of a pharmacopoeial monograph will furnish the pharmacopoeial authorities a complete validation dossier which will include:

- proof of structure by interpretation of spectral data;
- specification;
- justification for methods used and acceptance criteria applied;
- synthetic route and purification process;
- list of potential impurities with their chemical structures from the manufacturing process (including residual solvents);

- details of the separation technique used to detect and control the content of impurities including retention times, relative retention and response factors;
- assay procedure for content which is to be stability indicating when applied to the finished product (unless there is an adequate test for decomposition products listed in the specification);
- historical batch data.

Table 8-1 Validation characteristics required for assay validation [10].

Analytical Performance Characteristics	Assay Category I	Assay Category II		Assay Category III	Assay Category III	Assay Category IV
		Quantitative	Limit tests			
Accuracy	Yes	Yes	*	*	*	No
Precision	Yes	Yes	No	Yes	Yes	No
Specificity	Yes	Yes	Yes	*	*	Yes
Detection limit	No	No	Yes	*	*	No
Quantitation limit	No	Yes	No	*	*	No
Linearity	Yes	Yes	No	*	*	No
Range	Yes	Yes	*	*	*	No

*: may be required, depending on the nature of the specific test
Category I: Analytical procedures for the determination of the substance for pharmaceutical use either as the raw material or in the finished pharmaceutical product.
Category II: Analytical procedures for the determination of synthetic impurities or decomposition products in raw materials and finished pharmaceutical products. There may be limit or quantitative tests.
Category III: Analytical methods for the determination of performance characteristics by functionality tests (for example, dissolution).
Category IV: Analytical procedures for identification of the substance.

This data is assessed by the pharmacopoeial authorities and a monograph is prepared, the tests of which are experimentally verified (the mechanisms vary according to the pharmacopoeia). Particular attention is given to any separation technique employed for a test for impurities and/or for the assay of content for their transferability and robustness (see later).

When dealing with a multi-source substance different manufacturing processes may be employed and the impurity profile may be different, it will be necessary to verify that a given method is capable of separating and adequately controlling the known impurities from the different manufacturers. In such circumstances, it may be necessary to adapt an existing method or to propose a novel method adequately controlls all the impurities from the different manufacturers. Then a complete validation of the new or adapted method should be conducted. As part of the validation package for a pharmacopoeial method the inclusion of appropriate system suitability criteria (see Section 2.8) and the establishment of reference standards are essential.

8.2 Identification

The purpose of the identification section of a monograph is to confirm that the identity of the substance being examined corresponds to the substance described by the monograph. The test described is to be specific, otherwise a series of tests of different selectivity should be described that, when taken together, will ensure the specificity of the identification. The tests commonly described in the Identification section of a monograph may include one or more of the following analytical techniques:

- infrared spectrophotometry;
- ultraviolet spectrophotometry;
- melting point (freezing point or boiling point for liquids);
- optical rotation;
- chromatographic methods;
- electrophoretic methods;
- chemical reactions.

Some of these require comparison to a reference standard (CRS) for confirmation of identity of the substance by, for example,

- spectroscopy, usually infrared spectrophotometry, where the spectrum of the substance to be examined is compared with the spectrum of the CRS, or to the reference spectrum;
- separation techniques, where the retention times (or migration distance or migration time) of both the substance to be examined and the CRS are compared;
- identification by peptide mapping, which requires the use of both a CRS and its chromatogram.

Other techniques such as ultraviolet spectrophotometry, optical rotation and melting point, require the substance to be examined in order to comply with numerical limits derived from the pharmaceutical standard.

In both situations the standard must have been characterised appropriately by chemical attributes, such as structural formula, empirical formula and molecular weight.

A number of techniques are expected to be used including:

- nuclear magnetic resonance (NMR) spectroscopy;
- mass spectroscopy;
- infrared spectroscopy;
- the spectra are to be interpreted to support the structure;
- elemental analysis to confirm the percentage composition of the elements.

Having established the proof of the molecular structure, which is indicated in the manufacturer's validation package, it is then necessary to demonstrate that the use

of a single test (usually infrared spectroscopy) can discriminate between compounds of closely related substances. Often it can be demonstrated that the substance as an acid or a base can be specifically identified by a means of the infrared spectrum alone (comparison to a reference standard) but often for a salt, in particular for the sodium salt of an organic acid, or halide salt of an organic base, it is necessary to add a supplementary test to identify the ion. When the spectrum is not considered sufficiently different from compounds of similar structure it is necessary to add another test to ensure specificity usually a supplementary test, such as melting point, or TLC.

For identification tests relying on compliance with limits, for example, acceptance ranges for melting point, specific optical rotation and specific absorbance, the values must be determined on well-characterised substances of high purity. None of the other tests listed above can stand alone to ensure unequivocal identification of the substance and so several tests must be performed, the results of which when taken together will lead to specificity. However, tests must be well chosen. Examples of a strategy to follow is given for beta-blockers [14] and the benzodiazopines [15] where a selection of non-specific tests to achieve selectivity is illustrated in Tables 8-2 to 8-4. Both the Ph.Eur and the International Pharmacopoeias [13] often gives two series of identity tests. Thus, by a judicious selection of a number of simple tests, an unambiguous identification of a substance is possible. A second series for identification has been included in the European and International Pharmacopoeias so that they can be applied in pharmacies, where sophisticated instrumentation is not available, to identify the raw materials which are employed for the preparation of dispensed formulations. This is a legal requirement in some countries of Europe (France, Belgium and Germany) or in under-developed countries of the world. The World Health Organisation has also published a series of simple tests for the identification of the drugs given in the essential drugs list [16]. The tests have been validated by inter-laboratory testing. The tests that constitute the 'Second Identification' may be used instead of the test or tests of the 'First Identification', provided that it has been shown that the substance is fully traceable to a batch which has been certified to comply with all requirements of the monograph.

The pharmacopoeias also include tests for the identification of ions and groups and these general tests have been reviewed [17–20] with a view to harmonising them for the Japanese, European and United States Pharmacopoeias. These tests have been further investigated and validated for selectivity and to remove those which use toxic reagents. Proposals for harmonised test for these pharmacopoeias have been published [21–22] for comment within the international harmonisation process of the Pharmacopoeial Discussion Group (PDG).

When an identification series is being investigated, other similar compounds, whether or not they are subject to monographs of the pharmacopoeia, are to be examined to demonstrate that a particular combination of tests will successfully distinguish one from another.

Thus all tests prescribed for the identification of a substance for pharmaceutical use must be performed to ensure the unequivocal confirmation of its identity.

Table 8-2 Possible tests for inclusion in the alternative series for the identification of beta-adrenoceptor blocking agents [14].

	Test A	Test B		Test C	Test D	Test E
	Melting Point (°C)	UV absorbance (methanol as solvent)		(TLC-Rt)	Colour reation [1)]	Anion
		δ Max	$A_{1cm}^{1\%}$			
Alprenolol HCl	58	271 nm	12	12	Red-brown	Chloride
		277 nm	68			
Atenolol	153	275 nm	54	71	Yellow	–
		282 nm	46			
Labetolol HCl	156	303 nm	93	16	Violet [2)]	Chloride
Metoprolol tartrate	49	275 nm	45	29	Red	Tartrate
		282 nm	38			
Oxprenolol HCl	76	274 nm	75	22	Violet	Chloride
Pindolol	171	264 nm	366	48	Blue violet [3)]	–
		287 nm	189			
Propranolol HCl	94	289 nm	215	12	Blue/black	Chloride
		320 nm	68			
Timolol maleate	72	297 nm	244	31	–	Maleate

1 Marquis reagent
2 Test for phenol group using ferric chloride
3 Test for pyrrole group using 4-dimethylamnobenzylaldehyde

Table 8-3 Non-fluorinated benzodiazepines [15].

Substance	Melting Point (°C)	Spectrum UV			TLC : Rf Mobile phase			Chemical Reactions
		Solvent	δ max (nm)	Specific absorbance	A	B	C	
Bromazepam	246–251	CH_3OH	233	1020–1080	0.0	0.35	0.58	Identification of bromide
			331	58–65				NaOH: yellow
Chlordiazepoxide	240–242	0.1N HCl	246	1120–1190	0.06	0.34	0.60	Diazotatization
			308	316–336				
Chlordiazepoxide HCl	212–218	0.1 N HCl	246	996–1058	0.06	0.34	0.61	Diazotatization
			309	280–298				Identification of chloride
Clonazepam	237-240	CH_3OH	248	350–370	0.36	0.39	0.59	Diazotatization
			310	450–470				NaOH: yellow
Diazepam	131–135	H_2SO_4 0.5%	242	≈ 1020	0.25	0.52	0.76	Fluorescence (H_2SO_4)
		CH_3OH	285					
			366	140–155				Identification of chloride
Lorazepam	171–173 (decomp)	EtOH	230	≈ 1100	0.24	0.17	0.40	
			316					
Medazepam	101–104	0.1 N HCl	254	≈ 860	0.23	0.56	0.78	
Nitrazepam	226–230	H_2SO_4 0.5%	280	890–950	0.35	0.38	0.56	Diazotatization
		ds CH_3OH						NaOH: yellow

Table 8-4 Fluorinated benzodiazepines [15].

Substance	Melting Point (°C)	Spectrum UV			TLC : Rf Mobile phase			Chemical Reactions
		Solvent	δ max (nm)	Specific absorbance	A	B	C	
Fludiazepam	88–92 69–72							
Flumazenil	198–202	EtOH absolute	245	656–696	0.5			
Flunitrazepam	168–172	EtOH absolute	253 310	490–530 320–340	0.52	0.52	0.76	Identification of fluoride NaOH : yellow
Flurazepam	84–87	H_2SO_4 0.5% ds CH_3OH	239 284 362	600 270 80	0.0	0.42	0.69	Identification of chloride
Flutazolum	147 (decomp)	CH_3OH	246	295–315	0.8			
Halazepam	164–166				0.74			
Haloxazolam	185							
Midazolam	159–163	HCl 0.1N	258	356–378	0.05	0.6		
Quazepam	≈ 147							

8.3 Purity

The pharmacopoeias in the Tests section of the monograph include methods for the control of impurities, the selectivity (equivalent to discrimination as used in the ICH guidelines) of which depends on the method described and the purpose for which it is intended. The test method may be simply to indicate the general quality by an 'Appearance of Solution' test or may be specific for a known toxic impurity such as ethylene oxide [23]. The commonly described test methods include:

- appearance of solution;
- pH or acidity/alkalinity;
- optical rotation;
- ultraviolet/visible spectrophotometry;
- separation techniques (organic impurities);
- loss of drying/determination of water;
- foreign ions;
- heavy metals;
- atomic absorption/emission spectroscopy;
- sulphated ash;
- residual solvents/organic volatile impurities.

For the methods described in the test section of the monograph, it is essential to demonstrate that the selectivity and the sensitivity of the method are sufficient to

limit the level of impurity to that which is considered safe, based on toxicological studies, and subsequently to that which can be achieved under controlled production conditions. Two types of test are described. When a so-called *limit test* is described there is a direct comparison between a reference solution and a test solution and the measured response of the test solution should be less than the measured response of the test solution. In this case it is also necessary to determine the sensitivity of the method by means of the detection limit. When the *test is quantitative* it is necessary to demonstrate the quantitation limit is at or below the threshold limit [24] and that there is a linear relationship between the concentration of the impurity level and the response around the acceptance limit for the impurity. Precision is also to be established (repeatability, intermediate precision and reproducibility).

8.3.1 Appearance of Solution

These are subjective tests which compare the colour and/or opalescence of the test solution to a series of reference solutions. These tests are introduced to give a general assessment of the purity of the substance. When the impurity causing the colour or opalescence is known, the visual test should be validated by comparison to a quantitative analytical technique. Often, however, the impurity responsible for the permitted degree of colouration or opalescence is unknown and validation is based on the examination of data, from batches which would otherwise meet the requirements of the specification, which are supplied by the manufacturer.

For material intended for parenteral use and for highly coloured solutions, especially when the use of colour of solution test is contemplated, it is preferable to apply a limit of absorbance measured with a spectrophotometer at a suitable wavelength (usually between 400 and 450 nm). The concentration of the solution and the limit of absorbance must be stated. The conditions and limit must be based on knowledge of the absorbance curve in the range 400 – 450 nm and on results obtained with appropriate samples, including storage and degraded samples, as necessary.

8.3.2 pH or Acidity/Alkalinity

These are non-specific tests used for the control of protolytic impurities.

This test allows the limitation of acidic or alkaline impurities originating from the method of preparation or purification, or arising from degradation (for example, from inappropriate storage) of the substance. The test may also be used to verify the stoichiometric composition of certain salts.

Two types of test for protolytic impurities are used in the Pharmacopoeia: a titration experiment using indicators or electrometric methods to define the limits, the acidity–alkalinity test; or the measurements of pH.

pH measurement is preferred if the substance has buffering properties, otherwise a titrimetric procedure is recommended.

The question of whether to prescribe an acidity–alkalinity test or a pH measurement in a pharmacopoeial monograph can be decided on the basis of an estimation of the buffering properties of the material [11]. To this end a titration curve can be constructed for an aqueous solution (or, if necessary, extract) in the intended concentration (10 – 50 g/l) of a, preferably pure, specimen of the substance to be examined, using 0.01M hydrochloric acid and 0.01 M sodium hydroxide, respectively, and potentiometric pH measurements.

The inflection point of the titration curve is the true pH of the solution and will, for a pure compound, be at the point of intersection with the pH-axis. The measure of the buffering capacity of the solution to be examined is the total shift in pH (ΔpH), read from the titration curve as a result of adding, on the one hand, 0.25 ml of 0.01 M sodium hydroxide to 10 ml of the solution and, on the other hand, 0.25 ml of 0.01 M hydrochloric acid to another 10 ml portion of the same solution. The larger is ΔpH, the lower is the buffering capacity. For a sample that is not quite pure, a parallel displacement of the titration curve is to be performed so that the true pH of the solution is on the pH-axis before the ΔpH can be read from the curve.

The magnitude of ΔpH of the solution to be examined determines the choice of method for the limitation of protolytic impurities, according to the following scheme. The classification is based upon the observation that the colour change for most indicators takes place over a pH range of 2 units.

Class A	ΔpH>4	Acidity–alkalinity test utilising two appropriate indicators.
Class B	4>ΔpH>2	Acidity–alkalinity test utilising a single appropriate indicator.
Class C	2>ΔpH>0.2	A direct pH measurement.
Class D	ΔpH>0.2	The protolytic purity cannot be reasonable controlled.

Compounds that are salts consisting of ions with more than one acidic and/or basic function belong to Class D and, for these, pH measurement can contribute to ensuring the intended composition if the limits are sufficiently narrow.

In certain cases, a test for acidity–alkalinity cannot be performed with the use of indicators due to colouration of the solution to be examined or other complications, and the limits are then controlled by pH measurement. The addition of standard acid and/or base results in decomposition or precipitation of the substance to be examined may be necessary, regardless of the buffering properties, to prescribe a pH test.

If, a pH measurement has to be prescribed for solutions with little or no buffering capacity, the solution to be examined is prepared with carbon dioxide-free water. Conversely, the use of carbon dioxide-free water for preparing solutions that have sufficient buffering capacity to warrant a direct pH measurement, is not necessary since the required precision, which seldom exceeds one-tenth of a pH unit, will not be affected. When an acidity requirement corresponds to not more than 0.1 ml of 0.01 M sodium hydroxide per 10 ml of solution to be examined, the latter must be prepared using water free from carbon dioxide.

8.3.3 Specific Optical Rotation

Specific optical rotation may be used to verify the optical purity of an enantiomer. This method may be less sensitive than chiral LC. In the case where one enantiomer is to be limited by the measurement of specific optical rotation, then it is to be demonstrated that under the conditions of the test, the enantiomer has sufficient optical activity to be detected. Whenever possible the influence of potential impurities should be reported. Limits for the specific optical rotation should be chosen with regard to the permitted amount of impurities. In the absence of information on the rotation of related substances and when insufficient amounts of the related substances are available, the limits are usually fixed at ± 5% around the mean value obtained for samples which comply with the monograph. Samples of different origin should be examined whenever possible.

Measurement of an angle of rotation may be used to verify the racemic character of a substance. In that case, limits of + 0.10° to –0.10° are usually prescribed but it is to be demonstrated that, under the conditions of the test, the enantiomer has sufficient optical activity to be detected.

8.3.4 Ultraviolet Spectrophotometry

When ultraviolet spectrophotometry is used for a limit test for an impurity it is to be demonstrated that, at the appropriate wavelength, the related substance to be limited makes a sufficient contribution to the measured absorbance. The absorbance corresponding to the limiting concentration of the related substance must be established.

8.3.5 Limit test for Anions/Cations

These are simple and rapid tests which are to be shown to be appropriate by recovery experiments and/or comparison with other more sophisticated methods.

8.3.5.1 Sulphated Ash [25]

The sulphated ash test is intended as a global determination of cationic substances. The limit is normally 0.1%. This gravimetric test controls the content of foreign cations to a level appropriate to indicate the quality of production. This method is well established and no further validation is required.

8.3.5.2 Heavy Metals [26]

Appropriately low limits must be set for the toxic elements, many of which are controlled by the heavy metal test (for example, lead, copper, silver, mercury, cobalt, cadmium and palladium). This test is based on the precipitation of these heavy metals as the sulphides and visual comparisons with a standard prepared from a lead solu-

tion. Five different procedures are described [26] in the European Pharmacopoeia. Normally the limits are set at 10 ppm or 20 ppm. Lower limits may be set in which case Limits Tests E is to be used. Nevertheless, it is important that the appropriate procedure is chosen for the substance to be examined and that the response is verified at the proposed limit.

It must be noted that, for some of the procedures which require incineration, there is the risk of the loss of some heavy metals such as mercury, and lead in the presence of chloride [27]. This has been reported for methods C and D of the European Pharmacopoeia. If this is likely to be the case, then such metals may be controlled by using a closed mineralisation technique, for example, a Teflon bomb, followed by the application of the reaction to the sulphide or by an appropriate instrumental technique, for example, atomic absorption spectrophotometry.

The European Pharmacopoeia has recently published proposals [28] to revise the procedures for the testing of heavy metals by including a 'monitor' preparation and adding a method using microwave digestion (Method G)

Previously the test required the visual examination of the sulphide suspension produced but now, if it is difficult to distinguish the extent of the precipitation, usually it is proposed, as an option, to use a filtration technique and to examine the filtrates. By this means the sensitivity of the method is improved and the comparison is easier.

The proposed test for heavy metals is performed with the sample and the sample is 'spiked' with lead at the desired limit. The brown opalescence by the sample must be less than, and that produced by the 'spiked' sample must be equal to or more than the standard.

8.3.5.3 **Colour or Precipitations Reactions**

Limit tests are also described for individual cations and anions,which are based on visual comparison of a colour or opalescence. It is essential that it is demonstrated that:

- the colour or opalescence is visible at the target concentration (limit);
- the recovery of added ion is the same for the test and reference solutions (by visual observation and if possible by absorbance measurement);
- the response is sufficiently discriminatory around the target value by (50 percent, 100 percent and 150 percent of the target value) measuring the absorbances at an appropriate wavelength in the visible region.
- a recovery experiment at the target value is carried out six times and the repeatability standard deviation is calculated. Recovery should be greater than 80 percent and the repeatability RSD should be less than ±20 percent.

It would be desirable, when appropriate, to compare the results obtained from a recovery experiment, using the proposed limit test procedure, with a quantitative determination using a different method, for example, atomic absorption spectrophotometry for cations or ion chromatography for anions. The results obtained by the two methods should be similar (see Section 2.3.5).

8.3.6 Atomic Absorption Spectrometry

Atomic spectroscopy is exclusively employed in tests to determine the content of specific elements which are present in substances as impurities. The following validation requirements are pertinent to atomic spectrometric methods.

In principle, this technique is specific, using the appropriate source and wavelength, for the element to be determined, since the atom emits or absorbs radiation at discrete spectral lines. However, interferences may be encountered due to optical and/or chemical effects. Thus it is important to identify the interferences and, if possible, to reduce their effect by using appropriate means before starting the validation programme.

Such interferences may result in a systematic error if a direct calibration procedure is employed or may reduce the sensitivity of the method. The most important sources of error in atomic spectrometry are associated with errors due to the calibration process and to matrix interference.

Chemical, physical, ionisation and spectral *interferences* are encountered in the atomic absorption measurements and every effort should be made to eliminate them or reduce them to a minimum. Chemical interference is compensated for by addition of releasing agents or by using the high temperature produced by a nitrous oxide-acetylene flame; the use of ionisation buffers compensates for ionisation interference and physical interference is eliminated by dilution of the sample, matrix matching or through the method of standard additions. Spectral interference results from the overlapping of two resonance lines and can be avoided by using another resonance line. The use of a Zeeman or continuum source background correction also compensates for spectral interference and interferences from molecular absorption, especially in graphite furnace atomic absorption. The use of multi-element hollow-cathode lamps may also cause spectral interference.

Scatter and background in the flames/furnace increase the measured absorbance values. Background absorption covers a large range of wavelengths, whereas atomic absorption takes place in a very narrow wavelength range of about 0.002 nm. Background absorption can, in principle,be corrected by using a blank solution of exactly the same composition as the sample but without the specific element to be determined, although this method is frequently impractical.

Once the instrumental parameters have been optimised to avoid interferences so that sufficient sensitivity can be obtained (the absorbance signal obtained with the least concentrated reference solution must comply with the sensitivity specification of the instrument) the *linearity of response* against concentration is to be ascertained around the limiting concentration. No fewer than five solutions of the element to be determined at and around the limiting concentration should be prepared and the precision determined from six replicates at each concentration.

A calibration curve is constructed from the mean of the readings obtained with the reference solutions by plotting the means as a function of concentration, together with the curve which describes the calibration function and its confidence level. The residuals of all determinations, i.e., the difference between the measured and

estimated absorbance are plotted as a function of concentration. When a suitable calibration procedure is applied, the residuals are randomly distributed around the x-axis. When the signal variance increases with the concentration, as shown from either a plot of the residuals or with a one-tailed *t*-test, the most accurate estimations are made with a weighted calibration model. Both linear and quadratic weighting functions are applied to the data to find the most appropriate weighting function to be employed (see Section 2.4.1.1 and 2.4.2).

When aqueous reference solutions are measured to estimate the calibration function, it must be ensured that the sensitivity of both the sample solution and the aqueous solution are similar. When a straight-line calibration model is applied, differences in sensitivity can be detected by comparing the slopes of a standard addition and an aqueous calibration line. The precision of the estimation of the slopes of both regression lines depends on the number and distribution of the measurement points. Therefore, it is recommended to include sufficient measurement points in both regression lines and to concentrate these points mainly on the extremes of the calibration range.

The slopes of the standard addition line and the aqueous calibration line are compared, by applying a *t*-test, to check whether slopes of both regression lines are significantly different. If that is the case, then the method of standard additions is to be applied and, if not, then direct calibration can be employed.

For many applications a *pre-treatment of the sample* is required (for example, extraction or mineralisation) and so it is essential to perform a recovery experiment either from a similar matrix, which has been spiked, or a sample which has been fortified with the element to be determined. In both cases, the element is to be added to achieve a concentration of the element at the limit. The recovery experiment should be repeated six times and the mean and standard deviation should be determined.

When atomic absorption methods are prescribed in monographs, certainly in the European Pharmacopoeia, it is rare that a detailed procedure is prescribed and in fact it is the responsibility of the user, using the information provided, to elaborate a procedure which is suitable for their equipment. The user must therefore validate the procedure but it should conform to the requirements given in the General Chapter [29].

8.3.7
Separation Techniques (Organic Impurities)

These techniques are employed for the control of organic impurities (related substances). Related substances as defined by the European Pharmacopoeia include; intermediates and by-products from a synthetically produced organic substance; co-extracted substances from a natural product; and degradation products of the substance. This definition does not include residual organic solvents, water, inorganic impurities, residues from cells and micro-organisms or culture media used in a fermentation process. Normally, as indicated earlier in the chapter, the manufacturer will have validated the method for the control of impurities in the substance for

pharmaceutical use. Liquid chromatography is the technique most commoly employed and its selectivity to separate all known and potential impurities must be demonstrated.

Reference Standards

Reference standards of specified impurities, which are either synthesised or isolated from unpurified batches of the active ingredient, are to be available for the validation process. The substance for pharmaceutical use or impurity must be analysed:

1) to characterise the substance (proof of molecular structure) by appropriate chemical testing (as described for identification);
2) to determine the purity
   - determination of the content of organic impurities by an appropriate separation technique (such as gas chromatography (GC), liquid chromatography (LC) or capillary electrophoresis (CE));
   - quantitative determination of water (for example, micro or semi-micro determination);
   - determination of the content of residual solvents;
   - determination of loss on drying may in certain circumstances replace the determinations of water and residual solvents;
   - determination of the purity by an absolute method (for example, differential scanning calorimetry or phase solubility analysis, where appropriate. The results of these determinations are to support and confirm the results obtained from separation techniques. They are not included in the calculation of the assigned value);
   - determination of inorganic impurities (test for heavy metals, sulphated ash, atomic absorption spectrophotometry, ICP, X-ray fluorescence) – often the values obtained will have no consequence on the assignment of the purity of the standard.

These impurity reference standards are then employed to validate the chromatographic method where

- selectivity is to be demonstrated (lack of interferences);
- sensitivity is to be shown by the determination of the quantitation limit for each of the specified impurities. The response factors (correction factors) for each of the impurities relative to the substance for pharmaceutical use is to be determined;
- linearity of response should be apparent in the ranges of the reporting threshold to 120 percent, when normalisation of the limiting concentration is employed, and when an external standard is used;
- repeatability and intermediate precision is to be assessed;
- system suitability criteria for selectivity, sensitivity, accuracy and precision should be included.

A general chapter [30] 'Impurities' has been published in the European Pharmacopoeia, explaining the rationale for their control and how to interpret the monographs.

8.3.7.1 Liquid Chromatography

In general, the pharmacopoeias do not specify the brand name of the stationary phase employed (with the exception of the British Pharmacopoeia) but instead describe the column in general terms, for example, octadecylsilylsilica gel for chromatography R. It is essential, therefore, that the procedure is performed using a number of reverse-phase stationary phases with different characteristics. If the method is robust, then chromatography using the different stationary phases will be similar – the retention times and relative retentions of the substance for pharmaceutical use and its impurities, will essentially be the same as will the order of elution. At the present are approximately 600 different commercial C-18 stationary phases, exhibiting different characteristics, are available. The stationary phases may be characterised by type of silica employed (A or B), carbon loading, pore size, particle size, type of particles (irregular or spherical), the specific surface area, or the extent of blocking of silanol groups. The recent introduction of hybrid columns has complicated the situation further. A number of review articles have been published [31–33] of work performed in an attempt to categorise the different column types according to performance,and the USP has formed a working party to this end, but as yet no system has been found to a categorise columns adequately according to chromatographic performance (see Section 2.8.3.5 for a fuller discussion). Ideally a liquid chromatographic method for the control of impurities in a pharmacopoeia should be sufficiently robust so that the necessary selectivity should be achieved on any reverse-phase (C-18) stationary phase. Unfortunately, due to the differences in column performance from one type to another, this is not possible. Nonetheless, any method which has been developed and validated, should be tested on a number of stationary phases of an approximately similar type,based on their physical characteristics. In this regard it would be helpful to the user for the pharmacopoeias to describe better the stationary phases in the monographs. The USP has published lists of reverse-phase columns [34], which fall into the different categories (for example, L1 for octadecyl silane chemically bonded to porous silica or ceramic microparticles). The European Pharmacopoeia lists, as a footnote to proposed monographs published in Pharmeuropa, the commercial name(s) of the stationary phase(s) shown to be suitable during the development, evaluation and validation of the method. Subsequently, after publication in the Pharmacopoeia, suitable columns are listed on the website [35]. At this point the list may be more extensive if the chromatographic method was part of a collaborative trial to establish a pharmacopoeial assay standard.

Since greater selectivity is required to separate an increasing number of impurities, particularly when emanating from different manufacturing processes, there is an increasing propensity to employ gradient liquid chromatography. When gradient elution is described for the control of impurities, it is inadvisable to change the type of reverse-phase column and, in such a case, the column proposed by the manufacturer should be adequately described in the text of the test for related substances.

Reference standards of impurities may not be available or may only be available in insufficient quantities to establish pharmacopoeia reference standards, in which case the impurities will have to be controlled using a dilution of the test solution,

Table 8-5 Two liquid chromatographic systems required to control the impurities of trimethoprim. Impurities are identified by relative retention for application of correction factors [36].

Substance	Approx.RRT		Correction Factor
	Method A	Method B	
Trimethoprim	1 (RT = 5.2min)	1 (RT = 4.3 min)	1
Impurity A	1.5		1
Impurity B	2.3	1.3	0.43
Impurity C	0.8		1
Impurity D	2.0		1
Impurity E	0.9		0.53
Impurity F	4.0		1
Impurity G	2.1		1
Impurity H		1.8	0.50
Impurity I		4.9	0.28
Impurity J	2.7		0.66

A: A stainless steel column 0.25 m long and 4.0 mm in internal diameter packed with base-deactivated octadecylsilyl silica gel for chromatography R (5 µm).

B: A stainless steel column 0.25 m long and 4.6 mm in internal diameter packed with cyanopropylsilyl silica gel for chromatography R (5 µm) with specific surface area of 350 m^2/g and a pore diameter of 10 nm.

which makes it essential to determine the response factors. It will also be necessary to identify the impurities in the chromatogram of the test solution, particularly when the acceptance criteria for the impurities are different. In such cases, mixtures of impurities as reference standards may be required. These are prepared by the Pharmacopoeia and identify the impurities as well as demonstrating adequate selectivity of the system. There are also some cases where a single chromatographic method is incapable of separating and controlling all the impurities and so more than one chromatographic test is required. In the monograph for trimethoprim [36] eleven impurities are controlled by the application of two chromatographic methods. The impurities are identified by relative retention so that the necessary correction factors can be applied (Table 8-5). In method (A) the stationary phase is probably insufficiently described and there could be confusion in differentiating between impurities C and E, which have similar retentions, and only one which requires the application of a correction factor. The monograph for sumatriptan succinate [37] illustrates the use of mixtures of impurities. Figure 8-1 shows the chromatogram expected for sumatriptan for system suitability CRS, which is employed in related substances test A, to identify the specified impurities A and H. The chromatogram of sumatriptan impurity mixture CRS, shown in Figure 8-2, is used to identify the peaks of the impurities controlled by method B. Five peaks are obtained including that corresponding to sumatriptan. The area of the peak due to impurity E is about twice the area of the other impurity peaks. This is necessary to identify the peak of

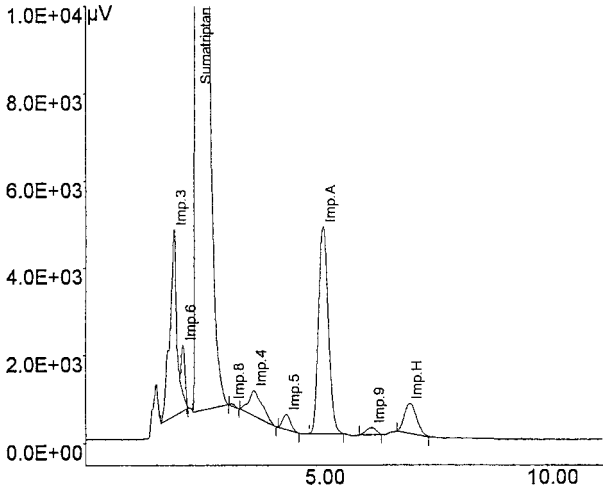


Figure 8-1 Chromatogram of sumatriptan for system suitability CRS to identify impurities A and H [37].

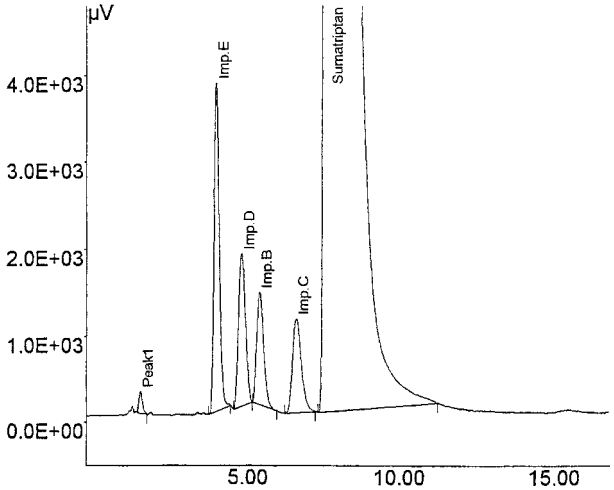


Figure 8-2 Chromatogram of sumatriptan impurity mixture CRS to identify impurities A, B, C, D and E [37].

impurity E, which may vary in retention time depending on the column used, and which is also limited to a different level from the other impurities.

Generally, the impurities are estimated by using an external standard of the substance itself, diluted to the limiting concentration to avoid the use of specific impurities, in which case the method will be checked for sensitivity, linearity and precision. The quantitation limit must be determined for the external standard, which is either a dilution of the substance to be examined, or a known impurity. When a peak of an impurity elutes close to the peak of the substance, particularly if

it elutes after the peak due to the substance, then the quantitation limit is to be determined for this impurity. The quantitation limit is to be at or preferably below the disregard level (reporting threshold) so it will demonstrate adequate sensitivity in the reference solution.

Stability data should also be verified to demonstrate the period of use of reference and test solutions.

When an extraction procedure is employed, a recovery experiment using known and available impurities is to be carried out under optimal conditions and the results reported. It is to be demonstrated that the recovery is consistent and has an acceptable precision.

Other separation techniques are also employed in the pharmacopoeias but to a very much lesser extent.

8.3.7.2 Gas Chromatography

The same requirements as described under liquid chromatography are required, except that limitation of impurities is usually determined by peak area normalisation, in which case linearity of response of the detector with concentration is to be demonstrated in a range from the disregard limit to 120 percent of the test solution concentration (see Section 2.5). The disregard limit is usually defined by a requirement for the signal-to-noise ratio, which is to be equal to or greater than the qualification limit (10).

An alternative approach is to employ an internal standard, in which case the ratio of the area of the secondary peak (impurity) to that of the internal standard is compared with the ratio of the peak areas of the reference substance to that of the internal standard.

8.3.7.3 Capillary Electrophoresis

Usually an internal standard is employed to improve the precision of the method. Evidence is to be provided that the method is sufficiently selective and sensitive (quantitation limit). The other requirements, as described for gas chromatography, are to be met.

8.3.7.4 Thin-layer Chromatography and Electrophoresis

Although thin-layer chromatography has been extensively used in the past, its application to the control of impurities is declining in favour of the aforementioned quantitative techniques. In fact, it is now the policy of the European Pharmacopoeia to replace TLC methods for related substances testing by quantitative separation techniques, especially liquid chromatography. Nonetheless, TLC may still be employed for specific impurities, which cannot be detected by other procedures.

When there is a test for related substances, a thin-layer chromatographic method is usually described in such a way that any secondary spot (impurity) in the chromatogram of the test solution is compared with a reference spot, equivalent to the limiting concentration. The intensity of the secondary spot should not be more intense or bigger than the reference spot. Of course, with visual examination it is not possible to estimate a total content of related substances when several are present.

The selectivity of the method is to be demonstrated, i.e., the capability of separating the specified impurities using plates of the same type but of different origin. The use of spray reagents should be universal unless the test is intended to limit a specific impurity, in which case a reference standard is to be employed for comparison. The sensitivity of the procedure is to be verified. When a visual method is applied, it is to be demonstrated that the quantity corresponding to the given limit is detectable. Data are also required to demonstrate linearity of response with concentration over an appropriate range, which incorporates the limit and repeatability and also the quantitation limit, when an instrumental procedure is to be applied. Usually the impurities are limited by comparison of the secondary peaks observed in a chromatogram of the test solution, with that of the principal peak obtained with the chromatogram of the reference solution. The area of the peak of the impurity should not be greater than the area of the peak (or a multiple of it) obtained with the reference solution. The use of an external standard is preferred to peak area normalisation, since the sensitivity can be increased by employing high concentrations of the substance to be examined in the test solution, even though the response of the principle peak is outside the linear range of the detector. The external standard solution is normally a dilution of the test solution at the limiting concentration of the related substances(s) or, in an increasing number of monographs, a solution of the specified impurity is employed.However, when the quantitation of impurity levels are required then linearity and precision need to be established.

8.3.8
Loss on Drying

When a loss on drying test is applied, the conditions prescribed must be commensurate with the thermal stability of the substance. The drying conditions employed should not result in loss of substance due to its volatility or decomposition.

Examination of the substance by thermogravimetric analysis will indicate water loss and decomposition. Usually, in the loss on drying test, the drying time is not defined in time, but drying is continued to constant weight which is considered to be when the difference in consecutive weighings do not differ by more than 0.5 mg, the second weighing following an additional period of drying.

8.3.9
Determination of Water

The semi-micro determination of water as described in the pharmacopoeias is the Karl Fischer titration, which is based on the quantitative reaction of water with sulphur dioxide and iodine in an anhydrous medium and which requires the presence of a base with sufficient buffering capacity. The titrant is the iodine-containing reagent and the end-point is determined by amperometry. Classically, pyridine was the base employed in the titrant but, because of its toxicity, has been replaced with non-toxic bases which are included in commercially available Karl Fischer reagents.

It is therefore necessary to ensure their suitability for use by means of a suitable validation procedure [11].

The result obtained when applying the method can be influenced by a number of parameters which affect its accuracy. For example, the sharpness of the end-point is affected by the composition of the reagent and the absolute amount of water in the sample [38]. The stabilisation time towards the end of the titration should be reduced to a minimum to avoid interference caused by side reactions. It is known that side reactions may occur in the presence of alcohols or ketones, especially in a poorly buffered or strongly alkaline reagents [39], and that penicillin acids may cause interference when using certain commercial reagents [40].

A number of approaches for the validation of the semi-micro determination of water have been published [41]. Examples were given using different substances and reagents. The substances chosen for examination were those known to present difficulties caused by interfering reactions and included erythromycin and its salts, folic acid, amoxicillin and isoprenaline. A standard validation method was considered to be the most appropriate to validate the Karl Fischer system by the European Pharmacopoeia and has been published in its 'Technical Guide for the Elaboration of monographs' [11].

The water content (m) of the sample is determined using the proposed conditions, after which to the same titration vessel a suitable volume of standardised water is added and titrated. At least five replicate additions and determinations should be performed. The regression base of cumulative water added against the determined water content is constructed and the slope (b), the intercept (a), with the ordinate and the intersection (d), of the extrapolated line with the abscissa,are calculated. The validation of the method is considered to be acceptable when

i) $b < 0.975$ and > 102.5,
ii) the percentage errors are not greater than 2.5 percent when calculated as follows:

$$e_1 = \frac{a-m}{m}100\% \qquad e_2 = \frac{d-m}{m}100\% \tag{8-1}$$

and

iii) the mean recovery is between 97.5 percent and 102.5 percent.

For erythromycin, in the direct determination of water with different titration systems, the repeatability was consistently poor when anhydrous methanol rather than a 10% m/v solution of imidazole in methanol was employed as the solvent (Table 8-6). In Table 8-7 the results of the determination of water using two different commercial reagents and two different solvents, using a standard addition, are presented. From this it can be concluded that results failing the acceptance criteria were only obtained when anhydrous methanol was employed as the solvent.

Table 8-6 Results for the determination of the water content of erythromycin base and esters employing different titration systems. Relative standard deviations are given in brackets [37].

Substance	Water content (relative standard deviation)											
	1a	1b	2a	2b	3a	3b	4a	4b	5a	6a	7a	7b
Erythromycin 1332	4.67 (0.21)	4.69 (0.13)	4.92	4.66 (0.36)	4.65 (0.58)	4.65 (0.27)	–	4.62 (0.19)	4.66 (0.45)	4.65 (0.22)	–	4.66 (1.09)
Erythromycin 9173	1.01 (6.9)	0.98 (0.85)										
Erythromycin estolate 3366	2.98 (1.18)	2.96 (0.28)	2.84 (1.94)	2.85 (0.43)	2.78 (1.16)	2.85 (0.43	–	2.77 (0.56)	2.76 (1.3)	2.74 (1.2)	–	2.72 (0.22)
Erythromycin ethylsuccinate 3367	1.39 (1.0)	1.35 (0.55)										
Erythromycin stearate 6133	1.98 (5.4)	1.83 (0.03)										

1–3 pyridine-free reagents from Riedel de Haen, BDH and Fluka respectively
4 pyridine-based reagent
5–7 pyridine-free reagents from Merck
a methanol as solvents
b 10% m/v imidazole in methanol as solvent

Table 8-7 Calculated values from the standard addition experiments (reproduced from [41]).

Sample	Titrant system	e_1	e_2	Recovery	Direct titration
Erythromycin	1a	0.04	–2.66	102.8	4.78
(1332)		0.09	–1.48	101.6	4.74
Erythromycin		–1.70	–3.01	101.4	1.05
(9173)		–0.85	–3.09	102.3	1.05
Erythromycin	1b	–0.59	–1.64	101.1	4.69
(1332)		–0.55	–1.15	100.6	4.69
Erythromycin		–1.97	–2.12	100.2	0.94
(9173)		–1.61	–1.41	99.8	0.93
Erythromycin	2a	–0.12	–5.05	105.2	4.79
(1332)		1.50	–3.77	105.5	4.88
Erythromycin		–1.11	–3.14	102.1	1.30
ethylsuccinate		–1.69	–3.96	102.4	1.29
Erythromycin	2b	0.02	–0.58	100.6	4.73
(1332)		0.32	1.60	98.7	4.72
Erythromycin ethyl		–0.04	–0.58	100.9	1.29
succinate		0.54	1.58	99.0	1.27
Proposed limits		<2.5	<2.5	97.5–102.5	

1 pyridine-free reagent (Riedel & Haen)
2 pyridine-free reagent (BDH)
a methanol as solvent
b 10% m/v imidazole in methanol

It is known that isoprenaline sulphate is slowly oxidised in the presence of iodine and it would be expected that the recovery and the slope would show positive deviations when this procedure was applied. This is demonstrated by the results in Table 8-8. Thus the procedure seems capable of detecting interferences. The examples illustrate the applicability of the validation procedure for the application of the Karl Fischer titration in the determination of water when using commercial pyridine-free reagents.

Table 8-8 Validation of semi-micro determination of water in a sample of isoprenaline sulphate by standard addition [41].

Experiment	W [mg]	m_{H2O} [mg]	C_{H2O} [%]	b	A [mg]	e_1 [%]	d [mg]	e_2 [%]	Recovery [%]
1	202.97	14.263	7.03	1.020*	14.402	1.0	14.114	–1.6	102.1#
2	250.65	18.079	7.21	1.023#	18.153	–0.4	17.741	–1.9*	102.4#
3	249.46	17.804	7.14	1.029#	18.089	1.6	17.587	1.2	103.4#
4	207.07	15.338	7.41	1.036#	14.494	1.0	14.962	–2.5#	103.5#
5	203.76	15.431	7.57	1.022#	15.628	1.3	15.288	–0.9	102.4#
Mean			7.27	1.026					1.028
Standard deviation			0.18	0.006					0.006

Four single standard addition experiments from five (marked by #) failed, one marked by *is very close to the given value.

W weight of the sample taken in mg
m_{H2O} water content of sample in mg
b slope of the regression line
a intercept in mg water
d intersection in mg water

Alternatively, the results obtained by the proposed method are not significantly different from results obtained by another method, such as gas chromatography, TGA, etc. Here again, the use of a particular pyridine-free reagent must be validated by the user.

The micro determination of water is used for the determination of small amounts of water in the range of 10 µm to 10 mg. In this case the iodine is generated electrochemically from iodate. The same procedures for the validation of the semi-micro determination of water are also applicable.

8.3.10
Residual Solvents or Organic Volatile Impurities

Although there may have been tests for specific solvent residues prescribed in individual monographs, the USP introduced a general test using gas-chromatographic methods [42], which were to be applied to the majority of active substances. The solvents to be controlled were considered to be carcinogenic or potentially carcinogenic and included benzene, carbon tetrachloride, chloroform, 1,4 dioxane and ethylene oxide. Subsequently, at the behest of the Pharmacopoeial Discussion Group

(PDG), the issue of residual solvents in pharmaceutical substances was investigated by ICH, resulting in the publication of ICH guideline Q3C 'Residual Solvents' [43]. The solvents employed to synthesise and/or purify substances for pharmaceutical use (active substances and excipients) were classified according to a toxicological assessment and set maximum acceptable limits [44]. Simultaneously, the European Pharmacopoeia developed and published a general methodology to identify and control these solvents [45]. The guideline also recommended that 'any harmonised procedures for determining the levels of residual solvents as described in the pharmacopoeias should be used, if feasible'. Consequently, the USP revised its list of organic volatile impurities to include the Class I residual solvents with the ICH limits, while the European Pharmacopoeia published the guideline as an information chapter [46] with an introduction explaining the application of the guideline to the pharmacopoeia. The European Pharmacopoeia has also published a monograph for 'Substances of Pharmaceutical Use' [42] which includes all general requirements for active substances and excipients, including those for residual solvents. Thus there is no specific test for 'Residual Solvents' given in individual monographs of the European Pharmacopoeia, unlike the USP. However, it is incumbent on the manufacturer to test for the solvents employed in the synthesis of the substance.

The general method in the European Pharmacopoeia for the identification and limitation of residual solvents has been developed [48] and investigated for selectivity (with respect to other residual solvents), sensitivity, precision and recovery using a restricted number of examples. Chromatograms for Class 1 and Class 2 residual

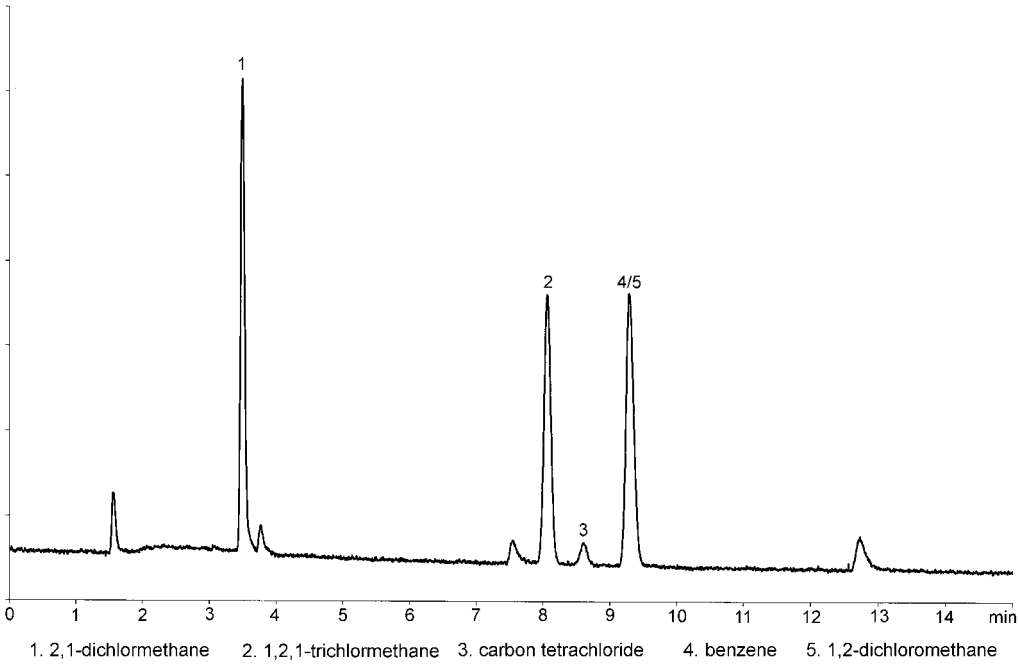
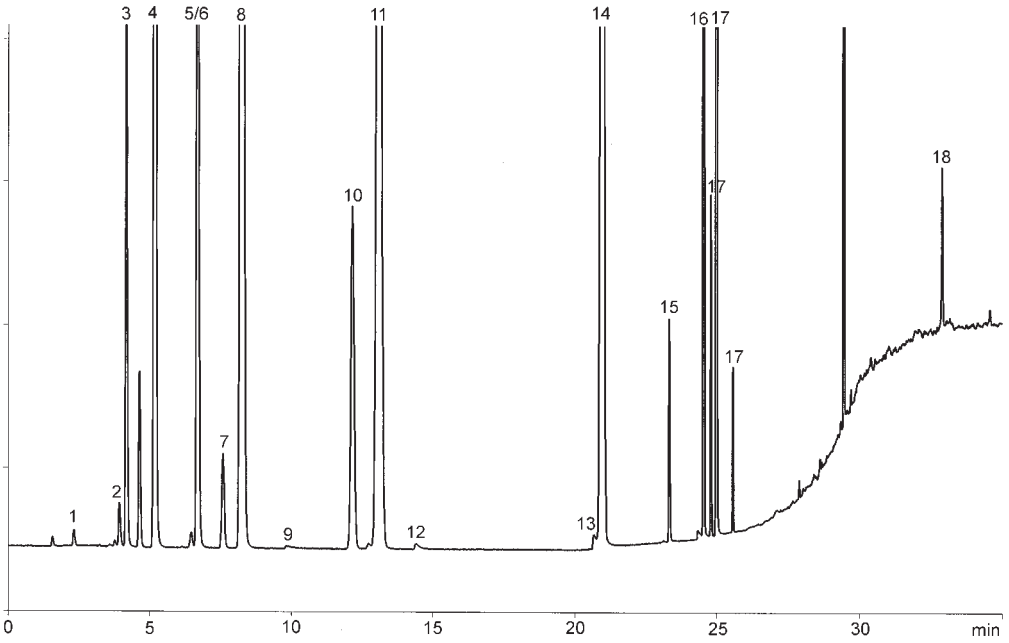


Figure 8-3 Chromatograms of Class 1 residual solvents using system A [45].

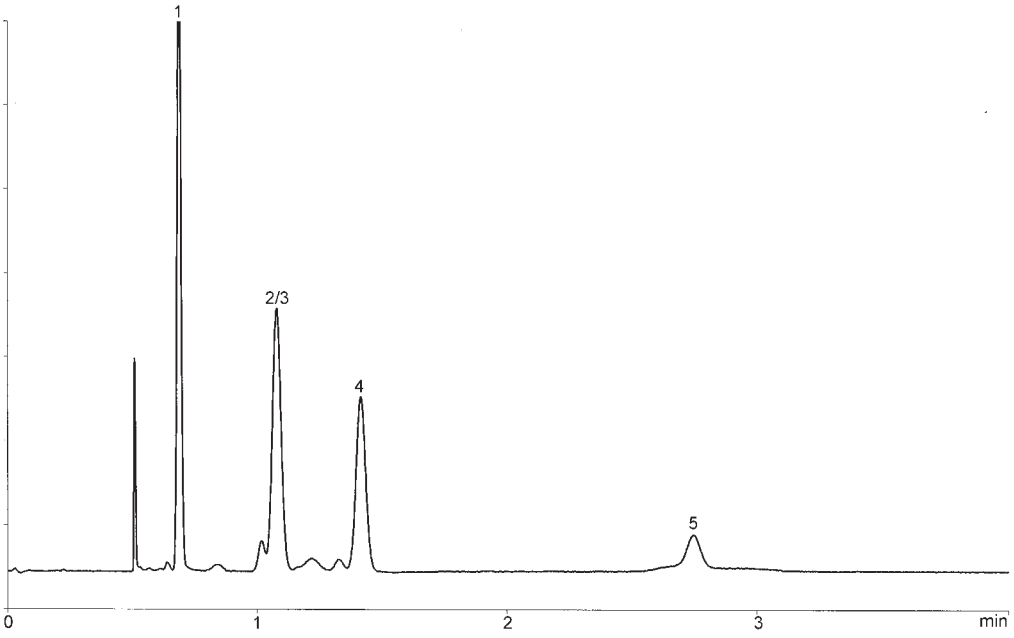


1. methanol
2. acetonitride
3. dichlormethane
4. hexane
5. *cis*-1,2-dichlorethene
6. nitromethane
7. chloroform
8. cyclohexane
9. 1,2-dimethoxymethane
10. 1,1,2-trichlormethane
11. methylcyclohexane
12. 1,4-dioxane
13. pyridine
14. toluene
15. 2-hexanone
16. chlorobenzene
17. xylene *ortho, meta, para*
18. tetralin

Figure 8-4 Chromatograms of Class 2 residual solvents using system A [45].

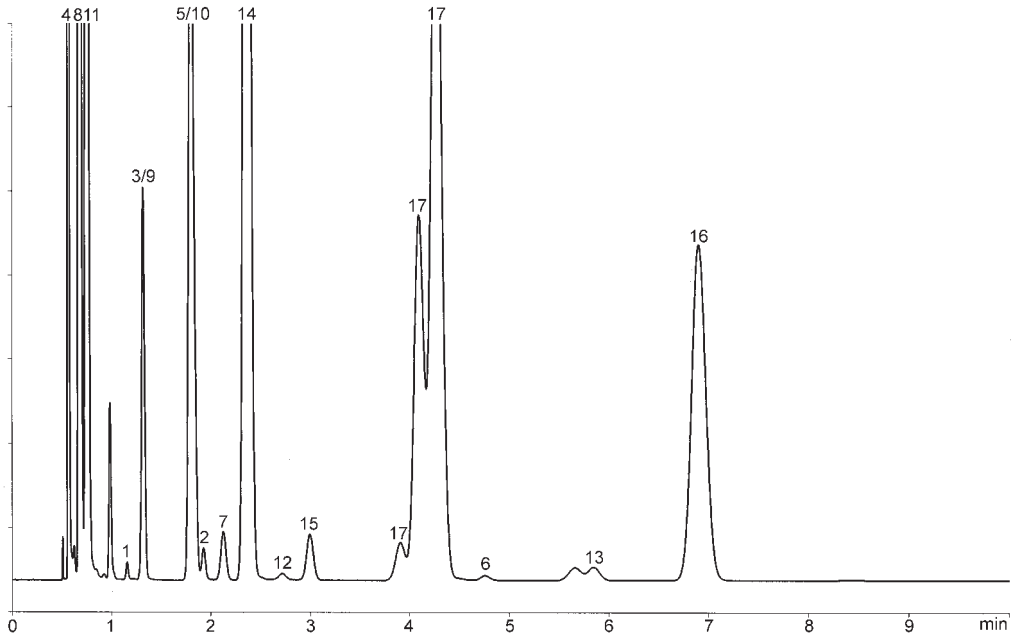
solvents using system A and system B are shown in Figures 8-3 to 8-6. However, it was impossible to validate the method for all the substances covered by a monograph in the pharmacopoeia. Thus when the pharmacopoeial method is to be employed quantitatively, further testing is required to validate the procedure for the substance under examination. It is clearly stated in the general method of the European Pharmacopoeia, that 'when the test procedure is applied quantitatively to control residual solvents in a substance, it must be validated'. The ICH validation guideline should be applied but with particular respect to:

- specificity, either using the two-column approach to confirm identification or by use of mass spectrometric relative abundance methods;
- quantitation limit, taking into account the effect of the matrix when employed in static head-space injection. In this regard, the application of multiple head-space extraction (MHE) is useful to determine suitable conditions [45, 49, 50]. MHE under ideal equilibration conditions, results in an inverse linear relationship between the logarithm of the signal response and the number of extractions from the head-space (Figure 8-7);



1. 1,1-dichloromethane 2. 1,1,1-trichloroethane 3. carbon tetrachloride 4. benzene 5. 1,2-dichloroethane

Figure 8-5 Chromatograms of Class 1 residual solvents using system B [45].



1. methanol
2. acetonitride
3. dichloromethane
4. hexane
5. *cis*-1,2-dichloroethene
6. nitromethane
7. chloroform
8. cyclohexane
9. 1,2-dimethoxymethane
10. 1,1,2-trichloroethene
11. methylcyclohexane
12. 1,4-dioxan
13. pyridine
14. toluene
15. hexanone
16. chlorobenzene
17. xylene *ortho, meta, para*
18. tetralin

Figure 8-6 Chromatograms of Class 2 residual solvents using system B [45].

– recovery experiments are to be validated under the conditions to be employed if it is intended to determine the solvent content quantitatively;
– repeatability is to be determined within the linear range.

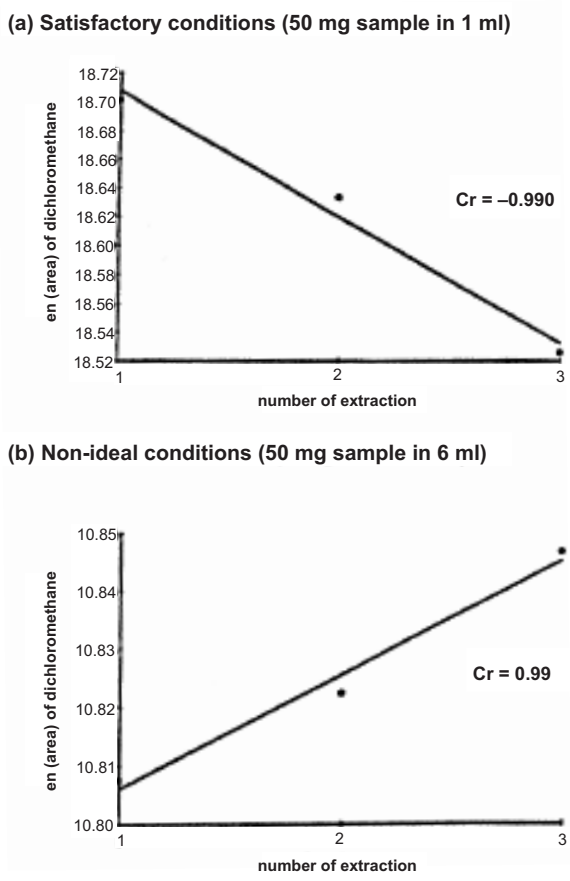


Figure 8-7 Multiple head-space extraction (MHSE) of methylene chloride in amoxicillin showing (a) satisfactory conditions and (b) non-ideal conditions [48].

8.4 Assay

Active substances, including biological substances and excipients, are determined for content in the raw material and in pharmaceutical preparations by a variety of techniques. For a chemical substance often the method employed is a volumetric titration, which is non-specific, but is an absolute method. Increasingly there is the use of comparative selective methods based on quantitative separation techniques, particularly liquid chromatography. Spectrophotometric methods based on ultravio-

let, visible or infrared spectroscopy, although still employed in some monographs, are decreasing in number.

8.4.1 Volumetric Titration

The use of volumetric titrations is applicable only when it has been demonstrated that

- impurities are present in low levels,
- the impurities are also titratable (if not then the limits set should be asymmetric to allow for the permitted content of the non-titratable impurity),
- the monograph includes a satisfactory test for related substances (organic impurities) based on a quantitative separation technique.

The advantage of a volumetric titration procedure is that, as it is an absolute method, a reference standard is not required and it usually exhibits higher precision.

When a volumetric titration procedure is well established it is sufficient to verify that the repeatability and accuracy of the titration are not greater than the limits given in Table 8-9 [11].

Table 8-9 Acceptance limits for repeatability and relative accuracy for different types of volumetric titrations (Reproduced from [11]).

Volume titration	Content limits (%)	Repeatability (RSD)	Relative accuracy (%)
Acid/base	± 1.0	0.33	± 0.67
Non-aqueous	± 1.0	0.33	± 0.67
Conjugate acid of base	± 1.0	0.33	± 0.67
Redox	± 1.5	0.5	± 1.0
Argentometric	± 1.5	0.5	± 1.0
Complexometric	± 2.0	0.67	± 1.33

Repeatability is expressed in the relative standard deviation determined from six replicates and the relative accuracy is calculated according to Eq.(8-2).

$$\Delta\overline{x} = \frac{\overline{x} - x}{x} \tag{8-2}$$

$\overline{x}$ = mean content
x = theoretical content

Another option when replacing one titration method by another is to demonstrate that the two methods do not give statistically different results.

In this case, the content of the substance is to be determined at least six times by both methods and there should be no statistically significant difference in the mean results obtained. Also, the repeatability of the replacement procedure should be bet-

ter than or not significantly different from the original procedure. For example, the European Pharmacopoeia has systematically replaced the non-aqueous titration of halide salts of organic bases [51, 52] by alternative titration procedures.

A third option has been described [53, 54] when developing a new volumetric assay procedure. The validation procedure requires the titration at seven different quantities under prescribed conditions in a randomised order to give end-point volumes in the range of 20 percent to 90 percent of the volume of the burette employed. Subsequently, the data are treated statistically and a number of criteria are to be fulfilled to permit acceptance of the titration procedure.

The relative error in reading the weight on the balance and the volume at the end-point is to be less than 0.5 percent of the given values.

The results, as end-point volumes (V_i) versus weight (m_i), are evaluated by linear regression. The regression line is calculated and characterised by the slope (b_{obs}), the extrapolated intercept (a_{obs}) and the precision as $sdv(v)$.

8.4.1.1 First Criterion – Proportional Systematic Error (Bias)

The calculated slope (b_{obs}), taking into account the titre of the standardised volumetric solution, is within 0.3 percent for potentiometric titrations (0.5 percent for visual titrations) compared with the theoretical value given as the titration constant (b_{theor}).

$$\left(\frac{b_{obs-btheor}}{b_{theor}}\right)100\% \text{ where } b_{theor} = \frac{Z}{MrCr} \tag{8-3}$$

Mr = relative molar mass
Z = stoichiometric factor of the chemical reaction
C_r = molar concentration of the titrant.

8.4.1.2 Second Criterion – Additional Systematic Error (Bias)

The extrapolated intercept (a_{obs}) is less than 0.4 percent for potentiometric titrations and 0.6 percent for visual titrations at the expected or target titration volume. This criterion may not be fulfilled when the titration is carried out too rapidly (potentiometric titration) or when an unsuitable indicator has been employed (visual titration).

$$\left(\frac{a_{obs}}{VT}\right)100\% \tag{8-4}$$

a_{obs} = extrapolated intercept of the regression line at zero
V_T = expected or target titration volume.

8.4.1.3 Third Criterion – Precision (Statistical Error)

The remaining estimated standard deviation ($sdv(v)$) is less than 0.3 percent for potentiometric titrations (0.5 percent for visual indicator titrations) of the mean titration volume of the end point using the titration procedure to be introduced in the monograph.

$$\left(\frac{sdv(v)}{VT}\right)100\% \tag{8-5}$$

$$\text{sdv (v)} = sdv(v) = \sqrt{\frac{\sum (V_i - a_{obs} - b_{obs}\, m_i)^2}{n-2}}$$

V_i = titrated volume,
m_i = mass of the substance titrated,
n = number of titrations performed.

8.4.1.4 Fourth Criterion – Practical Relative Error

Some titration procedures may not fulfil the first and second criteria but exhibit low and acceptable bias at the target titration volume (8 ml ± 1 ml for a 10 ml burette). Thus, if all the criteria given above are not met, then calculate the relative error at the target titration volume. It should be less than 0.3 percent or 0.5 percent for potentiometric titrations or visual end-point titrations, respectively.

$$\left|\left(\frac{a_{obs}}{VT} + \frac{b_{obs-btheor}}{b_{theor}}\right)\right| 100\% \tag{8-6}$$

8.4.2 Spectrophotometric Methods

These methods, although simple and rapid to perform, are being employed less and less for the determination of content, due to their lack of specificity. Nonetheless, when employed, they must be subjected to the full range of requirements for validation as described by ICH. Particular attention should be taken to assess the contribution of the absorbance of the known impurities in the substance. For quantitative infrared and colorimetric methods a reference standard is to be established, whilst for UV assay methods, either the specific absorption value or a reference standard is to be described.

8.4.2.1 Ultraviolet Spectroscopy

When the assay method is based on the measurement of the absorbance at a characteristic wavelength of maximum absorbance of the substance,then the suitability of the operating conditions including the solvent employed and its quality and the pH of the solution, etc., must be demonstrated. A linear relationship of measured absorbance to concentration is to be verified and,when used for the assay of an active ingredient in a preparation, the absence of interference is also to be demonstrated. When a reference standard is employed it must be of the highest purity attainable, which is estimated by applying a variety of techniques, including separation and absolute techniques. If a specific absorbance value is prescribed for the validation, then it is evaluated by an inter-laboratory trial using a batch of high purity [55]. An example is the determination of the specific absorbance of triamcinolone where the purity was determined before and after crystallisation using a variety of techniques

including differential scanning calorimetry, liquid chromatography and phase solubility analysis, followed by a collaborative trial to determine the specific absorbance.

8.4.2.2 **Colorimetric Assay Methods**

The operating conditions are optimised, after which validation is similar to that applied to UV measurement.

8.4.2.3 **Separation Techniques**

The most usual method applied is liquid chromatography, for determination of the content of raw materials and particularly for the content of the active ingredient(s) in pharmaceutical products. Many companies prefer to use the same method both for the control of impurities and for the assay. The analytical procedure should be stability-indicating [56] unless there is a suitable test for impurities included in the monograph. It is therefore essential to show that the method is capable of separating the decomposition products from the substance itself. Stability data should be available from the company supplying the method, from which the identity of the degradation products have been demonstrated. In some cases the liquid chromatographic method for the control of impurities and the liquid chromatographic method for the assay are not identical. The reason is normally to reduce the chromatographic run time for the assay. In such a case the assay method is less selective than the test for related substances but, nonetheless, it should be sufficiently selective to separate the potential decomposition products from the active substance. This difference in selectivity has an impact on the specification limits.

Reference Standards

For the application of the method, a suitable reference standard is to be established. Reference standards used in the assay have an assigned content and are generally primary standards, their content being assigned without comparison to another substance. The assigned content of the primary chemical reference standard is calculated from the values obtained from the analysis performed for the determination of purity, as described under 'Purity', and is verified by a calculation of the mass balance. The content is assigned on the basis of an inter-laboratory study, which may also serve to determine the reproducibility of the method [75].

As noted previously, a method which has already been validated by a manufacturer, is often indicated in the pharmacopoeial monograph of a substance for pharmaceutical use. The method is verified in a number of laboratories and adapted if necessary. However, the method is not fundamentally changed. Any changes are minor and are fully validated. There are occasions when major changes are required or a method has been developed for a particular application. In such a case the method development stage is followed by a complete validation including the organisation of an inter-laboratory trial to demonstrate the robustness of the method. Revision of the penicillin monographs of the Ph.Eur., where liquid chromatographic methods were introduced for both the control of impurities and for the assay for content, illustrates this approach [57–74] (Table 8-10). Not only was the robustness of the methods demonstrated, but the reference standards were also established simultaneously.

Table 8-10 Development and validation of liquid chromatographic methods for monographs of penicillins in the European Pharmacopoeia.

Penicillin	Reference
Amoxicillin	58
Amipicillin	57, 60
Benzathine benzylpenicillin	61, 62
Benzylpenicillin	63, 64
Cloxacillin	65, 66
Dicloxacillin	65, 67
Flucloxacillin	65, 68
Phenoxymethylpenicillin	70
Pivampicillin	71
Procaine benzylpenicillin	72
Ticarcillin	73, 74

Until now, reference standards for assay of the Ph. Eur. have been restricted to the determination of content of the substance for pharmaceutical use, employing the method prescribed in the pharmacopoeia, whereas reference standards of the USP and many of the BP are also employed for the assay of pharmaceutical preparations for which monographs appear in these compendia. Unless otherwise indicated on the label, the standards of the USP are considered to be 100 percent pure for the purposes of the assay methods. The Ph. Eur. has proposed [75] that the reference standard with an assigned content can be used for the determination of the content of the active ingredient(s) of a pharmaceutical preparation provided that the following conditions are fulfilled.

- The same liquid chromatographic method is used as is described in the monograph of the active substance.
- The method is stability indicating.
- Any pre-treatment (for example, extraction) has been validated by the user. In particular, there are no interferences from the matrix and, in the case of an extraction, the precision and recovery are satisfactory.

Of course, such considerations also apply when methods described in the BP and USP are employed for a pharmaceutical preparation, since it is not feasible that the BP and USP check all the pharmaceutical preparations in the market which might comply with the description of the monograph but which might have a different formulation.

8.5 Conclusions

In general, all tests and assays described in individual monographs of the pharmacopoeias have been validated for the substance for pharmaceutical use.

However, if pyridine-free reagents are employed for the Karl Fischer determination of water, the use of the method is to be validated. Application of the methods described for the quantitative determination of residual solvents will also require validation, as do the methods using atomic absorption spectroscopy.

All tests and assays based on separation techniques described in the pharmacopoeias must comply with the system suitability criteria which are described in the general chapter and/or in the individual monographs. Provided that these criteria are fulfilled further validation should not be necessary.

The methods described in monographs for pharmaceutical preparations do require validation.

General methods of the pharmacopoeia have been validated for the substances and preparations of the individual monographs but, if they are to be applied to non-compendial substances or preparations, their use must be validated.

Acknowledgement

My sincere thanks are due to Miss Emma Polland for her patience, tolerance and dedication during the preparation of the texts.

8.6 References

[1] European Pharmacopoeia 5th Ed (2005), Council of Europe, Strasbourg, France.
[2] European Treaty Series (1964) No 50 Convention on the Elaboration of the European Pharmacopoeia amended by the protocol to the convention (European Treaty Series No 134).
[3] British Pharmacopoeia (2003), HMSO, London, U K.
[4] United States Pharmacopoeia 26 (2003), United States Pharmacopoeial Convention Inc, Rockville MD, USA.
[5] European Union Directive 2001/83 and 2003/63.
[6] Section 501, Federal Food, Drugs and Cosmetic Act.
[7] Good Manufacturing Practice Regulations, 21 CFR 211.194(a).
[8] ICH Guideline (1994) Validation of Analytical Methods (Definitions and Terminology) I.F.P.M.A. Geneva.
[9] ICH Guideline (1997) Validation of Analytical Procedure: Methodology I.F.P.M.A. Geneva.
[10] Validation of Compendial Methods (1225), United States Pharmacopeia 26 (2003), United States Pharmacopeial Convention Inc, Rockville MD, USA.
[11] Technical Guide for the Elaboration of Monographs 3rd Ed, Pharmeuropa, Special Issue (1999).
[12] European Pharmacopoeia 4th Ed (2002), Council of Europe, Strasbourg, France.
[13] International Pharmacopoeia, World Health Organisation, Geneva, Switzerland.

[14] J.H.McB. Miller, J. Fuchs and J. Pijnenburg: The identification of Beta-adrenoceptor Blocking Agents by simple tests, Pharmeuropa **2** (1990) 220–226.
[15] J.L. Roberts S. Weber, J. Krippers, F. Schnyder: Flunitrayepam. Identification et specificité, Pharmeuropa **2** (1990) 158–160.
[16] Basic tests for drugs, World Health Organisation, Geneva, Switzerland.
[17] R. Verneil, C. Adam, A. Becuart, C. Ellie, S. Forget, C. Gorden, P. Loiseau, P. Michaud, F. Rodriguez and G. Thauvin: Harmonisation des methods générales d'analyse des pharmacopées américaine, européene et japanaise. Rapport d'une commission SFSTP, STP Pharma Pratiques **4** (1994) 5–18.
[18] R. Verneil, F. Adam, M. Bertucat, C. Ellie, S. Forget, C. Gorden, P. Loiseau, J. Renaurd, P. Thauvin and M.H. Vilijin: Harmonisation des methods d'identification des ions des pharmacopées américaine, européene et japanaise. Rapport d'une commission SFSTP, STP Pharma Pratiques **5** (1995) 127–156.
[19] R. Verneil, F. Adam, M. Bertucat, C. Ellie, S. Forget, C. Gorden, P. Loiseau, J. Renaurd, P. Thauvin and MH. Vilijin: Harmonisation des methods d'identification des ions des pharmacopées américaine, européene et japanaise. Rapport d'une commission SFSTP, STP Pharma Pratiques **5** (1995) 319–330.
[20] R. Verneil, F. Adam, M. Bertucat, C. Ellie, S. Forget, C. Gorden, P. Loiseau, J. Renaurd, P. Thauvin and M.H. Vilijin: Harmonisation des methods d'identification des ions des pharmacopées américaine, européene et japanaise. Rapport d'une commission SFSTP, STP Pharma Pratiques **6** (1996) 307–327.
[21] International Harmonisation, Identification Reactions for Ions and Groups, Part 1, Pharmeuropa **9** (1997) 101–106.
[22] International Harmonisation, Identification Reactions for Ions and Groups, Part 2, Pharmeuropa **10** (1998) 416–417.
[23] 2.4.25 Ethylene oxide sodium, European Pharmacopoeia 4th Ed (2002), Council of Europe, Strasbourg, France.
[24] ICH Guideline, Impurities in New Drug Substances Q (R), I.F.P.M.A. Geneva.
[25] 2.4.14 Sulphated Ash, European Pharmacopoeia 5th Ed (2005), Council of Europe, Strasbourg, France.
[26] 2.4.8 Heavy Metals, European Pharmacopoeia 5th Ed (2005), Council of Europe, Strasbourg, France.
[27] K.B. Blake: Harmonisation of the USP, EP and JP Heavy Metals Testing Procedure, Pharm Forum **21** (1995) 1632–1637.
[28] Heavy Metals (2.4.8), Pharmeuropa **15** (2003) 359–361.
[29] 2.2.23 Atomic absorption spectrophotometry, European Pharmacopoeia 5th Ed (2005), Council of Europe, Strasbourg, France.
[30] 5.10 Impurities, European Pharmacopoeia 5th Ed (2005), Council of Europe, Strasbourg, France.
[31] S.D. Rogers and J.G. Dorsey: Chromatographic silanol activitiy test procedures: a quest for a universal test, J. Chromatogr.A **892** (2000) 57–65.
[32] C. Stella, S. Rudaz, J.L. Veuthey and A. Tepapla: Silica and other materials as supports in liquid chromatography: Chromatographic tests and their importance for evaluating these supports, Parts I & II, Chromatographia **53** (2001) S113–S140.
[33] D. Visky, Y. Vander Heyden, T. Iványi, P. Barten, J. De Beer, B. Noszal, E. Roets, D.L. Massart and J. Hoogmartens: Characterisation of reverse phase liquid chromatographic columns by chromatographic tests, Pharmeuropa **14** (2002) 288–297.
[34] M.R.C. Marques: HPLC packings used in the USP-NF, Pharm Forum **26** (2000) 273–288.
[35] http//www.pheur.org

[36] Monograph 0060 Trimethoprim, European Pharmacopoeia 4th Ed (2002), Council of Europe, Strasbourg, France.
[37] Monograph 1573 Sumatriptan succinate, European Pharmacopoeia 4th Ed (2002), Council of Europe, Strasbourg, France.
[38] M. Müller, Deutsche Apotheker Zeitung **41** (1987) 2043.
[39] Th. Cachet & J. Hoogmartens: The determination of water in erythromycin by Karl Fischer titration, J Pharm Biomed Anal **6** (1998) 461.
[40] J. Hoebus, E. Roets and J. Hoogmartens, The interference of penicilloic acids with Karl Fischer tiration of penicillins, J Pharm Biomed Anal **15**, 461 (1998).
[41] S. Ebel, C. Engel, J.H.Mc.B. Miller, E. Pel, and J.Y. Concessa: A Validation Procedure for the Semi-micro Determination of Water, Pharmeuropa **11** (1999) 578–585.
[42] Organic Volatile Impurities (467), United States Pharmacopoeia **26** (2003), United States Pharmacopeial Convention Inc, Rockville MD, USA.
[43] ICH Guideline, Residual Solvents Q3C, I.F.P.M.A. Geneva.
[44] J.C. Connelly , R. Hasegawa, J.V. McArdle and M.L. Tucker: ICH Guideline Residual Solvents Toxicity Data Assessment, Pharmeuropa **9** Suppl (1997) 51–568.
[45] 2.4.24 Identification and Control of Residual Solvents, European Pharmacopoeia 5th Ed (2005), Council of Europe, Strasbourg, France.
[46] 5.4 Residual Solvents, Limiting Residual Solvent Levels in Active Substance, Excipients and Medicinal Products, European Pharmacopoeia 5th Ed (2005), Council of Europe, Strasbourg, France.
[47] Monograph 2034, Substances for Pharmaceutical Use, European Pharmacopoeia 4th Ed. Council of Europe, Strasbourg, France.
[48] M. Bauer, J. Bertrams, D. De Kaste, J. Hoogmartens, J. Kaufmann, E. Keller, J.H.McB. Miller and G. Sekules: Progress report of the working party on Residual Solvents (Technical) of the European Pharmacopoeia Commission, Pharmeuropa **8** (1996) 573–592.
[49] B. Kolb: Multiple headspace extraction – a procedure for eleminating the influence of the sample matrix in quantitative head-space GC. Chromatographia **15** (1982) 587–594.
[50] B. Kolb: Calibration techniques for quantitative analysis of organic volatile impurities using equilibration head-space gas chromatography, Pharm Forum **20** (1994) 6956–6960.
[51] J.P. Billon, F. Pellerin and J.H.McB. Miller: Assay of halogen salts of organic bases. Comparison of different assay methods, Pharmeuropa **1** (1988) 97–98.
[52] J.H.McB. Miller: Determination of halide salts of organic bases and quaternary ammonium compounds by titration, J Pharm Biomed Anal **7** (1989) 771–775.
[53] S. Ebel, B. Meyer and A. Wegener: Validation of titrations, Pharmeuropa **7** (1995) 133–136.
[54] E. Ebel, A.G.J. Daas, J.-P. Fournier, J. Hoogmartens, J.H.McB. Miller, J.-L. Robert, F. J. vander Vaart and J. Vessman: Inter-laboratory trials to assess a validation procedure for volumetric titrations, Pharmeuropa **12** (2000) 18–25.
[55] U. Rose and J. Fuchs: Validation of the value for specific absorbance in monographs of the European Pharmacopoeia, Pharmeuropa **10** (1998) 146–148.
[56] ICH Guideline, Specifications QC6, I.F.P.M.A. Geneva.
[57] Z. Zhu, E. Roets, Z. Ni, M.L. Morens, E. Porgueras and J. Hoogmartens: Evaluation of liquid chromatographic methods for the separation of ampicillin and its related substances, J Liq Chrom and Rel Technol **19** (12996) 1893–1908.
[58] Z. Zhu et al.: J. Liq Chrom and Rel Technol **19** (1996) 3305–3314.
[59] Z. Zhu et al.: Journal Pharm Biomed Anal **14** (1996) 631–639.
[60] Z. Zhu, M.L. Moreno, E. Porquero, E. Bourke, A. Bruzzi, M. Aletrari, P. Kanari, D. Partasidou, J. Nienhaus, W. Ferigo, J.L. Robert, J.H.McB. Miller,

J.-M. Spieser, E. Roets and J. Hoogmartens: Interlaboratory study of the analysis of ampicillin by liquid chromatography, Journal Pharm Biomed Anal **14** (1996) 1151–1156.

[61] N. Cillat , G. Decristo, P. Inarma, R.Macas, A. Mayhofer, J.H.McB. Miller, C. van der Vlies and M. Weiser: Liquid chromatographic method for the assay of benzathine benzlypenicillin and the determination of impurities, Pharmeuropa **10** (1998) 510–518.

[62] J.H.McB. Miller, H. Binder, R. De Richter, G. Decristoforo, A. Mayrhofer, E. Roets, S. Saarlo and C. van der Vlies: Collaborative study to assess the reproducibility of a reverse-phase LC method to determine the content and to estimate the impurities of benzathine benzylpenicillin, Pharmeuropa **12** (2000) 3–7.

[63] Z.Zhu, A. Verhasselt, E. Roets, A. Percz, E. Porqueras and J. Hoogmartens: Evaluation of liquid chromatography methods for the analysis of benzylpenicillin and its related substances, J Chromatog A **773** (1997) 147–156.

[64] Z.Zhu, E. Roets, J. Loidel, P. Inama, A. Perez, E. Porqueras, E. Wachberger, C. Lebbe, A.J.W. Smelt, C. van der Vlies, J.H.McB. Miller, J.M. Spieser and J. Hoogmartens: Interlaboratory study of the analysis of benzylpenicillin by liquid chromatography, J. Chromatog. B**707** (1998) 189–193.

[65] J.H.McB. Miller and C.J. Nap: Development of a liquid chromatographic method for assay of content of oxacillins in bulk drug and a test for related substances, Pharmeuropa **8** (1996) 187–192.

[66] J.H.McB. Miller, W. Backe, J. Diago, M.G. Kister, A. Mayrhofer and E. Porquero: A collaborative study of a liquid chromatographic method for the assay of content and for the test for related substancesof oxacillins. Part I, Cloxacillin sodium. Pharmeuropa **8** (1996) 427–434.

[67] J.H.McB. Miller, T. Burat, D. Demmer, K. Fischer, M.G. Kister, A. Kleman, F. Partenheimer, E. Souli and L. Thomsen: Collaborative study for a liquid chromatographic method for the assay of the content and for the test for related substances of oxacillins, Part III, Dicloxacillin sodium, Pharmeuropa **9** (1997) 129–134.

[68] J.H.McB Miller, A. Aranda, J. Burat, K. Fischer, A. Kleman, M.G. Kister and A. Mayrhofer: A collaboratorive study of a liquid chromatographic method for the assay of content and for the test for related substances of oxacillins. Part II, Flucloxacillin sodium, Pharmeuropa **8** (1996) 434–444.

[69] Z. Zhu E. Merken, M.P. Arevalo, E. Porqueras, E. Roets, J. Hoogmartens, Evaluation of LC methods for the separation of phenoxymethylpenicillin and its related substances. Eur. J. Pharm. Sci. 6, (1998) 39–45.

[70] Z. Zhu, E. Roets, B. Trippen, C.P. Christiansen, M.P. Arevalo, E. Porqueras, B. Maichel, P. Inama, S. Soderholm, J.H.McB. Miller, J.M. Spieser and J. Hoogmartens: Interlaboratory study for the analysis of phenoxymethyl penicillin by liquid chromatography, Chromatographia **47** (1998) 152-156.

[71] G. Kister, J. De Beer, S. Lauktien, J.H.McB. Miller, L. Thomsen and L. Turakka: Collaborative study of a liquid chromatographic method for the assay of the content of pivamicillin sodium, Pharmeuropa **9** (1997) 732-733.

[72] J.H.McB. Miller, T. Burat, C.P. Christiansen, N. Cillat, J. Hoogmartens, P. Inama, R.K. De Richter, A. Mayrhofer, E. Roets and K. Sinivuo: A collaborative study of a liquid chromatographic method for the assay of content of procaine and benzylpenicillin in procaine benzylpenicillin and the test for related substances, Pharmeuropa **10** (1998) 149–153.

[73] J.H.McB. Miller and C.J. Nap: Liquid chromatographic method for the assay of ticarcillin sodium and the determination of impurities, Pharmeuropa **8** (1996) 179–186.

[74] J.H.McB. Miller , J. De Beer, V. Hartofylax, J. Hoogmartens, M.G. Kister, B. Serckx and Y. Zhy: A collaborative study of a liquid chromatographic method for the assay of content and for the test for related substances of ticarcillin sodium, Pharmeuropa **8** (1996) 443–450.

[75] 5.12 Pharmaceutical Reference Standards, Pharmeuropa **16**, (2004).

9
Analytical Procedures in a Quality Control Environment

Raymond A. Cox

One of the most important aspects of controlling a manufacturing process and assuring that a quality product is produced, is the proper use and application of the testing procedures that were developed and validated during the Research and Development stage [1]. The basic assumption is that all the test procedures were properly developed to assure ruggedness and validated to identify those variables associated with the test procedure that must be monitored and controlled. If this assumption is not true, the task of controlling the manufacturing process and assuring the ultimate quality of the product becomes increasingly difficult. All good quality control laboratories will have access to a validation report and will assure themselves that it is available and complete, with all supporting data, spectra, and representative chromatograms. This becomes an invaluable tool when investigating any discrepancies or variances discovered during use of the method.

This chapter does not give exhaustive text on each of the topics discussed but should provide a sound basis for understanding the principles. The reader is encouraged to utilise other references, which go into greater detail and offer more explicit examples.

9.1
Monitoring the Performance of the Analytical Procedure

There are many ways to monitor the performance of the analytical procedure being used. These range from the use of check or control samples with known values, to recording the output or variable attribute of the instrumentation used and the use of titration blanks. Additionally, a system suitability test can be used to monitor many test procedures.

9.1.1
Utilization of Blanks

For titration analysis procedures, evaluation of a blank should be included. This is commonly referred to as a blank titration. Any response due to the blank is corrected for when the sample is analysed. If there is a problem with the reagents or, in some

Method Validation in Pharmaceutical Analysis. A Guide to Best Practice. Joachim Ermer, John H. McB. Miller (Eds.)

ISBN: 3-527-31255-2

cases, if the instrument is not functioning properly, this can be detected when the blank is run.

Ultraviolet absorptivities measurement also utilizes a blank to correct for or to zero out any absorptivity resulting from the reagents and solvent.

Solvent blanks are also utilized in high-performance liquid chromatographic analysis to correct for any interferences arising from peaks due to impurities in the solvent [2].

9.1.2 System Suitability Test Parameters and Acceptance Criteria

System suitability test parameters and acceptance criteria are based on the concept that the equipment, electronics, analytical operations, and samples to be analysed, constitute an integrated system [3]. System suitability testing ensures that the system is working properly at the time of analysis and is recommended as a component of any analytical procedure, not just those that involve chromatographic techniques. The system suitability tests selected should be based on the type of analysis, the intended use of the procedure and the stage of development. Regardless of the type of analytical procedure, testing should be used to confirm that the system will function correctly independent of the environmental conditions. Chromatographic analysis systems must have a system suitability requirement. For further details, see also Section 2.8.

9.1.2.1 Chromatographic and Quantitative System Suitability Parameters

System suitability parameters included in chromatographic procedures are needed to ensure the quality of separation. Quantitative methods also require parameters for variability. The exact subset of parameters selected for a given method will be based on the type of analysis, the intended use of the method, and the stage of development. Acceptance criteria should be established based on historical data observed during method development, method validation, and method transfer, or should be otherwise justified. Before any test data from a chromatographic analysis can be used, system suitability test requirements must be met. Parameters typically evaluated for chromatographic system suitability testing are listed in Table 9-1 [2–5].

Table 9-1 System suitability parameters.

Parameter	Recommendation [2–5]
Trailing/asymmetry factor	≤ 2.0
Capacity factor	Generally > 2.0
Theoretical plates	Generally > 2000
Resolution	Generally > 2.0
Relative Standard Deviation (RSD)	USP: RSD < 2.0 for $n = 5$
	Ph. Eur.: Dependent on values of n
Limit of Detection (LoD)	0.03 %
Limit of Quantitation (LoQ)	0.05 %

Performance of proper system suitability tests during the analyses also ensures both Operational Qualifications (OQ) and Performance Qualifications (PQ) which are part of the concepts of analytical quality assurance (see Chapter 4). The chromatographic systems are qualified routinely through the concepts of system suitability in chromatography.

9.1.3 Use of Check or Control Samples

One of the simplest means to monitor the performance of an analytical procedure is to use a check or control sample that has a well-established known value for the attribute which is being monitored. For monitoring any parameters such as colour, particle size, impurity level, or assay, one just needs to establish a typical production sample as the check or control sample and then document its suitability for use for this purpose.

9.1.3.1 Closure and Storage Conditions for Check or Control Samples

There are several items that need to be addressed in order to use check or control samples. One is to determine what storage conditions and container closure system is needed in order to protect the material from change. Typically, a more protective container closure system along with a more protective environmental storage condition is a good way to help ensure protection of the check or control sample. Usually these conditions can be identified from the stability studies already conducted on the drug substance or drug product, from normal stability programs, and from the analytical validation of test procedures where material is stressed to validate the stability-indicating potential of the method. The primary environmental stress conditions that the material needs to be protected against are moisture, heat, and light. Most organic compounds are stable if protected from these three conditions.

9.1.3.2 Continued Suitability Testing

Once the container closure system and storage conditions are chosen, a program must be implemented to assess the continued suitability of this material for use as a check or control sample. This should be a documented program with well-defined quality systems to ensure integrity of the program and test data. This material is then suitable for use in monitoring the performance of the analytical procedure as long as the continued suitability testing supports its suitability. An example of data, which support the continued suitability of material for use as an assay check sample, is shown in Figure 9-1.

The continued suitability testing is accomplished by running the test on this sample every time the test is performed on this type of sample, or when a series of samples are tested. Instead of routinely running the control or check sample each time the analysis is run, it is also possible to periodically test the check or control sample if prior history has shown the process or procedure to have long-term assay stability. If the interval between running the check or control sample is extended to, say, weekly or monthly, one must remember that this extended length of time means a more exhaus-

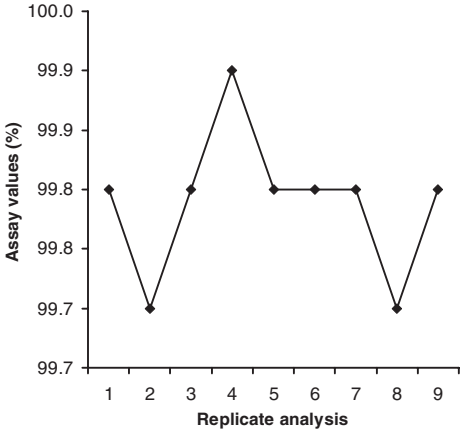


Figure 9-1 Continued suitability testing.

tive investigation, should the check or control sample results indicate a problem. The results are compared to the known value assigned to the check or control sample to see if the value obtained during the performance of the testing agrees with the assigned value within the experimental error of the method. The experimental error is determined from the validation data on the method intermediate precision and accuracy or by using historical data from previous testing on this sample. If the value obtained is within the expected range of the known value, it provides evidence that the analyst has performed the method properly and that the method is providing valid results within the validation parameters for that method. An example of data resulting from the use of a check sample for a perchloric acid titration is shown in Figure 9-2.

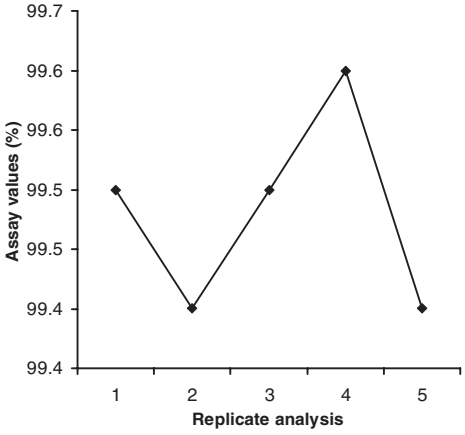


Figure 9-2 Perchloric acid titration values for check sample.

9.1.3.3 Utilization of Standard Preparation

Another variation of this is to save a previous preparation of the standard and assay it as a sample. This provides a convenient means of establishing a check or control sample which has a known value (the concentration at which the standard was pre-

pared) so this can be used to monitor the performance of the assay. In addition, it provides the stability data needed to support extended use of a stock reference standard preparation to help minimise the cost of preparation of reference standards. An example of data resulting from the use of a 10 mg/ml standard preparation as the check sample, is shown in Figure 9-3.

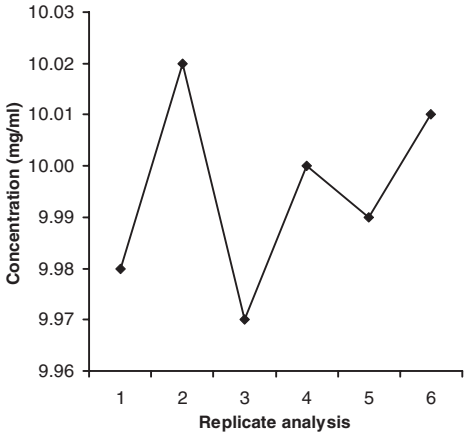


Figure 9-3 Utilisation of a 10 mg/ml standard preparation as control sample.

9.1.4 Analyst Performance

Another important monitor of the performance of an analytical procedure is the performance of the analyst. Laboratory errors occur when the analyst makes a mistake in following the method of analysis, uses incorrect standards, and/or simply miscalculates the data. The exact cause of analyst error or mistake can be difficult to determine specifically and it is unrealistic to expect that analyst error will always be determined and documented. Minimisation of laboratory errors is accomplished by assuring that the analyst is properly trained in quality systems and test procedure techniques [1].

9.1.4.1 Following Basic Operating Procedures

The quality systems include a thorough understanding of the importance of adherence to procedures and following all Basic Operating Procedures related to the laboratory operation, instrument operations, calibration, preventative maintenance, and documentation. It is important that the analyst be initially trained on these systems and procedures as well as retrained at appropriate intervals or whenever there is a significant change to any system or procedure.

A means of monitoring the laboratory performance of the analyst with respect to laboratory errors or deviations from established procedures is therefore needed. This can be accomplished through appropriate *Corrective Action and Preventative Action* (CAPA) programs. These programs identify and document any deviation or suspect test result, the investigation associated with it, the cause, and the corrective

and preventative actions taken to ensure that the result is valid and that any potential future occurrence of an incorrect result is prevented [11]. This information is analysed over time and trended to identify when there is a high probability that another problem might occur. Since this program identifies the cause of the problem as well as the analyst associated with the problem, it enables management to identify when additional analyst training or other corrective action is needed. Perhaps the procedure is unclear or the method's ruggedness is questionable. If this is the case, a new more rugged and precise method should be developed and validated.

9.1.5 Instrumental Performance

Instrumental performance is also a key factor in monitoring the performance of instruments used with an analytical procedure. Instrument performance is dependent on proper use and care of the instrument as provided in the manual and the related basic operating procedure. Instrument performance can be monitored using instrumental outputs associated with the type of instrument used. In high-pressure liquid chromatography, this may be the absorbance output of the ultraviolet detector or the column pressure experienced by the pumping system.

The *instrument manual* is the most important document associated with any instrument. It not only gives all the specifications for the instrument, but lists such key items as the physical location and environmental conditions under which the instrument will operate properly and the maintenance and preventative maintenance needed to maintain performance, in addition to key calibration ranges and the proper cleaning of the instrument.

As well as the instrument manual, it may also be appropriate to have *standard operating procedures* which give much greater detail on the use and operation of the instrument with respect to a specific analytical procedure. The standard operating procedure could be used to clarify key points in the use of the instrument and the interpretation of results.

Another example is the monitoring of *chromatographic system performance* by plotting the performance of the column with respect to variables , which could change with a deterioration in the column performance. These could include theoretical plates, column pressure, changes in mobile phase composition needed to obtain system suitability, or absolute retention times for components. As can be seen from Figure 9-4, the absolute retention time for the selected component separated, is increasing. This suggests that the column performance is changing and the analyst should start to investigate why this change is occurring. The change could be caused by a leaking pump, an impurity build-up on the column, thus changing its column efficiency, or evaporation of the organic modifier in the mobile phase.

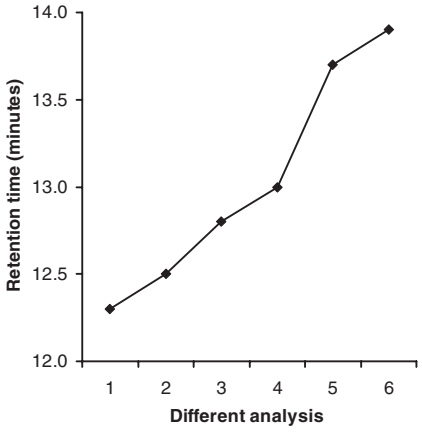


Figure 9-4 Example of tracking absolute retention time.

9.1.6
Reagent Stability and Performance

Reagent stability and performance is critical for the proper performance of the analytical procedure. This includes standardised titrants or any reagent or supply that can deteriorate in performance with time or improper storage and use, and chromatographic columns. A good example of monitoring reagent stability and performance is plotting the standardisation of acid/base solutions. If one of these reagents is routinely standardised by each analyst that uses it, a simple plot of the average standardised value versus time or the replicate values versus time for each analyst, can provide information on the stability of the reagent and the performance of the analyst. One would see a trend up or down or significant differences in values obtained by one analyst versus another.

9.1.7
Internal Limits and Specifications

The use of internal limits and specifications [6] is also a useful tool to monitor the performance of an analytical procedure. These are expected ranges or tighter limits than the regulatory limits for product release and stability testing. They are useful to trigger an investigation whenever a result approaches these limits or slightly exceeds them, but is still within the regulatory limit. Internal limits should be established once the operating performance of an analytical procedure is determined. This means that, once there is enough data to establish normal and acceptable ranges for results and the expected variability of the procedure is calculated, an internal limit would be set, taking into account this normal variability [12, 13]. This would usually be by setting the limit at +/- 2 standard deviations determined from the average value of multiple-lot assays. This tool is only useful for manufacturing processes which are under control so that the process produces a product with a consistent

quality. If the process itself were variable, the limits suggested would be so large that its utility would be questionable. Under these circumstances, other controls, such as more replicate analysis using the means as the variable to monitor performance, would be more appropriate.

9.2 Use of Control Charts

Control charts [9, 10] are simply a table of results as a graphical presentation of the data on the y-axis for each result or sample identification on the x-axis. Control charts are extremely valuable in providing a means of monitoring the total performance of the analyst, the instrument, and the test procedure and can be utilised by any laboratory. The statistical description of the stability of data over time requires that the pattern of variation remain stable, not that there should be no variation in the variable measured. A variable that continues to be described by the same distribution when observed over time is said to be in statistical control, or simply in control. Control charts work by distinguishing the natural variation in the process from the additional variation, which suggests that the process has changed. A control chart sounds an alarm when there is too much variation.

9.2.1 Examples of Control Charts

Control charts can be made from absolute data, data ranges, standard deviations for replicate analysis, and CUSUM data. The different types of charts [9] are often classified according to the type of quality characteristic that they are supposed to monitor: there are quality control charts for variables and control charts for attributes. Specifically, the following charts are commonly constructed for controlling variables.

- *X-bar chart.* In this chart the sample means are plotted in order to control the mean value of a variable (e.g., size of piston rings, strength of materials, etc.).
- *R chart.* In this chart, the sample ranges are plotted in order to control the variability of a variable.
- *S chart.* In this chart, the sample standard deviations are plotted in order to control the variability of a variable.
- S^2 *chart.* In this chart, the sample variances are plotted in order to control the variability of a variable.
- Cumulative Sum (CUSUM) Chart. If one plots the cumulative sum of deviations of successive sample means from a target specification, even minor, permanent shifts in the process mean will eventually lead to a sizeable cumulative sum of deviations. Thus, this chart is particularly well-suited for detecting such small permanent shifts that may go undetected when using the X-bar chart.

9.2.1.1 Production Process Out of Control

Figure 9-5 is an example of using a control chart of absolute data for the assay of a production process which is out of control. There is a continuous downward trend in purity values. This could be caused by a change in the purity of the material with succeeding lot production.

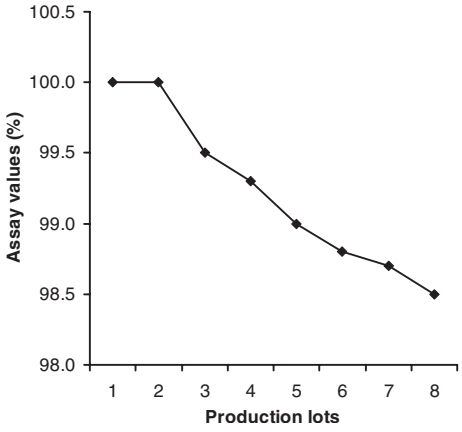


Figure 9-5 Example of production process out of control.

9.2.1.2 Shift in the Quality of Production Process or Change in Assay

Figure 9-6 is an example of the use of a control chart to detect a shift in the quality of the production lots. There is a distinct difference between the purity values for samples 1–5 compared with samples 6–9. This could be caused by a change in the manufacturing process or by a change in the assay procedure, i.e., a different reference standard being used.

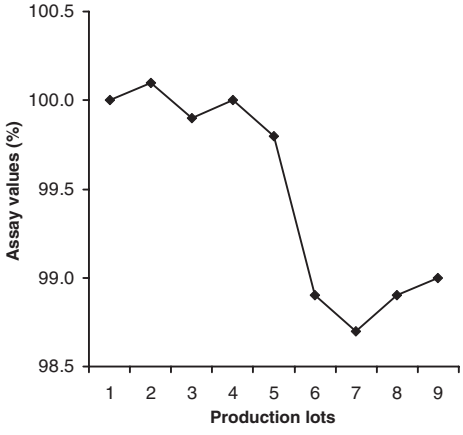


Figure 9-6 Example of a shift in production process quality or change in assay.

9.2.1.3 Procedure Out of Control

An example of a procedure out of control is as follows. For duplicate analysis, values of 98.5 and 99.8 percent are obtained for Sample 106 in Table 9-2. The specifications are not less than 98.0 and not more than 102.0 %. These results are within the specification limit, but when one looks at these results in comparison with results obtained from previous data one can see that there is something unusual about these results (see Table 9-2). The process average was running 99.5%. The current assay had an average of 99.2% – not far from the process average and well within the specification limit. If one only looked at the individual results and the average values, nothing in the data would necessarily trigger a concern. All results pass and the average value passes.

Table 9-2 Tabulation of assay results for production lots.

	Assay 1	Assay 2	Average value	Absolute difference
Sample 101	99.4%	99.5%	99.5	0.1%
Sample 102	99.5%	99.6%	99.6	0.1%
Sample 103	99.3%	99.5%	99.4	0.2%
Sample 104	99.6%	99.7%	99.7	0.1%
Sample 105	99.4%	99.6%	99.5	0.2%
Sample 106	98.5%	99.8%	99.2	1.3%
	Process average		99.5%	

However, if one were also looking at the difference between duplicate results, a significant difference for one set of data can be detected (Table 9-2). The individual results suggest that something might be out of control. The agreement within the individual set of results for sample 106 is different than with the other five sets of data. This set has an absolute difference of 1.3% compared with the largest absolute difference of the other sets of data of 0.2%.

Using just a simple table of results, one is able to detect a set of results which suggest that a problem has occurred with the assay and that something is starting to go out of control. Of course the critical factor is to know what to look for and track. In this example, one of the key pieces of information is the absolute difference between results since only this attribute gives an indication of a potential problem. A simple control chart of this same information offers the user the ability to detect a change before the change gets out of control as indicated in Figure 9-7. It is obvious that something is wrong. Either there was a small weighing error or one of the reagents or test conditions is starting to degrade. It is unlikely that there is a change in the product, but prudence dictates that a review be made of the manufacturing process to be sure that it has not changed. This early investigation can avoid significant costs, which would be incurred if a lot of material had to be re-tested and reworked or destroyed, because the change was out of control.

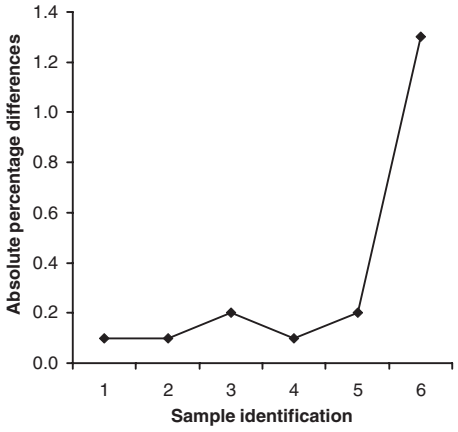


Figure 9-7 Example of variability of simple sample out-of-control test results.

9.2.2 Population in Control Charts

The population in the control chart setting consists of all items that would be produced or tested by the process if it ran on forever in its present state. The items actually produced form samples from this population. We generally speak of the process rather than the population. We choose a quantitative variable, such as assay, that is an important measure of the quality of an item. The process mean μ is the long-term average value of this variable; μ describes the centre or aim of the process. The sample mean x of several items estimates μ and helps us judge whether the centre of the process has moved away from its proper value. The most common control chart plots the means x of small samples taken from the process at regular intervals over time as shown in Figures 9-5 and 9-6.

9.2.3 Cost of Control Charts

There is practically no cost associated with tabulated lists of the historical data or simple data control charts, and the payback is an early warning of change, which could impact the validity of the data. Of course more sophisticated control charting programs are available, where the data can be automatically sent to a program that is capable of statistically analysing the results and providing reports. The draw-back to these is the cost. In addition to the cost of program development or purchase price for commercially available programs, the systems need to be validated. The validation costs can add significantly to the total cost of using these programs but once installed and validated, they are less influenced by operator error.

9.3 Change Control

Every manufacturing process, all test procedures, calibration, preventative maintenance, and documentation, need to have change control [1]. This is critical to make sure that the current validated system and all documents related to it remain within the validation parameters at any time during use. If there is any change to any of these systems, this must be documented. Included in the documentation should be a note of what was changed, why the change was made, who authorised the change, the impact of the change, which systems need to be revalidated as a result of the change, an evaluation of the change itself, and the final approval of the change. Quality Assurance must be part of the change control approval process.

Here we will only address change control as it relates to test procedures, calibration and preventative maintenance, and documentation.

9.3.1 Basic Elements of Test Procedure Change Control

There are five basic elements associated with change control. These are protocol, protocol approval, evaluation and validation, documentation, and final approval.

When a test procedure itself, the calibration and preventative maintenance, and the documentation are changed significantly, or a new test procedure is developed and validated to replace an existing test procedure, this change must be authorized and approved. This process is started by writing a protocol outlining the proposed change, the procedure to be followed to evaluate the change, the results which are expected as a result of the change, and the approvers of the change [1].

9.3.1.1 Protocol

The protocol would be written by either Research and Development or laboratory management, but would be reviewed and approved by all the units involved. The protocol must include specific requirements for any validation work and must specify the actual test procedure to be followed. The protocol could have a first review and approval after initial laboratory study, with an addendum protocol added, reviewed and approved for a modified test procedure if it is discovered that a modified procedure would be more appropriate. If validation were needed, this would be completed by either the Research and Development laboratory or could be done by the Quality Control laboratories. All procedures and data must be documented, reviewed and approved. Once the change or new procedure is documented, it must be transferred into the Quality Control laboratory. This must be documented with a protocol, which would include the samples to be tested, and the expected results. The protocol is approved prior to performing any testing for the transfer (see Chapter 7).

9.3.1.2 Protocol Approval

The approvers are usually Research and Development, manufacturers, laboratory management, and Quality Assurance. Research and Development are involved

because this is the group that normally validates changes to test procedures or develops and validates new test procedures. Manufacturers are involved because they are responsible for the product which is evaluated and released by the testing, and because modification or a new procedure might possibly give new information about the quality of the product. Laboratory management are involved because they have responsibility for implementing the change and they must ensure that the laboratory personnel are properly trained and that the laboratories have the necessary instruments and reagents needed to run the test. Quality Assurance are involved because they oversee product quality and are ultimately responsible for product release.

9.3.1.3 Evaluation and Validation

Based upon protocol design, any validation required along with any requirements for data generation should be completed and documented. The validation report and data documentation could be a separate report attached to the protocol for final review and approval. Once the method has been transferred and the results of any testing documented and compared with the expected results, this is again reviewed and approved to authorise implementation of the change.

9.3.1.4 Documentation

It is critical that all aspects of change control be documented [1]. This is needed to ensure that other competent analysts will be able to follow the same path and that the proper review and approval has been completed at the appropriate time.

9.3.1.5 Final Approval

The final step is changing and approving the appropriate operating documents, i.e, standard control procedures and specifications. This approval must be done by all the parties which approved the initial protocol, research and development, manufacturing, laboratory management, and quality assurance.

9.3.2 Change Control for Calibration and Preventative Maintenance

Change control for calibration and preventative maintenance (PM) follows a similar scenario. Calibration and preventative maintenance are initially established by the use of a calibration (also termed Performance Qualification, see Chapter 4) or preventative maintenance, request. This request must include the proper identification of the instrument or system including its location, the area responsible for the calibration, and the procedures to be used for calibration. The attributes to be calibrated and the frequency must be indicated with the calibration limits set, based upon the instrument manufacturer's recommendations for the specific use of the instrument to provide the needed accuracy for that use. An example would be a balance being used only to weigh out reagents for preparation of solutions capable of accuracy to ± 2 mg but the reagent weight might only need to be within ± 100 mg. It would not make sense to demand tolerances of ± 2 mg when all one needs is ± 100 mg. If the calibration of the balance is only to ± 100 mg, it must never be used to weigh masses

where accuracy to ±2 mg is needed. Of course you might argue that a less costly balance could have been purchased and used, which only provided accuracy to ±100 mg.

Once the instrument is calibrated or preventative maintenance has been performed, this is documented and reviewed and approved by both the calibration/maintenance area and quality assurance. All calibrations and preventative maintenance must have specified intervals for re-calibration or preventative maintenance. Any future calibration or PM must be performed before the indicated interval, or the instrument or equipment must be taken out of service until this is completed.

9.3.3
Future Calibration and Preventative Maintenance

If future calibration or PM indicate that the output of the instrument or equipment is outside the acceptable tolerance and must be repaired or adjusted, this repair or adjustment must be approved by Maintenance and Quality Assurance. An investigation and *impact analysis* must be done to assure that the out-of-tolerance condition did not allow for the release of any product which would not have met specifications had the instrument or equipment been within tolerances. This investigation and impact analysis must cover each of the lots of material released since the last approved calibration or PM and must be documented.

9.4
When is an Adjustment Really a Change?

The European Pharmacopoeia has addressed acceptable adjustments to quantities being weighed for analysis purposes. This acceptable weight range for quantities being accurately weighed for an assay or test procedure is within ±10 percent of the stated mass in the test procedure. This means that, if an assay procedure states to perform the test on a 100 mg sample, it would be acceptable to use any mass between 90.0 mg and 110.0 mg as long as the actual mass taken was known. Some laboratories have extrapolated this allowed variance to mean that other attributes such as temperature, time, and in some cases volumes, could be changed, as long as the change was within ±10 percent of the stated value. This is not specifically allowed for in the Pharmacopoeia and any of the changes must be validated, if not covered in the robustness studies (see Section 2.7). Therefore, if one had a test procedure which required a mixing time of 30 minutes, one must either mix for 30 minutes or validate that a different mixing time gives equivalent results. Validation would also be needed for changes in temperature and volumes, if applicable.

9.4.1 Chromatographic Adjustments versus Changes

The only other test procedure with a suggested allowable adjustment is a chromatographic analysis. The European Pharmacopoeia published an article in the Reader's Tribune on System Suitability [7]. This article discussed how the system suitability test, if designed to control critical separation of components of the sample, could also be used to allow for adjustments in the operating conditions in order to obtain satisfactory system suitability, even when parameters of the procedure were changed. It should be pointed out that these allowed ranges should be addressed as part of the validation of robustness (see Section 2.7) to document acceptable performance of the method when adjustments are made.

9.4.1.1 Typical Attribute Adjustments Allowed

To obtain system suitability without changing the method design, certain parameters of the chromatographic system may be varied, prior to determining system suitability. The magnitude of the allowed changes should be judged by the adjusted system's ability to separate the desired components, and are not recommended to supersede the method validation (i.e., robustness) ranges. The parameters which are usually adjusted to obtain system suitability are:

- pH of the mobile phase (± 1 depending on pKa of analyte);
- the concentration of salts in the buffer (± 10 percent);.
- the ratio of solvents in the mobile phase (± 30 percent relative or ± 2 percent absolute, whichever is larger);
- the column length (± 70 percent);
- the column inner diameter (± 25 percent);
- the flow rate (± 50 percent);
- the particle size of the stationary phase (may be reduced by up to 50 percent);
- the injection volume (may be increased by up to 2 fold or reduced);
- the column temperature (± 10 percent for GC: ± 40 °C for LC);
- the oven temperature program, GC (± 20 percent).

9.5 Statistical Process Control (SPC)

It is the responsibility of management to reduce common cause or system variation as well as special cause variation [8]. This is done through process improvement techniques, investing in new technology, or re-engineering the process to be more rugged as well as more accurate and precise. Control charts and statistical process control are used to identify successful process improvements, advantageous new technology, and process re-engineering which produces a better quality product or higher yields.

Process improvement techniques form an entire subject by themselves. There are many references and training seminars for this topic (e.g. [14]).

Investing in new technology could be a new mixing technology or drying apparatus. Other examples might include cell culture production versus fermentation.

Re-engineering the process involves looking at the process steps to see what changes could be made to produce a better product in higher yields. It could involve such things as developing a continuous process or changing the particle size or coating process to give a formulation which has better dissolution characteristics.

9.5.1 Purpose of Control Charts

The purpose of a control chart is not to ensure good quality by inspecting most of the items produced. Control charts focus on the manufacturing process, in-process controls, raw material quality, intermediate step quality, final active ingredient, and final product. Process here means any aspect of testing, calibration, and preventive maintenance. By checking the process at regular intervals, we can detect disturbances and correct them quickly. This is called statistical process control.

9.5.2 Advantages of Statistical Process Control

A process that is in control is stable over time, but stability alone does not guarantee good quality. The natural variation in the process may be so large that many of the products are unsatisfactory. Nonetheless, establishing control brings a number of advantages.

- In order to assess whether the process quality is satisfactory, we must observe the process operating in control, free of breakdowns and other disturbances.
- A process in control is predictable. We can predict both the quantity and the quality of items produced.
- When a process is in control we can easily see the effects of attempts to improve the process, which are not hidden by the unpredictable variation which characterises a lack of statistical control.

A process in control is doing as well as it can in its present state.

9.6 Revalidation

Revalidation must be performed whenever there is a significant change, to ensure that the analytical procedure maintains its characteristics (for example, specificity) and to demonstrate that the analytical procedure continues to ensure the identity, strength, quality, purity, and potency of the drug substance and drug product, and the bio-availability of the drug product [1]. The degree of revalidation depends on the nature of the change.

If, during each use an analytical procedure can meet the established system suitability requirements only after repeated adjustments to the operating conditions stated in the analytical procedures, then the analytical procedure must be re-evaluated and amended, and that amendment revalidated, as appropriate.

Test procedures must be reviewed periodically (this could be a five-year interval) from the last documented and approved validation or change control, to determine whether anything has changed since the last validation or documented approved change control. This can easily be done by reviewing the data from the lot history and comparing the current test procedure with the validated test procedure to see if any changes have crept in, or reviewing the change history for that test procedure to ensure that all changes were properly executed. If there are no abnormalities and all documentation is complete, the only thing needed is to document this review and the acceptable performance of the test procedure up to the time of the evaluation. Obviously, if something were detected which would indicate a change, the change control procedure would have to be followed. This 'paper' review is needed to ensure that no changes have inadvertently occurred over time without being detected.

Table 9-3 Revalidation requirements for changes.

Changes	Degree of revalidation
In the synthesis of the drug substance	
Different synthetic route	Major[1)]
Different purification solvent	Major
Different manufacturing solvent	Minor[2)]
In the composition of the finished product	
New excipient	Major[1)]
Intermediate dose size	Minor[3)]
Change to particle coating	Intermediate[4)]
In the analytical procedure	
New method	Major[1)]
Variation in chromatographic parameters within original validation limits	None
Change in mobile phase composition of organic solvent or modifiers	Major
In limits which the test method supports	
Control of new impurities	Intermediate[5)]
Significantly lower impurity limits	Major[1)]
Slightly higher impurity limits within validation limits	None
New applications of the test method	
Chromatographic analysis of a different compound	Major[1)]
Karl Fischer titration in new formulation	Minor[2)]
Head Space analysis of solvent residue in new formulation	Minor

1 see Chapter 1, Table 1.1 for validation characteristics
2 test for solvent levels in drug substance and related substances
3 verify method on known samples
4 validate for dissolution
5 show separation, run linearity and precision for new impurity

9.6.1 Revalidation Summary

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 analytical procedure.
- Changes in limits which the test method supports.
- New application of the test method.

The degree of revalidation required depends on the nature of the changes. Some examples of this are presented in Table 9-3.

Using control charts to routinely monitor the performance of a test procedure or manufacturing process, along with documented periodic review of the control charts, can negate the need for a periodic revalidation.

Acknowledgement

The author would like to acknowledge Sandra Sands for her invaluable help in the organisation, content structure, and editorial support in writing this chapter.

9.7 References

[1] International Conference on Harmonization Q7A Guideline: Good Manufacturing Practice Guide for Active Pharmaceutical Ingredients.
[2] International Conference on Harmonization Q2B Guideline: Validation of Analytical Procedures: Methodology.
[3] Chapter <621> Chromatography; US Pharmacopeia 23, United States Pharmacopeial Convention, Inc., Rockville MD 1994.
[4] International Conference on Harmonization Q2A Guideline: Text on Validation of Analytical Procedures.
[5] 2.2.46; European Pharmacopoeia, 4th Edition, Council of Europe, F-67075 Strasbourg, France.
[6] International Conference on Harmonization Q6A: Test Procedures and Acceptance Criteria for New Drug Substances and New Drug Products.
[7] Raymond Cox, Gopi Menon, System Suitability, Readers' Tribune, PharmEuropa 10.1, March 1998.
[8] Quality and Statistical Process Control, Prof Sid Sytsma , Ferris State University, [Unterstrichen]www.sytsma.com/tqmtools/ctlchtprinciples.html
[9] Copyright Statsoft, Inc. 1984–2003, Statistical Quality Control Charts, www.statsoft.com/qccharts.html
[10] W.H. Freeman: The Basic Practice of Statistics, Second Edition, 2000.
[11] Guide to Inspection of Quality Systems, FDA, ORA, August 1999.
[12] Taylor, J.K., Quality Assurance of Chemical Measurements, Lewis Publishers, Inc. 1987.
[13] Miller, J.C., J.N.Miller, and E. Horwood: Statistics for Analytical Chemistry, 3rd Edition, Prentice Hall, 1993.
[14] G. Vorley and F. Tickle: Quality Management Tools and Techniques.

10
Aberrant or Atypical Results

Christopher Burgess

10.1
Laboratory Failure Investigation

The purpose of an analysis of a sample for a particular analyte is to predict the value of that property for the entire lot or batch of product from which the sample was taken. Assuming that the sample is both representative and homogeneous, the sample is analysed using an analytical procedure. This procedure is itself a process, just as the manufacturing operation is a procedure [1]. All analytical measurements are subject to error. We are therefore faced with the situation of using one process (the analytical one) to judge the performance of another, the manufacturing process. Ideally we would like to use a measurement process which is infinitely precise and of known accuracy. If this were the case, any aberrant or atypical result (AAR) would be attributed to sampling or manufacturing process variation and not to the measurement process itself. From a regulatory perspective, the concern is primarily whether an out-of-specification result relates to the manufacturing process which would lead to batch rejection, or whether it results from some other assignable cause. The possible assignment of attributable cause is a major part of laboratory failure investigations as required particularly by the FDA [2]. Failure to identify or establish attributable analytical cause within the laboratory triggers a full-scale failure investigation (Fig. 10-1).

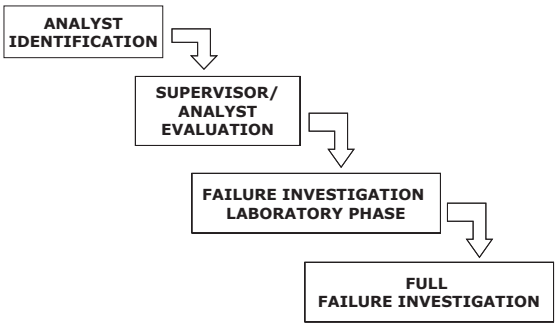


Figure 10-1 Stages for the investigation of atypical or aberrant results.

Method Validation in Pharmaceutical Analysis. A Guide to Best Practice. Joachim Ermer, John H. McB. Miller (Eds.)

ISBN: 3-527-31255-2

The role and responsibilities of the analyst and the supervisor are critical to the performance of within-laboratory failure investigations. The analyst's role and responsibilities are as follows:

1. The first responsibility for achieving accurate laboratory testing results lies with the analyst who is performing the test.
2. The analyst should be aware of potential problems that could occur during the testing process and should watch for problems that could create AARs.
3. The analyst should ensure that only those instruments meeting established specifications are used and that all instruments are properly calibrated [3] (see also Chapter 4).
4. Analytical methods that have system suitability requirements which, if not met, should not be used or continued. Analysts should not knowingly continue an analysis they expect to invalidate at a later time for an assignable cause (i.e., analyses should not be completed for the sole purpose of seeing what results can be obtained when obvious errors are known).
5. Before discarding test preparations or standard preparations, analysts should check the data for compliance with specifications.
6. When unexpected results are obtained and no obvious explanation exists, test preparations should be retained and the analyst should inform the supervisor.

The analyst's direct line manager or supervisor must be informed of an AAR occurrence as soon as possible. The supervisor is then involved in a formal and documented evaluation. Their role and responsibilities are as follows:

1. To conduct an objective and timely investigation and document it.
2. To discuss the test method and confirm the analyst's knowledge of the procedure.
3. To examine the raw data obtained in the analysis, including chromatograms and spectra, and identify anomalous or suspect information.
4. To confirm the performance of the instruments.
5. To determine that appropriate reference standards, solvents, reagents and other solutions were used and that they met quality control specifications.
6. To evaluate the performance of the testing method to ensure that it is performing according to the standard expected based on method validation data.
7. To document and preserve evidence of this assessment.
8. To review the calculation.
9. To ascertain, not only the reliability of the individual value obtained, but also the significance of these AARs in the overall quality assurance program.
   Laboratory error should be relatively rare. Frequent errors suggest a problem that might be due to inadequate training of analysts, poorly maintained or improperly calibrated equipment or careless work.
10. When clear evidence of laboratory error exists, the laboratory testing results should be invalidated.

When evidence of laboratory error remains unclear, a laboratory failure investigation should be conducted to determine what caused the unexpected results. This process could include the following points:

1. Re-testing the original solutions.
2. Re-testing a portion of the original laboratory sample – the decision to re-test should be based on sound scientific judgement.
3. Use a different analyst in conjunction with the original analyst.
4. A predetermined testing procedure should identify the point at which the testing ends and the product is evaluated. Testing into compliance is objectionable under the CGMPs.
5. If a clearly identified laboratory error is found, the re-test results would be substituted for the original test results.
6. The original results should be retained, however, and an explanation recorded.
7. The results and conclusions should be documented.

This chapter is concerned not only with out-of-specification analytical measurements, but also those that do not meet expectations or are discordant. In order to discuss whether or not a result is aberrant or atypical, it is firstly necessary to define what a result is and secondly to specify what constitutes typical behaviour. Once these criteria have been defined it is possible to review the methods available for detecting and evaluating atypical behaviour. We need to be concerned about AARs because, when they are included in our calculations, they distort both the measure of location (usually but not always the mean or average value) and the measure of dispersion or spread (precision or variance).

10.2 Basic Concepts of Measurement Performance

Analytical measurements are the outcomes of scientifically sound analytical methods and procedures. These methods and procedures are themselves dynamic processes. It is important to recognise that, when analyses are carried out with the objective of measuring manufacturing process performance, the problem is essentially of one process being used to assess another. For the purposes of this discussion we will ignore the sampling process and assume that the test sample, drawn from a laboratory sample from which the analytical signal derives, is representative of the lot or batch of material under test. In order to describe the characteristics of analytical measurements and results, a basic vocabulary of unambiguous statistical terms needs to be firmly established. Concepts such as accuracy and precision are widely misused and misunderstood within the analytical community [4]. The importance of a commonly agreed terminology cannot be underestimated. Figure 10-2 illustrates some of the basic concepts and definitions.

All measurements and responses are subject to error. These errors may be random or systematic, or a combination of both. As an example, we will assume that

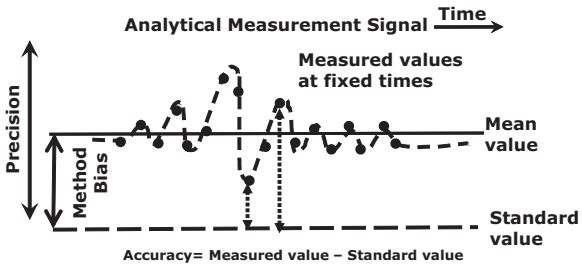


Figure 10-2 Basic definitions and concepts for analytical measurements.

the analytical measurement signal shown in Figure 10-2, represented as a varying black line, is the analogue voltage output from a UV spectrophotometric absorbance measurement of a sample solution. This signal is sampled or recorded as a series of measurement values, in time, represented by the dots. This might be by an A/D converter, for example. The amplitude of the natural and inherent variability of the instrument measurement process allows an estimate of the random error, associated with the measurement, to be made. The random error estimate is a measurement of *precision*. There are many types of precision (see Section 2.1.2). The one estimated here is the measurement or instrument-response precision. This represents the best capability of the measurement function.

As analytical data are found to be [5] or assumed to be normally distributed in most practical situations, *precision* may be defined in terms of the measurement variance V_m, which is calculated from the sum of squares of the differences between the individual measurement values and the average or mean value. For a measurement sequence of n values this is given by Eq. (10-1).

$$V_m = \sum_{i=1}^{n} (X_i - \bar{X})^2$$

and hence the standard deviation is given by

$$s_m = \sqrt[2]{V_m}$$

and the Relative Standard Deviation by

$$RSD = \frac{s_m}{\bar{X}} 100 \qquad (10\text{-}1)$$

Precision is about the spread of data under a set of predetermined conditions. There are other sources of variability within an analytical procedure and hence different measurements of precision from the one discussed above and these will be discussed later (Section 10.4). However, it should be noted that the instrumental or measurement precision is the best which the analytical process is capable of achieving. With increasing complexity, the additional variance contributions will increase the random component of the error.

Accuracy is defined in terms of the difference between a measured value and a known or standard value. In our example, this would be the difference between the measured absorbance value and the assigned value of a solution or artefact established by, or traceable to, a National Laboratory (for example, NIST or NPL). This definition implies that the accuracy of measurement varies across a measurement sequence and contains elements of both random and systematic error.

For this reason, it is best analytical practice to combine a number of measurements by the process of averaging in order to arrive at a mean value. Conventionally, the difference between this mean value and the standard or known value is called the *bias*. However, the International Standards Organisation (ISO) have defined a new term, *trueness* [6] to mean the closeness of agreement between an average value obtained from a large series of measurements and an accepted reference value. In other words, trueness implies lack of bias [7].

In addition the term 'accuracy' cannot be strictly applied to methods or procedures. This is because the outcome of such processes is subject to an estimate of measurement uncertainty [8]. This measurement uncertainty estimate contains contributions from both systematic and random errors and is therefore a combination of accuracy and precision components.

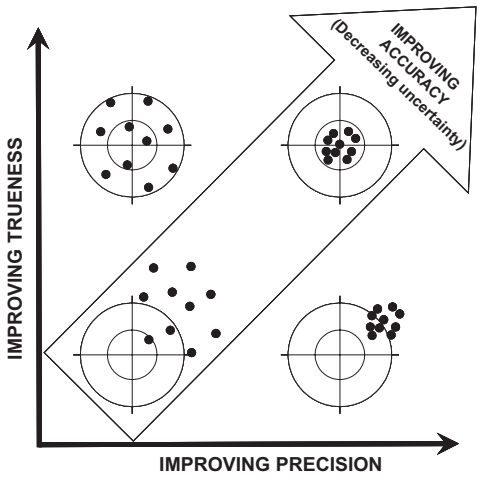


Figure 10-3 Accuracy, precision and trueness (redrawn from [4]).

Examination of Figure 10-3 reveals that the traditional method of displaying accuracy and precision, using the well known target illustration, is not strictly correct. It is *trueness* (or lack of bias) which is relatable to *precision* not *accuracy*.

10.3
Measurements, Results and Reportable Values

Thus far we have only considered the instrumental measurement process and basic statements of measurement performance. We need to extend these ideas into the

overall analytical process from the laboratory sample to the end result or reportable value [9]. The purpose of any analysis is to report upon the sample provided. This entails comparing the *reportable value(s)* relating to the sample and comparing it (them) to a set of limits (a specification). This implies that the selected analytical method or procedure is fit for its intended purpose.

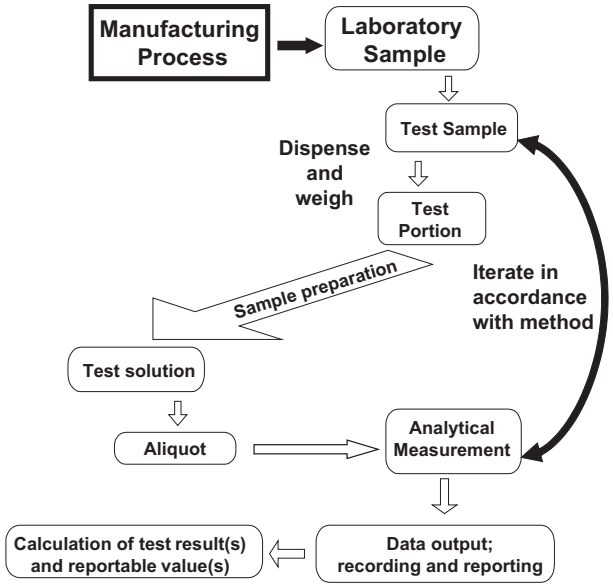


Figure 10-4: Analytical process flow.

From a regulatory perspective, 'fitness for purpose' means that all methods and procedures are validated and that this validation has been performed using equipment and systems which have been qualified and calibrated. In addition, all computerised systems involved in generating data and results have been subjected to adequate verification and validation.

Although the analytical measurement is at the heart of the analytical process, it is not the only source of error (systematic or random) which affects the overall trueness of the end result. Consider the analytical process flow shown in Figure 10-4. It is apparent that one analytical measurement does usually not constitute a route to a reportable value. Additionally, there are variance contributions which arise from other parts of the process, particularly in sample preparation and sub-sampling.

Generally speaking, analytical measurements are derived from the sampling of an analytical signal or response function. Analytical results are based upon those analytical measurements given a known (or assumed) relationship to the property which is required, such as a concentration or a purity value. *Reportable values* are predetermined combinations of analytical results and are the only values that should be compared with a specification.

An analytical method or procedure is a sequence of explicit instructions that describe the analytical process from the laboratory sample to the reportable value. Reportable values should be based on knowledge of the analytical process capability determined during method validation. This will be discussed in Section 10.5.

10.4 Sources of Variability in Analytical Methods and Procedures

Examination of Figure 10-4 reveals some of the additional sources of variability which affect an analytical method. The ICH Guidelines [10] define three levels of precision when applied to an analytical procedure that need to be established during method validation; i.e., repeatability, intermediate precision and reproducibility. The magnitude of these precisions increases with the order. In the laboratory, a fourth kind of precision is encountered, that of instrument or measurement precision. This represents the smallest of the precisions and is an estimate of the very best that an instrument can perform, for example, the precision obtained from a series of repeated injections of the same solution in a short space of time. This measurement of instrument repeatability is often confused with the ICH repeatability, which refers to a complete sample preparation.

The most important factors in the determination of repeatability, intermediate precision and reproducibility are, for a given method: laboratory, time, analyst and instrumentation [1] (Table 10-1).

Repeatability is the closest to the instrument precision discussed earlier. This is determined using a series of replicate measurements over a short time period (at least six at a concentration level, or nine if taken over the concentration range) on the same experimental system and with one operator. *Intermediate precision* is a measure of the variability within the development laboratory and is best determined using designed experiments. *Reproducibility* is a measure of the precision found when the method is transferred into routine use in other laboratories. The determination of reproducibility is normally achieved via a collaborative trial. The random error component increases from repeatability to reproducibility as the sources of variability increase (Table 10-1).

Table 10-1 Factors involved in precision determinations.

Type of precision to be determined	Factors to control	Factors to vary
Repeatability	L, T, A, I	
Intermediate precision (within-laboratory reproducibility)	L	T, I and A
Reproducibility (between-laboratory reproducibility)		L, T, A, I

Abbreviations:
L = laboratory
T = time
A = analyst
I = instrumentation

These measurements of precision are made either at one concentration or over a narrow range of concentrations. In the latter, it is assumed that the variance does not change over the concentration range studied. This is a reasonable assumption for analytical responses which are large. If the analytical responses approach the limit of quantitation, for example, with impurities, then this assumption should be checked using an *F* test for homogeneity of variances.

Analytical chemists have long been aware that the relative standard deviation increases as the analyte concentration decreases. Horwitz [11] at the FDA undertook the analysis of approximately 3000 precision values from collaborative trials which led to the establishment of an empirical function, $RSD = \pm 2^{(1-0.5\log C)}$, which when plotted yields the Horwitz trumpet. This function is illustrated in Figure 10-5 and clearly shows that the assumption of constant variance with concentration is only reasonable at high concentrations and narrow ranges.

These considerations lead us to the idea that analytical process capability is critical in defining an aberrant or atypical result.

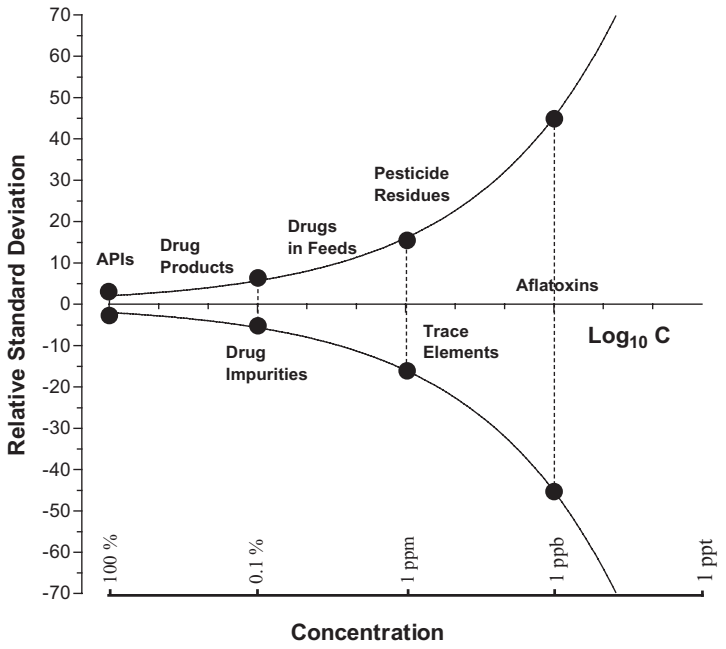


Figure 10-5 Horwitz 'trumpet' function.

10.5 Analytical Process Capability

Process capability is a statistical concept. It requires two things:

1. a knowledge of the randomness and trueness of the process;
2. a set of boundary conditions under which the process is required to operate.

The first of these requirements have been discussed in the first two sections. The second requirement is normally called a specification or tolerance limit. Our definitions of AARs will depend upon the type of boundary condition imposed on the process. The different types will be discussed in Section 10.6.

For the moment, let us assume a specification for release of a drug product of 95% – 105% of labelled claim of an active material. Let us also assume that the analytical method we are using to generate analytical measurements is unbiased, i.e., the mean of many results generates a 'true' value. The spread of results is indicated by the precision, as defined by a standard deviation, arising from all the sources of variability considered. The analytical process undertaken in shown in Figure 10-4.

In our example we will define that a reportable value is derived from a single analytical result. For our purposes, let us assume that the analytical process standard deviation lies between 1 and 3% (note that 2% is a value often found for HPLC methodologies, see Section 2.1.3.2). We use the symbol s as the estimate of the population standard deviation σ. This estimate is normally obtained from the intermediate precision.

We can now calculate what the distribution of (single) reportable values would look like by generating the normal distribution curves for each of the standard deviations and marking the upper and lower specification limits. The resulting plot is shown in Figure 10-6. By visual inspection, it is immediately apparent, without the necessity for further calculation, that if our analytical process had an $s = 1\%$ then we would be reasonably confident that, if a value lay outside the specification limits, it was unlikely to be due to the inherent variability in the method. In contrast, when $s = 3\%$, such a method would not be suitable because a large percentage (in this instance about 10.6%) would lie outside the limits due to the measurement process itself. Clearly it is scientifically unsound to attempt to monitor a manufacturing process with a defined analytical process which is not fit for that purpose. For $s = 2\%$ we have the situation where only a small amount of data will lie outside the limits (approximately 1.5%). So this begs the question: how good does our method have to be?

Any analytical method must be capable of generating reportable values which have a sufficiently small uncertainty to be able to identify variations within the man-

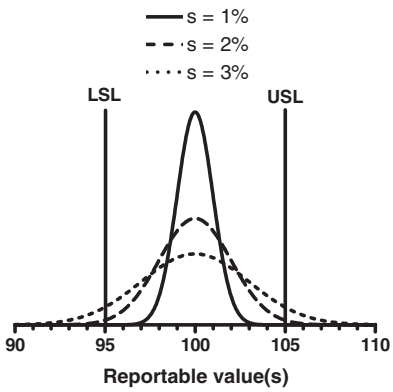


Figure 10-6 Simulation of (single) reportable values for s =1, 2 and 3%.

ufacturing process. This leads us naturally into measures for process capability. The process capability index, C_p, is calculated from Eq.(10-2).

$$C_p = \frac{USL - LSL}{6s} \tag{10-2}$$

If we substitute our values into this equation, it becomes:

$$C_p = \frac{10}{6} = 1.67 \text{ for s} = 1\%$$

$$C_p = \frac{10}{12} = 0.83 \text{ for s} = 2\%$$

$$C_p = \frac{10}{18} = 0.56 \text{ for s} = 3\% \tag{10-3}$$

From the theory of statistical process control [12] (SPC), it is known that, from a control viewpoint, the value of C_p can be used as a confidence measure (Table 10-2).

Table 10-2 Effectiveness indicators for analytical process capabilities, C_p.

Value of C_p	Effectiveness of control
<1	Hopeless
1	Poor and process not in statistical control
1.2	Marginal
1.6	Good
1.8	Very good
2.0 or more	Ideal

What happens if our measurements are biased either knowingly or, even worse, are unsuspected? Here we need to consider an alternative process capability measure C_{pk}.

$$C_{pk} = \frac{USL - \bar{X}}{3s} \text{ or } \frac{\bar{X} - LSL}{3s} \text{ whichever is the smaller} \tag{10-4}$$

Suppose that the process mean is biased by 1.5% from the target of 100%. The situation shown in Figure 10-6 is no longer the case and the plot becomes as in Figure 10-7. Such a bias would not be unusual if a recovery problem occurred.

Note that this shift could also be due to the manufacturing process itself. It is therefore important to establish analytical process capabilities on samples in which such effects have been minimised or eliminated.

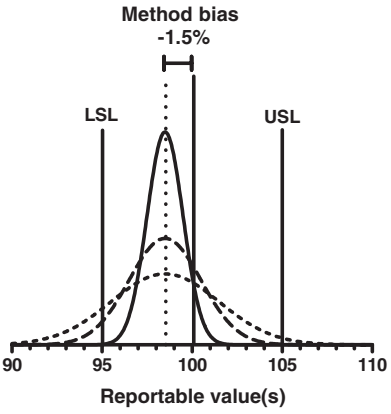


Figure 10-7 Process capability for a process mean of 98.5% (1.5% bias).

We can now recalculate C_{pk} based upon this new information.

$$C_{pk} = \frac{\bar{X} - LSL}{3s} \text{ as this is the smaller}$$

$$C_{pk} = \frac{98.5 - 95}{3s} \begin{array}{l} \text{which is 1.17 for s} = 1 \\ \text{which is 0.58 for s} = 2 \end{array}$$

and 0.39 for s = 3 (10-5)

Hence, a different situation is apparent even for s=1 for which the C_{pk} is now poor-to-marginal. Clearly we need to control bias as well as precision in assessing analytical process performance.

Thus far we have assumed that one analytical result equates with one reportable value. This is clearly not sensible in the light of the values for s found with chromatographic methods in the region of 2%. Indeed, it is part of analytical folklore that assays should be performed in duplicate! The combination of two results by taking the average to be the reportable value, is common practice. However, the important concept of the standard error of the mean, SEM, plays a large part in determining how to arrive at a scientifically sound reportable value. The SEM is defined as $s/\sqrt{n}$ where n is the number of replicates (i.e., the number of iterations shown in Fig. 10-4). We can use the value for the SEM to replace s in order to calculate C_p (or C_{pk}) for the number of replicates required, so that we can generate a reportable value which will be effective in controlling the manufacturing process (Table 10-3).

The degree of replication should be established and justified as part of the method- development process (see also Section 1.4.1 and Chapter 6).

Table 10-3 Effect of replication on process capability.

Replicates (*n*)	SEM	C_p
1	2.00	0.83
2	1.41	1.18
3	1.15	1.44
4	1.00	1.67
5	0.89	1.86
6	0.82	2.04

10.6 Classification of Atypical or Aberrant Results

We have seen that an AAR is identifiable as a value or measurement lying outside the range expected for a given process. The definition of this expected range is important in classifying AARs. The most common AAR from a regulatory perspective is the Out-Of-Specification or OOS result. However an OOS occurs only when the reportable value lies beyond the registered specification limits for release. However, all AARs are by definition outside pre-defined limits.

The other types of AAR are:

- OOT; Out-Of-Trend;

and

- OOE; Out-Of-Expectation.

The difference between these three types may be illustrated by way of an example. HPLC is the most widely used analytical technique in pharmaceutical analysis and is a ready source of AARs. Figure 10-4 illustrates the overall analytical process. The analytical measurement generator at the heart of this process is subject to many sources of variation and requires extensive control [13].

We will take as the example a simple external standard method for a single substance. The core of the HPLC analytical process is the sampling protocol and the injection sequence.

Let us assume that the analytical method specifies duplicate sample weighings to be taken from the laboratory sample and that, following the sample treatment, duplicate dilutions are taken. Each dilution is then injected twice onto the HPLC column. The sampling 'tree' is shown in Figure 10-8.

In addition, our analytical method requires the preparation of duplicate standard weighings at a single level and an analysis sequence using the 'rolling bracket' technique. In this commonly used technique, the sample injections are bracketed by both standards to minimise errors caused by chromatographic drift. The standard sequence is illustrated in Figure 10-9.

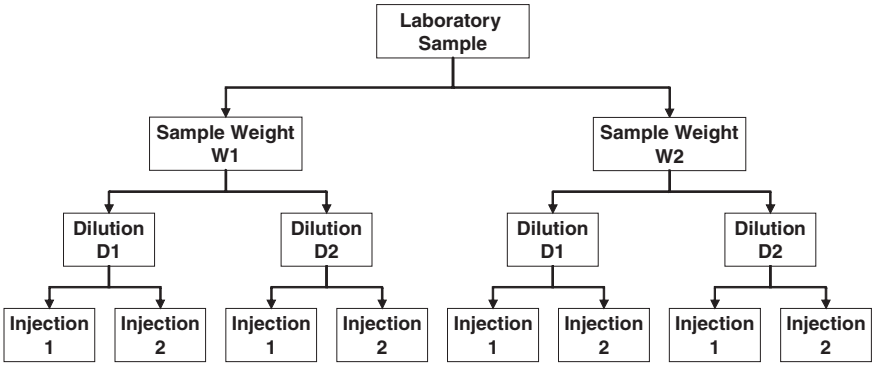


Figure 10-8 HPLC example: sampling diagram.

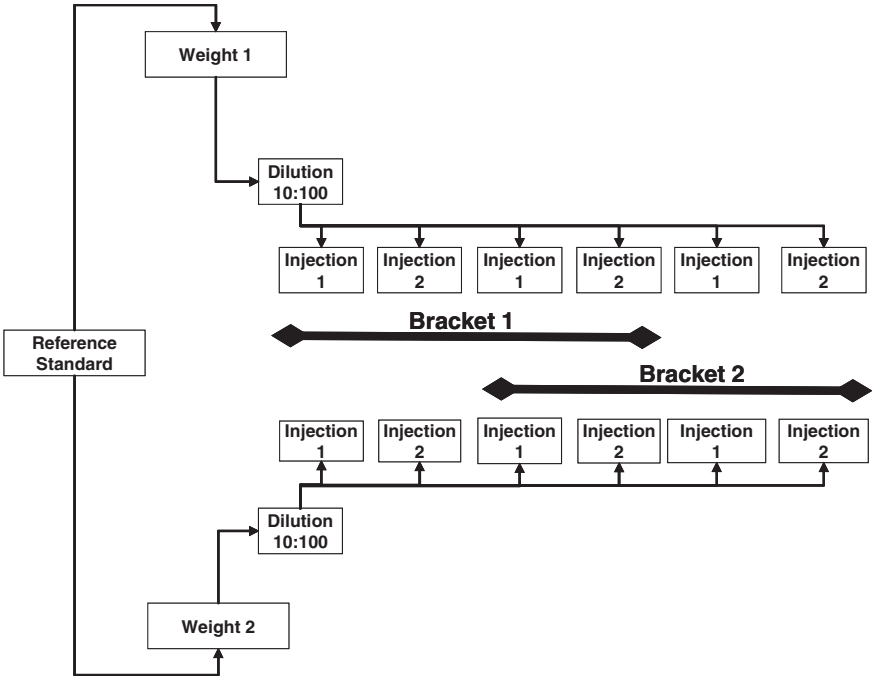


Figure 10-9 HPLC example: standard bracketing scheme.

In addition, the analytical method specifies the injection sequence and the calculation method. These are shown diagrammatically in Figures 10-10 and 10-11. Note that the common industry practice of averaging the standard responses has been avoided. This is because averaging reduces data spread and one of the intents is to find out the maximum data spread for OOE evaluation (Fig. 10-11). Averaging is the process we use to gain the best least-squares estimate of the sample concentration (reportable value) as the final step in the analytical procedure.

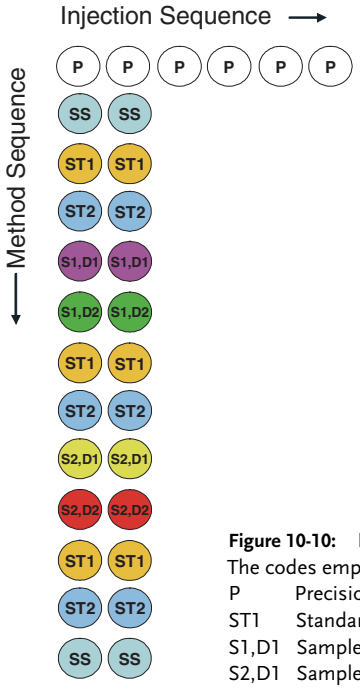


Figure 10-10: HPLC example: injection sequence.
The codes employed are as follows:

P	Precision standard	SS	System Suitability Standard
ST1	Standard preparation 1	ST2	Standard preparation 2
S1,D1	Sample 1 dilution 1	S1,D2	Sample 1 dilution 2
S2,D1	Sample 2 dilution 1	S2,D2	Sample 2 dilution 2

Figure 10-11 clearly illustrates the complexity of the analytical and calculation processes which lead to the generation of a reportable value. However, this is probably one of the simplest QC protocols in practical use.

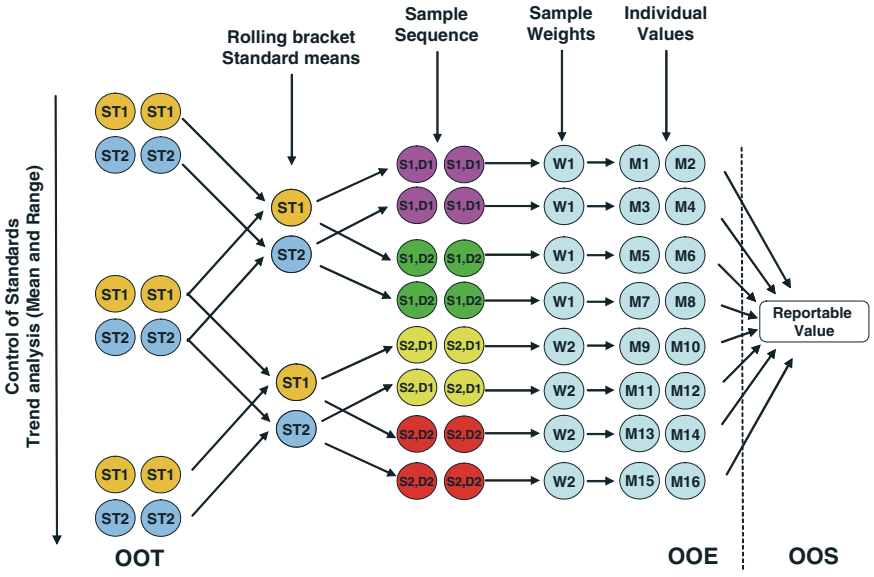


Figure 10-11 HPLC example; calculation method.

Table 10-4 Peak area data for HPLC example.

Injection #	Weights	Type	Preparation	Injection	Peak Area
1	0.1077	Standard	1	1	45298112
2	0.1077	Standard	1	2	44880224
3	0.1064	Standard	2	1	44730224
4	0.1064	Standard	2	2	44090118
5	2.1744	Sample	W1, D1	1	44214432
6	2.1744	Sample	W1, D1	2	44298626
7	2.1744	Sample	W1, D2	1	39018993
8	2.1744	Sample	W1, D2	2	44177892
9	0.1077	Standard	1	1	44693116
10	0.1077	Standard	1	2	40943696
11	0.1064	Standard	2	1	44616692
12	0.1064	Standard	2	2	44707718
13	2.0335	Sample	W2, D1	1	40599602
14	2.0335	Sample	W2, D1	2	40143560
15	2.0335	Sample	W2, D2	1	40236666
16	2.0335	Sample	W2, D2	2	41111162
17	0.1077	Standard	1	1	44070114
18	0.1077	Standard	1	2	44759924
19	0.1064	Standard	2	1	44793450
20	0.1064	Standard	2	2	44311080

Ignoring, for the moment, the possibility of AAR results arising from the precision and system suitability standards, the peak areas for the 20 injections for the main analytical sequence are shown in Table 10-4. From these data the calculation proceeds as follows.

1. Calculate the 12 normalised standard responses from injections 1–4, 9–12 and 17–20 from Eq. (10-6).

$$ST_{\#}^{NR} = \frac{ST_{\#}^{\text{Weight}}}{10ST_{\#}^{\text{PeakArea}}} \tag{10-6}$$

2. For each of the six pairs of normalised standard responses for each preparation, calculate the mean value
3. Calculate the eight normalised sample responses from injections 5–8, and 13–16 from Eq. (10-7).

$$SM_{\#}^{NR} = \frac{10SM_{\#}^{\text{PeakArea}}}{SM_{\#}^{\text{Weight}}} \tag{10-7}$$

4. Calculate the two pairs of mean values for the normalised standard responses across each of the brackets for Standards 1 and 2 ST_1^{meanNR} and ST_2^{meanNR}.

5. Calculate the 16 individual measurements from each of the eight sample injections using the mean values calculated in step 4 from Eq. (10-8).

$$V_{\#}^{ST_1} = SM_{\#}^{NR}\, ST_1^{\text{meanNR}} \text{ and } V_{\#}^{ST_2} = SM_{\#}^{NR}\, ST_2^{\text{meanNR}} \tag{10-8}$$

6. Calculate the reportable value from the average of these 16 individual measurements.

The results of these calculations are shown in Table 10-5. The specification for the product is registered as 0.0500%m/v as the target and a range of 0.0475–0.0525% m/v. The reportable value found, based upon the registered method, is 0.0481, which lies within these limits. On this basis alone, the batch result would be suitable for release without an investigation.

Table 10-5 Calculation table for HPLC example.

					Rolling bracket		Rolling bracket	
	Normalised areas		Standards		Mean normalised areas		Individual values	
Injection #	Standards	Samples	Mean	Range	Standard 1	Standard 2	Standard 1	Standard 2
1	2.3776E-10							
2	2.3997E-10		2.3887E-10	–2.2138E-12				
3	2.3787E-10							
4	2.4132E-10		2.3960E-10	–3.4534E-12				
5		203340839			2.4544E-10	2.3892E-10	0.04991	0.04858
6		203728045			2.4544E-10	2.3892E-10	0.05000	0.04867
7		179447172			2.4544E-10	2.3892E-10	0.04404	0.04287
8		203172792			2.4544E-10	2.3892E-10	0.04987	0.04854
9	2.4098E-10							
10	2.6304E-10		2.5201E-10	–2.2067E-11				
11	2.3848E-10							
12	2.3799E-10		2.3823E-10	4.8554E-13				
13		199653809			2.4726E-10	2.3853E-10	0.04937	0.04762
14		197411163			2.4726E-10	2.3853E-10	0.04881	0.04709
15		197869024			2.4726E-10	2.3853E-10	0.04892	0.04720
16		202169471			2.4726E-10	2.3853E-10	0.04999	0.04822
17	2.4438E-10							
18	2.4062E-10		2.4250E-10	3.7663E-12				
19	2.3753E-10							
20	2.4012E-10		2.3883E-10	–2.5858E-12				

Reportable Value	0.0481
Min	0.0429
Max	0.0500
Range	0.0071
Std. dev.	0.0020
RDS	4.3%

However, there may be OOE results within the data set used to calculate the reportable value. It is analytical best practice to specify the acceptable limits for data spread which are allowed in the calculation of reportable values, based upon the method development studies. This would usually be by way of limits for the range and/or RSD. These should be based upon the process capability of the method as determined during the method development and its subsequent validation. However, there are currently no generally agreed methods for setting such in-house limits. Laboratories should arrive at their own scientifically sound conclusions and be prepared to defend them under regulatory scrutiny.

For our example, the method validation report determined the intermediate precision of the method, *s*, to be 2.4%. Assume that we set an OOE limit for the range of individual measurements of $\pm 3.3\,s$ which, in our example, equates ± 0.0040 and hence an acceptable range of 0.0460–0.0540. Inspection of Table 10-5 reveals two analytical measurements that lie outside this range; 0.0440 and 0.0429, both associated with injection 7. The presence of these OOE results have affected the RSD, which is much larger at 4.3% than would be expected from the method validation (2.4%).

10.7 Statistical Outlier Tests for Out-of-Expectation Results

The use of outlier testing is a useful tool when investigating OOE results. The range measurement criterion for the OOE makes it likely that any OOEs will have a probability of less than 5 in a 1000 of belonging to the analytical population. One simple method of assessing outlying data is to calculate and plot the *z* scores.

z scores are simply the residuals (differences between the individual values and the sample mean) divided by the sample standard deviation;

$$z = \frac{(X_i - \bar{X})}{s} \tag{10-9}$$

In this way it becomes easy to see the shape of the data set. The calculations are shown in Table 10-6 and the dot plot in Figure 10-12.

Examination of the top plot in Figure 10-12 shows that the two OOE measurements are low and lie between –2z and –3z. The remaining 14 values are distributed between ± 1z. The two OOE measurements are sufficiently far away to suspect that they may be erroneous. If we exclude them from the calculation of the sample mean and standard deviation and recalculate the z scores, the results are shown in Table 10-6 and the lower plot of Figure 10-12.

Although the mean value changes slightly, the RSD reduces dramatically to 2%. The revised dot plot now shows that the 14 data points lie within $\pm 2z$ and the two OOE data points are now at –4.8 and –5.9, which is very strong evidence that they do not belong to the same sample population as the remaining 14. This statistical evidence is strongly indicative that injection 7 may have some analytical error associated with it and requires investigation.

Table 10-6 *z* scores for the original and reduced analytical measurements in the OOE investigation.

		All data		Without #3 & #11	
#	**X**	**Residual**	**z score**	**Residual**	**z score**
1	0.0499	0.0018	0.88	0.0011	1.14
2	0.0500	0.0019	0.93	0.0012	1.24
3	0.0440	–0.0041	**–1.98**	–0.0047	**–4.75**
4	0.0499	0.0018	0.86	0.0011	1.10
5	0.0494	0.0013	0.61	0.0006	0.60
6	0.0488	0.0007	0.34	0.0000	0.04
7	0.0489	0.0008	0.40	0.0002	0.15
8	0.0500	0.0019	0.92	0.0012	1.22
9	0.0486	0.0005	0.23	–0.0002	–0.19
10	0.0487	0.0006	0.28	–0.0001	–0.10
11	0.0429	–0.0052	**–2.55**	–0.0059	**–5.92**
12	0.0485	0.0004	0.21	–0.0002	–0.23
13	0.0476	–0.0005	–0.24	–0.0011	–1.15
14	0.0471	–0.0010	–0.50	–0.0017	–1.69
15	0.0472	–0.0009	–0.44	–0.0016	–1.58
16	0.0482	0.0001	0.06	–0.0005	–0.55
Mean	0.0481			0.0488	
Std dev	0.0020			0.0010	
RSD	4.3%			2.0%	

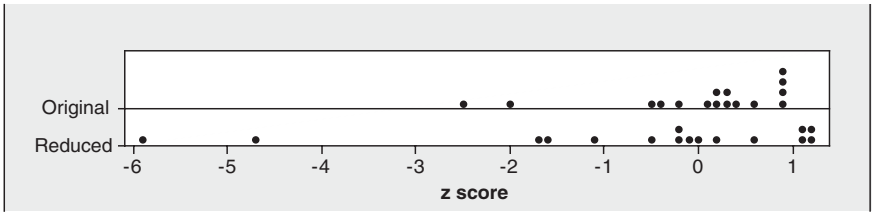


Figure 10-12: Dot plots of *z*-scores for OOE investigation.

There are other more powerful statistical methods for investigating outliers in data sets. These include Grubb's test (also known as the extreme studentised deviate test), Dixon's test and Hampel's test.

Grubb's test [14] follows directly from the *z* score method. The z scores are ranked (that is sorted into order) and the largest absolute z score is tested for statistical significance at the 95 or 99% (α=0.05 or 0.01) level. The test statistic used has been

generalised by Rosner [15] for multiple outliers. The test statistic [16] is given by Eq. (10-10).

$$\lambda_i = \frac{t_{n-i-1,p}(n-i)}{\sqrt{(n-i-1+t^2_{n-i-1,p})(n-i+1)}}$$

where

$i = 1, \ldots$ r outliers

$t_{\nu,p}$ is the $100p$ percentage point of the t distribution

with ν degrees of freedom and $p = 1 - \left[\frac{\alpha}{2(n-i+1)}\right]$ (10-10)

Fortunately this test statistic is readily calculated and is displayed in tabular form for n between 10 and 100 (Table 10-7).

This procedure requires that a decision is made in advance about the number of outliers, r, to be tested for. The value of r selected must be equal to or larger than the number of outliers in the data set,otherwise one might be missed. In our example, the number of OOE results is two so we select $r = 3$ to run the test. For first calculation steps the z score is calculated for all 16 data values as was previously done, and

Table 10-7 Rosner's ESD test statistic table for n=10 – 100 for up to five outliers.

	$\alpha = 0.05$ for $r = 1 - 5$					$\alpha = 0.01$ for $r = 1 - 5$				
n	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**
10	2.29	2.22	2.13			2.48	2.39	2.27		
11	2.35	2.29	2.22			2.56	2.48	2.39		
12	2.41	2.35	2.29			2.64	2.56	2.48		
13	2.46	2.41	2.35			2.70	2.64	2.56		
14	2.51	2.46	2.41			2.76	2.70	2.64		
15	2.55	2.51	2.46			2.81	2.76	2.70		
16	2.59	2.55	2.51			2.85	2.81	2.76		
17	2.62	2.59	2.55			2.89	2.85	2.81		
18	2.65	2.62	2.59			2.93	2.89	2.85		
19	2.68	2.65	2.62			2.97	2.93	2.89		
20	2.71	2.68	2.65	2.62	2.59	3.00	2.97	2.93	2.89	2.85
25	2.82	2.80	2.78	2.76	2.73	3.14	3.11	3.09	3.06	3.03
30	2.91	2.89	2.88	2.86	2.84	3.24	3.22	3.20	3.18	3.16
35	2.98	2.97	2.95	2.94	2.92	3.32	3.30	3.29	3.27	3.25
40	3.04	3.03	3.01	3.00	2.99	3.38	3.37	3.36	3.34	3.33
45	3.09	3.08	3.07	3.06	3.05	3.44	3.43	3.41	3.40	3.39
50	3.13	3.12	3.11	3.10	3.09	3.48	3.47	3.46	3.46	3.45
60	3.20	3.19	3.19	3.18	3.17	3.56	3.55	3.55	3.54	3.53
70	3.26	3.25	3.25	3.24	3.24	3.62	3.62	3.61	3.60	3.60
80	3.31	3.30	3.30	3.29	3.29	3.67	3.67	3.66	3.66	3.65
90	3.35	3.34	3.34	3.34	3.33	3.72	3.71	3.71	3.70	3.70
100	3.38	3.38	3.38	3.37	3.37	3.75	3.75	3.75	3.74	3.74

their absolute values are determined and ranked. The highest value is deleted and the z scores are recalculated. This process is repeated a second time. The values for the test statistics for n= 16 are obtained from Table 10-7.

Table 10-8 ESD test on HPLC data using Rosner's test statistic at 95% confidence.

	Stage 1		Stage 2		Stage 3	
#	X	$\frac{\lvert X_i-\bar{X}\rvert}{s}$	X	$\frac{\lvert X_i-\bar{X}\rvert}{s}$	X	$\frac{\lvert X_i-\bar{X}\rvert}{s}$
16	0.0482	0.06	0.0482	0.15	0.0482	0.55
12	0.0485	0.21	0.0485	0.05	0.0485	0.23
9	0.0486	0.23	0.0486	0.08	0.0486	0.19
13	0.0476	0.24	0.0476	0.54	0.0476	1.15
10	0.0487	0.28	0.0487	0.14	0.0487	0.10
6	0.0488	0.34	0.0488	0.23	0.0488	0.04
7	0.0489	0.40	0.0489	0.30	0.0489	0.15
15	0.0472	0.44	0.0472	0.81	0.0472	1.58
14	0.0471	0.50	0.0471	0.88	0.0471	1.69
5	0.0494	0.61	0.0494	0.59	0.0494	0.60
4	0.0499	0.86	0.0499	0.91	0.0499	1.10
1	0.0499	0.88	0.0499	0.94	0.0499	1.14
8	0.0500	0.92	0.0500	0.99	0.0500	1.22
2	0.0500	0.93	0.0500	1.00	0.0500	**1.24**
3	0.0440	1.98	0.0440	**2.84**		
11	0.0429	**2.55**				
Mean	0.0481		0.0485		0.0488	
Std dev	0.0020		0.0016		0.0010	
Rosner values		**2.59**		**2.55**		**2.51**

The calculation results are shown in Table 10-8. As the maximum value clearly exceeds the test statistic at Stage 2 but not at Stage 3, we conclude that both 0.0429 and 0.0440 are outliers. Note that this would also be the case at 99% confidence.

An alternative approach which has been used extensively for small analytical data sets (n = 3–30) is the test due to Dixon [17, 18] which is commonly called Dixon's Q test. This test, or more correctly, series of tests, is based upon the ratio of differences. It is arithmetically simple and is widely described in analytical chemistry text books. To carry out these tests, the n sample data need to be ranked (ordered) in increasing order, i.e., $x_1 < x_2 < \ldots x_n$. There is considerable confusion amongst the critical values to be used and the exact equations to be selected,in many text books. However, Rorabacher [19] has provided a concise and updated account of the appropriate critical values and equations. There are two families of equations: the first where the suspected outlier(s) is low; and the second where the reverse is the case. These are summarised in Table 10-9.

The r_{10} equations simply compare the difference between the suspected outlier and the value next to it as a fraction of the overall range (spread) of the data set. This

is the commonest form of the Dixon test used in analytical chemistry and is used for small data sets ($3 \leq n \leq 7$). The other equations are designed to avoid the influence of a second (or more) outliers in the data. Clearly the potential for additional outliers becomes increasingly possible as n increases. If the additional potential outliers are suspected at the other end of the ordered data set, the use of r_{11} or r_{12} is indicated. If both ends are suspect r_{21} and r_{22} can be used. If an additional outlier is suspected at the same end as the first, then r_{20} should be used. The ranges for n in Table 10-9 are generally the recommended ones based on Dixon's calculations.

Table 10-9 Dixon test equations and recommended ranges.

Condition	Equation		Sample size range
	Test lowest	Test highest	
For a single outlier x_i*	$r_{10} = \frac{x_2 - x_1}{x_n - x_1}$	$r_{10} = \frac{x_n - x_{n-1}}{x_n - x_1}$	$3 \leq n \leq 7$
For an outlier x_i avoiding x_n	$r_{11} = \frac{x_2 - x_1}{x_{n-1} - x_1}$	$r_{11} = \frac{x_n - x_{n-1}}{x_n - x_2}$	$8 \leq n \leq 10$
For an outlier x_i avoiding x_n, x_{n-1}	$r_{12} = \frac{x_2 - x_1}{x_{n-2} - x_1}$	$r_{12} = \frac{x_n - x_{n-1}}{x_n - x_3}$	$11 \leq n \leq 13$
For an outlier x_1 avoiding x_2	$r_{20} = \frac{x_3 - x_1}{x_n - x_1}$	$r_{20} = \frac{x_n - x_{n-2}}{x_n - x_1}$	$8 \leq n \leq 10$
For an outlier x_1 avoiding x_2 and x_n	$r_{21} = \frac{x_3 - x_1}{x_{n-1} - x_1}$	$r_{21} = \frac{x_n - x_{n-2}}{x_n - x_2}$	$11 \leq n \leq 13$
For an outlier x_1 avoiding x_2 and x_n, x_{n-1}	$r_{22} = \frac{x_3 - x_1}{x_{n-2} - x_1}$	$r_{22} = \frac{x_n - x_{n-2}}{x_n - x_3}$	$n \geq 14$

* r_{10} is Dioxin's Q test

In our HPLC example, n is 16 so we select Dixon's test equation r_{22}. The calculation from the ranked data is shown in Table 10-10. The critical values are found in Table 10-11. The value 0.0429 is identified as an outlier. After its removal and the recalculation, 0.0440 is not identified as an outlier as it was in the Rosner ESD test. This is an important finding as the use of different outlier tests can yield different results. W. Edwards Deming fully supported RA Fisher's remark that "*a point is never to be excluded on statistical grounds alone*" [20] .

Both the ESD test and Dixon's test are sequential tests and are susceptible to the influence of the very outliers they are trying to detect. Robust methods based upon the median rather than the mean have been developed, that are relatively unaffected by the number of outliers. One of these is the Hampel test [21, 22]. It has the advantage also of usually being a one-time test for data sets with only a few potential outliers. This is because the median is relatively unaffected by an outlying observation.

Table 10-10 Dixon's test on HPLC example data.

	Ranked					Ranked			
	#	X				#	X		
1	11	0.0429	x_3-x_1	0.0042	1	3	0.0440	x_3-x_1	0.0032
2	3	0.0440	x_{14}-x_1	0.0070	2	14	0.0471	x_{13}-x_1	0.0059
3	14	0.0471	r_{22}	0.599	3	15	0.0472	r_{22}	0.538
4	15	0.0472	Crit. value	0.548	4	13	0.0476	Crit. value	0.568
5	13	0.0476			5	16	0.0482		
6	16	0.0482			6	12	0.0485		
7	12	0.0485			7	9	0.0486		
8	9	0.0486			8	10	0.0487		
9	10	0.0487			9	6	0.0488		
10	6	0.0488			10	7	0.0489		
11	7	0.0489			11	5	0.0494		
12	5	0.0494			12	4	0.0499		
13	4	0.0499			13	1	0.0499		
14	1	0.0499			14	8	0.0500		
15	8	0.0500			15	2	0.0500		
16	2	0.0500				Mean	0.0485		
	Mean	0.0481				Std dev	0.0016		
	Std dev	0.0020				RSD	3.2%		
	RSD	4.3%							

Table 10-11 Critical values for Dixon's test equations at 95 and 99% confidence [19].

	r_{10}		r_{11}		r_{12}		r_{20}		r_{21}		r_{22}	
n	**95%**	**99%**	**95%**	**99%**	**95%**	**99%**	**95%**	**99%**	**95%**	**99%**	**95%**	**99%**
3	0.970	0.994										
4	0.829	0.926	0.977	0.995			0.983	0.996				
5	0.710	0.821	0.863	0.937	0.980	0.996	0.890	0.950	0.987	0.998		
6	0.625	0.740	0.748	0.839	0.878	0.951	0.786	0.865	0.913	0.970	0.990	0.998
7	0.568	0.680	0.673	0.782	0.773	0.875	0.716	0.814	0.828	0.919	0.909	0.970
8	0.526	0.634	0.615	0.725	0.692	0.797	0.657	0.746	0.763	0.868	0.846	0.922
9	0.493	0.598	0.570	0.677	0.639	0.739	0.614	0.700	0.710	0.816	0.787	0.873
10	0.466	0.568	0.534	0.639	0.594	0.694	0.579	0.664	0.664	0.760	0.734	0.826
11	0.444	0.542	0.505	0.606	0.559	0.658	0.551	0.627	0.625	0.713	0.688	0.781
12	0.426	0.522	0.481	0.580	0.529	0.629	0.527	0.612	0.592	0.675	0.648	0.740
13	0.410	0.503	0.461	0.558	0.505	0.602	0.506	0.590	0.565	0.649	0.616	0.705
14	0.396	0.488	0.445	0.539	0.485	0.580	0.489	0.571	0.544	0.627	0.590	0.674
15	0.384	0.475	0.430	0.522	0.467	0.560	0.473	0.554	0.525	0.607	0.568	0.647
16	0.374	0.463	0.417	0.508	0.452	0.544	0.460	0.539	0.509	0.580	0.548	0.624
17	0.365	0.452	0.406	0.495	0.438	0.529	0.447	0.526	0.495	0.573	0.531	0.605
18	0.356	0.442	0.396	0.484	0.426	0.516	0.437	0.514	0.482	0.559	0.516	0.589
19	0.349	0.433	0.386	0.473	0.415	0.504	0.427	0.503	0.469	0.547	0.503	0.575
20	0.342	0.425	0.379	0.464	0.405	0.493	0.418	0.494	0.460	0.536	0.491	0.562

Table 10-11 Continued.

	r_{10}		r_{11}		r_{12}		r_{20}		r_{21}		r_{22}	
n	**95%**	**99%**	**95%**	**99%**	**95%**	**99%**	**95%**	**99%**	**95%**	**99%**	**95%**	**99%**
21	0.337	0.418	0.371	0.455	0.396	0.483	0.410	0.485	0.450	0.526	0.480	0.551
22	0.331	0.411	0.364	0.446	0.388	0.474	0.402	0.477	0.441	0.517	0.470	0.541
23	0.326	0.404	0.357	0.439	0.381	0.465	0.395	0.469	0.434	0.509	0.461	0.532
24	0.321	0.399	0.352	0.432	0.374	0.457	0.390	0.462	0.427	0.501	0.452	0.524
25	0.317	0.393	0.346	0.426	0.368	0.450	0.383	0.456	0.420	0.493	0.445	0.516
29	0.312	0.388	0.341	0.420	0.362	0.443	0.379	0.450	0.414	0.486	0.438	0.508
27	0.308	0.384	0.337	0.414	0.357	0.437	0.374	0.444	0.407	0.479	0.432	0.501
28	0.305	0.380	0.332	0.409	0.352	0.431	0.370	0.439	0.402	0.472	0.426	0.495
29	0.301	0.376	0.328	0.404	0.347	0.426	0.365	0.434	0.396	0.466	0.119	0.489
30	0.298	0.372	0.324	0.399	0.343	0.420	0.361	0.428	0.391	0.460	0.414	0.483

It is based upon the calculation of the median of absolute residuals. One method of calculation is as follows [23].

1. Calculate the median, x_m, for the data set.
2. Calculate the absolute residuals of the median, $r_i = |x_i - x_m|$.
3. Calculate the median of these absolute residuals, $r_{m.}$
4. Calculate the ratios of the absolute residuals, $\frac{r_i}{5.06 r_m}$.

If the value of the ratio is greater than 1, then the data point is considered an outlier at 95% confidence. If there is a value or values close to unity then a recalculation, omitting any points greater than 1, may be used.

Table 10-12 Hampel's method applied to the HPLC example data.

#	x_i	r_i	$\frac{r_i}{5.06 r_m}$
1	0.04991	0.00128	0.23
2	0.05000	0.00138	0.24
3	0.04404	0.00458	0.81
4	0.04987	0.00124	0.22
5	0.04937	0.00074	0.13
6	0.04881	0.00018	0.03
7	0.04892	0.00030	0.05
8	0.04999	0.00136	0.24
9	0.04858	0.00005	0.01
10	0.04867	0.00005	0.01
11	0.04287	0.00575	**1.01**
12	0.04854	0.00009	0.02
13	0.04762	0.00100	0.18
14	0.04709	0.00154	0.27
15	0.04720	0.00143	0.25
16	0.04822	0.00040	0.07
		r_m	
MEDIAN	0.04863	0.00112	

The calculation is shown in Table 10-12. The data are shown to five places of decimals to show the calculation more clearly. Again data point 11, 0.0429, is declared an outlier whereas data point 3, 0.0440, is not. This conclusion is in agreement with the Dixon test.

It is not possible to say whether the z score and ESD conclusions are correct in identifying data point 3 as an outlier and that the Dixon and Hampel test are wrong in not identifying it as such on statistical grounds alone. My recommendation would be to use the most conservative approach in such cases and the robust method would be my choice for these data. Whatever the method chosen, it must be defended as being scientifically sound under regulatory scrutiny [24].

10.8 Trend Analysis for Quality Control

In the previous section we concentrated on the statistical investigation of OOE results in our HPLC example. However, the monitoring of analytical performance on an on-going basis is an important aspect of quality management in the laboratory [25]. The use of control charting and trend analysis is an essential component of such activities. For example, the application of trend analysis to detect AARs is widely used and applied to time-related data, such as system suitability data and other method and instrument performance criteria. The use of simple Shewhart control charts is well established for such data (see, for example, [26, 27]). The purpose of such charts is prophylactic and, properly used, they should reduce the number of OOT and OOE results. These preventative measures are best practice for minimising the occurrence of OOS results (see also Chapter 9).

From an AAR perspective, the investigation of historical time series data can be very helpful. For example, the trend analysis of system suitability data can identify significant changes and these can be correlated with known changes within the laboratory, such as columns, operators, reagents, etc. One of the most sensitive methods of detecting changes is CuSum analysis. Consider the HPLC system suitability, R_s, data shown in Table 10-13. The data are derived from the mean values of duplicate injections before and after the standard and sample sequence (Fig. 10-10).

Table 10-13 HPLC System Suitability, R_s, data for 50 consecutive runs.

#	*Rs*	#	*Rs*
1	2.200	26	2.300
2	2.150	27	2.325
3	2.250	28	2.325
4	2.225	29	2.300
5	2.300	30	2.275
6	2.150	31	2.350
7	2.250	32	2.250
8	2.150	33	2.325

Table 10-13 Continued.

#	Rs	#	Rs
9	2.250	34	2.200
10	2.225	35	2.300
11	2.175	36	2.200
12	2.225	37	2.300
13	2.225	38	2.250
14	2.175	39	2.250
15	2.300	40	2.275
16	2.175	41	2.175
17	2.275	42	2.275
18	2.275	43	2.200
19	2.250	44	2.325
20	2.150	45	2.225
21	2.300	46	2.300
22	2.350	47	2.250
23	2.250	48	2.225
24	2.225	49	2.250
25	2.325	50	2.225

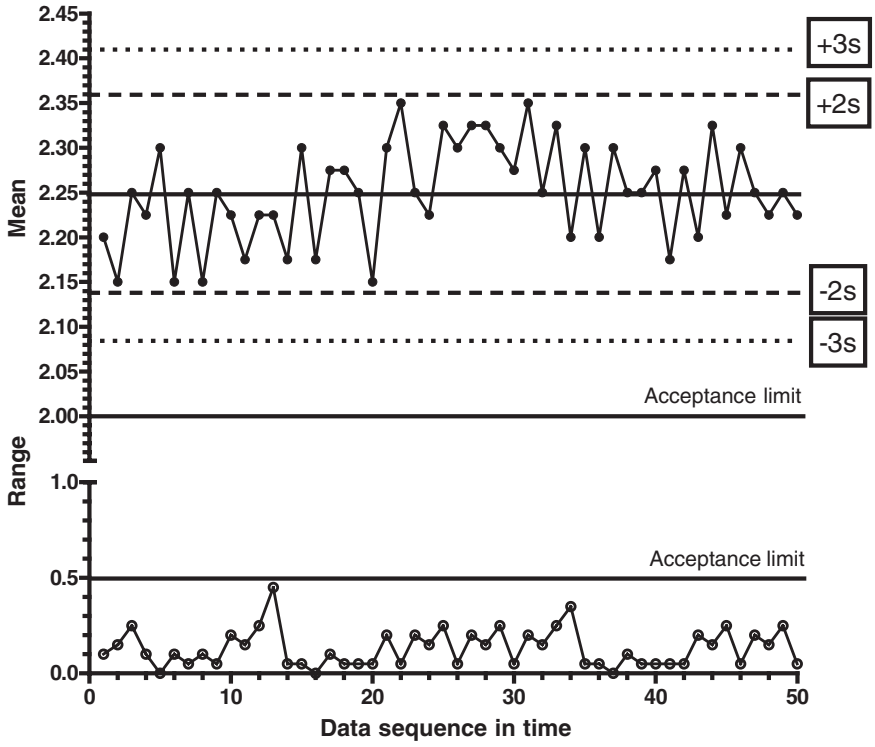


Figure 10-13 Shewhart mean and range control chart of R_s data over 25 consecutive runs.

During the method development and validation processes, it was established that the system suitability criteria for the mean resolution, *Rs*, had to be a minimum of 2.00 and have a range over the analysis run of not greater than 0.5. The laboratory operates a Shewhart mean and range control chart system and this is shown in Figure 10-13.

All the mean data are in statistical control and within the acceptance criterion. The range data are also within its acceptance criterion, with a few values approaching the limit. The mean data show considerable variability with respect to time and may be the result of subtle changes within the analytical testing which could lead to OOE or OOS results. The use of CuSum analysis as a post mortem technique often enables the correlation of statistically significant changes to physical events such as operator, column or mobile phase changes, for example, which warrant investigation.

10.9 CuSum Analysis of System Suitability Data

The technique of CuSum analysis was developed in the 1950s in the UK. It was used for process analysis and troubleshooting [28, 29]. It is basically a very simple calculation whereby the cumulative effect of successive differences from some target value is calculated. This target value could be a specification value or, as we will use for post-mortem purposes, the mean value. An example calculation for the CuSum is shown in Table 10-14 for the first seven data points and the resultant plot for all the 50 data points is shown in Figure 10-14. The mean of all 50 data points was 2.250.

Table 10-14 CuSum example values for the HPLC R_s values.

#	*Rs*	Difference from mean	CuSum
1	2.200	–0.050	–0.050
2	2.150	–0.100	–0.150
3	2.250	0.000	–0.150
4	2.225	–0.025	–0.175
5	2.300	0.050	–0.125
6	2.150	–0.100	–0.225
7	2.250	0.000	–0.225

The important features of a CuSum chart are the changes in slope rather than the numerical values. An upward trend indicates an increase in the parameter being monitored and vice versa. The question that needs to be addressed is: when is a change in slope due to an effect or merely due to the noise in the data? Inspection of Figure 10-14 indicates that the largest change occurs at #20 and another one may be at #33.

The method of testing the significance of each turning point is based on comparing the ratio of the maximum value of the CuSum to the local standard deviation, s_L,

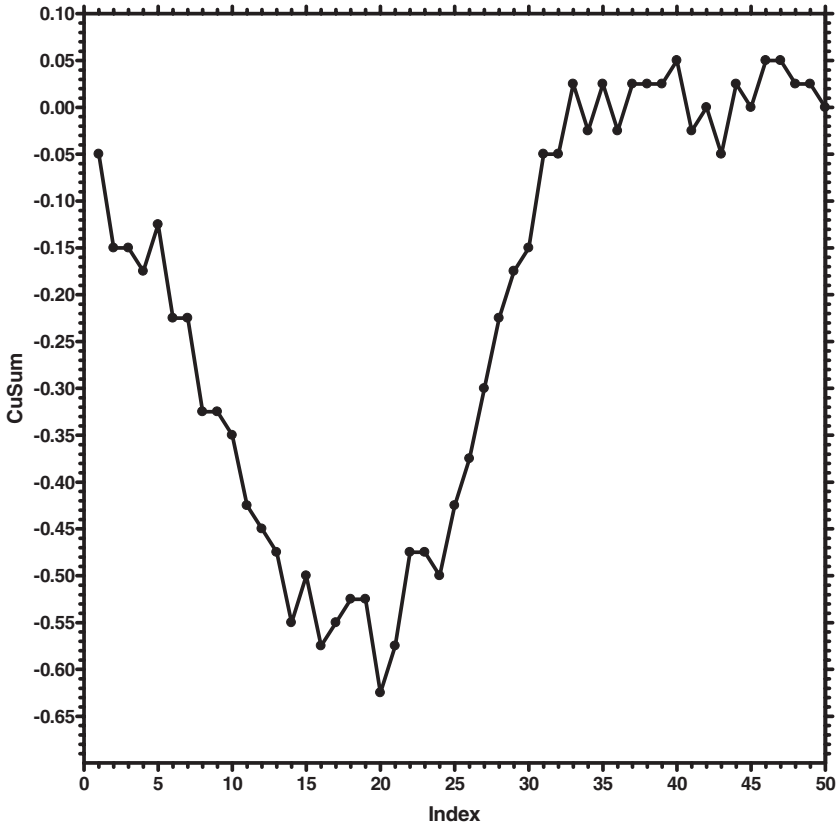


Figure 10-14 CuSum plot for HPLC R_s values.

with a set of critical values at 95 or 99% confidence. The method has not been as widely used as it might because these critical values are not readily available in conventional textbooks, etc. This method is arithmetically simple, but is a powerful tool for identifying changes

The method used for post mortem analysis of the HPLC data is as follows.

1. Calculate the sum of the squares of the differences between successive values and divide it by 2(n-1). The localised standard deviation of the data is the square-root of this value.

$$s_L = \sqrt{\frac{\sum_{i=1}^{i=n} (x_i - x_{i+1})^2}{2(n-1)}} \tag{10-11}$$

2. Find by inspection the absolute value of the maximum CuSum for the data set, and note the index number.

3. Calculate the test statistic $\frac{|\text{CuSum}_{\max}|}{s_L}$ (10-12)
4. Compare this value with the critical value for the span.
5. If this change point is significant, divide the CuSum plot into two groups by drawing two lines from the maximum CuSum to the extremities of the plot. These are the new baselines.
6. Inspect these regions for the next largest CuSum to be tested. Note that this is the vertical distance from the turning point to be tested to the new baseline,
7. Repeat steps 4 – 6 until no significant turning points remain.

The method is best illustrated with the HPLC R_s example data. Inspection of Figure 10-14 shows that the maximum CuSum value for the full span of 50 values is at #20 and has an absolute value of 0.625. To calculate the test statistic we need the local standard deviation as shown in step 1 above and illustrated in Table 10-15.

Table 10-15 Calculation of the local standard deviation, s_L, of the HPLC R_s CuSum data.

#	R_s	Difference from previous value	Squares of the difference
1	2.200		
2	2.150	–0.050	0.002500
3	2.250	0.100	0.010000
4	2.225	–0.025	0.000625
5	2.300	0.075	0.005625
6	2.150	–0.150	0.022500
7	2.250	0.100	0.010000
↓			
45	2.225	–0.100	0.010000
46	2.300	0.075	0.005625
47	2.250	–0.050	0.002500
48	2.225	–0.025	0.000625
49	2.250	0.025	0.000625
50	2.225	–0.025	0.000625
Mean	2.250	Sum of squares of the differences	0.310625
Std Dev	0.055	s_L	0.056

Starting with the largest CuSum, #20, the test statistic is now calculated from

$$\frac{|\text{CuSum}_{\max}|}{s_L} = \frac{0.625}{0.056} = 11.1 \quad (10\text{-}13)$$

and compared with the critical values given in Table 10-16 for a span of 50 data points. These values are 8.6 and 10.4 for 95% and 99% confidence, respectively. The test statistic is much larger than both of these critical values and therefore is highly significant.

Construct two baselines from the #20 turning point to cover the whole data set, as shown in Figure 10-15. Inspection of the region from 1 to 20 does not reveal any potentially significant changes. However, in the region 20 – 50, there is a change in the vicinity of #33. The absolute value of the second maximum CuSum is 0.385, which yields a test statistic value of 6.8. The critical values for a span of 30 are 6.7 and 8.0. Hence, this turning point is significant at 95% confidence, but not at 99% confidence. Given this finding there will not be any further significant turning points.

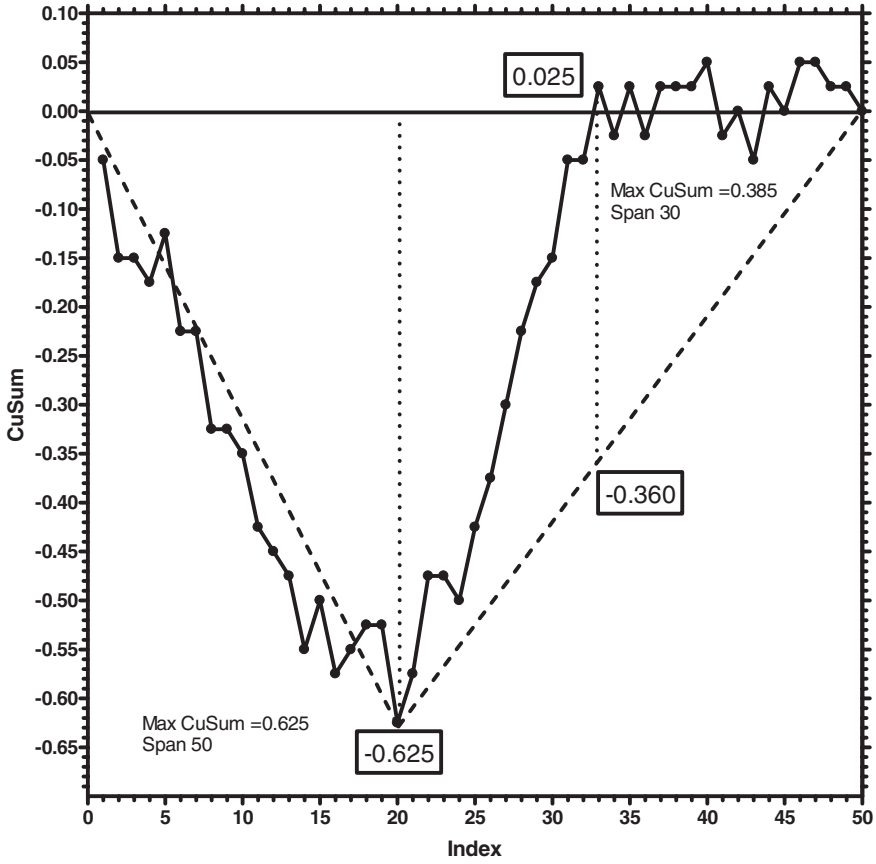


Figure 10-15 Analysis of CuSum turning points for HPLC R_s data.

Effectively, the analysis has shown that there are three statistically distinct regions within this data set, #1 – 20, 21 – 33 and 34 – 50. Based upon this information, the means and standard deviations may be calculated for each region.

If we now review our original Shewhart data in terms of these three distinct statistical regions, we can annotate it with these means and RSD values. The drawing of the means on such a plot is often referred to as a Manhattan plot (Fig. 10-16). Based upon our analysis we can ask a number of questions such as:

Why has the overall precision remained relatively constant over the 50 runs but the mean has not?

What changes occurred that might have resulted in the shifts in the R_s values?

Table 10-16 Critical values for the CuSum span test [30].

Span	Critical value		Span	Critical value	
	95%	99%		95%	99%
2	1.6	2.1	**14**	4.6	5.6
3	2.0	2.5	**15**	4.8	5.8
4	2.3	2.9	**20**	5.6	6.8
5	2.7	3.3	**25**	6.0	7.3
6	3.0	3.6	**30**	6.7	8.0
7	3.2	4.0	**40**	7.8	9.3
8	3.5	4.3	**50**	8.6	10.4
9	3.7	4.6	**60**	9.5	11.3
10	3.9	4.9	**70**	10.3	12.2
11	4.1	5.1	**80**	10.8	12.9
12	4.3	5.3	**90**	11.3	13.6
13	4.5	5.5	**100**	11.8	14.3

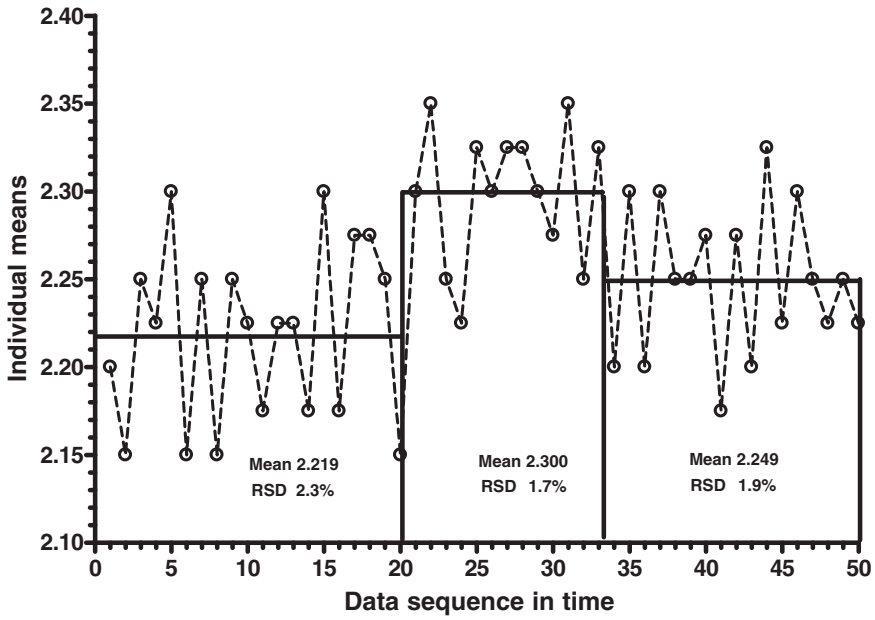


Figure 10-16 Manhattan plot and regional means and RSDs for HPLC R_s data.

10.10 Summary

Following the Wolin Judgement in the Barr Laboratories case, the regulatory focus, particularly by the FDA, in analytical laboratories, has been primarily on OOS results. However, much of the problem has been generated by the laboratories themselves, in that they have not controlled their analytical processes and methods in an appropriate manner. As shown in this chapter, laboratory quality management of OOE and OOT results and the proper definition of reportable values based upon analytical process capabilities, should minimise any OOS results arising from analytical causes. As Judge Wolin remarked, OOS results should be rare. If this is not the case then time-consuming analytical and manufacturing process investigations are inevitable.

10.11 References

[1] C. Burgess: Valid Analytical Methods and Procedures. The Royal Society of Chemistry, 2000, ISBN 0-85404-482-5.
[2] Draft Guidance for Industry; Investigating Out of Specification (OOS) Test Results for Pharmaceutical Production, FDA Center for Drug Evaluation and Research (CDER), September 1998.
[3] US CGMP regulations (21 CFR§ 211.160 (b)(4)).
[4] AMC Technical Brief No. 13, The Royal Society of Chemistry, September 2003, (http://www.rsc.org/lap/rsccom/amc/amc_index.htm).
[5] M. Thompson and R.J. Howarth. Analyst 105 (1980) 1188–1195.
[6] ISO 3534:3.12.
[7] ISO 3534:3.13.
[8] EURACHEM/CITAC Guide, Quantifying Uncertainty in Analytical Measurement, S L R Ellison, M Rosslein, A Williams (Eds), 2nd Edition, 2000.
[9] L.D. Torbeck: Reportable values for out-of-specification test results. Pharmaceutical Technology, Special supplement, February 1999.
[10] ICH Harmonised Tripartite Guideline Q2B, Validation Of Analytical Procedures: Methodology, 1996.
[11] W. Horwitz: Analytical Chemistry, 54 (1982) 67A–76A.
[12] J.S. Oakland, R.F. Followell: Statistical Process Control, 2nd edition, Heinemann Newnes, London, 1990.
[13] Center for Drug Evaluation and Research (CDER) Reviewer Guidance, Validation of Chromatographic Methods, November 1994.
[14] F.E. Grubbs: Technometrics, 11(1) (1969)1.
[15] B. Rosner: Technometrics, 25(2) (1983) 165.
[16] B. Iglewicz and D.C. Hoaglin: How to detect and handle outliers. Section 5.2.1, American Society for Quality Control (1993).
[17] W.J. Dixon: Biometrics, 9 (1953) 74–89.
[18] R.B. Dean, W.J. Dixon: Analytical Chemistry, 23 (1951) 636–638.
[19] D.B. Rorabacher: Analytical Chemistry, 63 (1991) 139–146.
[20] W. Edwards Deming: Statistical Adjustment of Data. 1943, 171 (Dover Reprint).
[21] P.L. Davies: Fresenius Z Anal Chem, 331 (1988) 513 –519.
[22] P.L. Davies: J Amer. Stat. Soc. 88 (1993) 782 –801.

[23] T.P.J. Linsinger, W. Kandler, R .Krska, M. Grasserbauer: Accred. Qual. Assur. 3 (1998) 322–327.
[24] 21 CFR§160(b).
[25] C. Burgess in Analytical Chemistry, R. Kellner, J.-M. Mermet, M. Otto, M. Valcárel, H.M. Widmer (Eds.), 2nd Edition, Chapter 6, Analytical Quality Management, 67–89, 2004, Wiley-VCH, ISBN 3-527-30590-4
[26] G. Kateman, L. Buydens: Quality Control in Analytical Chemistry, 2nd Edition, Section 3.4.2, John Wiley & Sons Ltd., 1993.
[27] P.C. Meier, R.E. Zund: Statistical Methods in Analytical Chemistry, 2nd Edition, Section 1.8.4, John Wiley & Sons Ltd., 2000.
[28] R.H. Woodward, P.L. Goldsmith: Cumulative Sum Techniques, Oliver and Boyd for Imperial Chemical Industries Limited, Edinburgh, 1964.
[29] Juran's Quality Handbook, 5th Edition, McGraw Hill, New York, Section 45.17, 1999.
[30] Values derived from the nomogram (Figure 12) of British Standard BS 5703 Part II (1980) which was devised using numerical simulation.

11
Future Trends in Analytical Method Validation

David Rudd

11.1
Introduction

Analytical method validation has evolved successfully over a number of years into the formalised, well-founded process which we recognise today. The key elements of method performance (accuracy, precision, sensitivity, etc.) are well-established, at least for laboratory-based pharmaceutical analysis, and are universally recognised as the main indicators by which 'fitness for purpose' of the analytical method may be demonstrated. For routine application of methods in a quality control or stability testing environment, for example, it is questionable whether there is any need for method validation concepts to evolve from this current position. While advances continue to be made in measurement technology within existing, well-established pharmaceutical analytical techniques (high pressure liquid chromatography or UV-visible spectrophotometry, for example), the basic requirements of method validation for such techniques remain largely unchanged.

However, there is, at present, a substantial shift occurring in the way in which pharmaceutical products and manufacturing processes are being developed – with a consequent impact on the type of analytical methodology required to support this new approach. Emerging regulatory [1] and business factors [2] are placing much greater emphasis on incorporating a 'Quality by Design' philosophy into pharmaceutical development and manufacture, such that the traditional approach of demonstrating and sanctioning product quality using laboratory-based finished product testing is no longer considered wholly appropriate.

The 'Quality by Design' principle seeks to overcome the shortcomings identified within the current approach to pharmaceutical development and manufacture, namely that finished product testing can only *reveal* problems relating to product quality, but can do little to rectify or prevent such problems occurring. For example, testing a batch of tablets in accordance with a pharmacopoeial specification for Uniformity of Content of individual dosage units will reveal that, perhaps, the powder blending process prior to compression of the tablets may have failed to achieve chemical homogeneity but, although the batch may be prevented from reaching the unsuspecting public, it remains unsatisfactory in terms of quality and safety and will need to be discarded or destroyed. More significantly, unless additional informa-

Method Validation in Pharmaceutical Analysis. A Guide to Best Practice. Joachim Ermer, John H. McB. Miller (Eds.)

ISBN: 3-527-31255-2

tion is gathered regarding the cause of the powder blending issue, there is every likelihood that subsequent batches will suffer from the same problems. Finished product testing may, therefore, be regarded as a 'safety net', there to catch the unfortunate faller, but doing little to prevent the fall in the first place.

In contrast, the 'Quality by Design' concept sets out to ensure that the finished product will possess the appropriate, pre-defined level of quality based on the application of a robust, scientifically well-understood manufacturing process coupled with assurance of the quality and suitability of the input raw materials.

This is achieved by appropriate characterisation of raw materials, not just from the traditional chemical and physical standpoint (for example chemical purity, water content, particle size, etc.), but also by establishing the suitability for subsequent processing (or 'processability'). As an example, it may be important to establish the mixing characteristics for materials which need to be blended with one another, or to confirm the flow properties where material transfer during processing is unavoidable.

In turn, and in order to be truly classified as 'robust', the manufacturing process or processes will need to be able to accommodate the inevitable variations in raw material quality and/or characteristics. Even with stringent controls based on chemical and physical specifications, as well as aspects of processability, raw materials continue to exhibit some variability from batch to batch, or from supplier to supplier, in terms of quality. As long as a specification parameter contains allowable ranges (for example, chemical purity might be specified as a range from 98 to 102% by weight), variation in quality remains inevitable.

Thus, manufacturing processes must be able to deal with these typical variations in such a way that the quality of their output remains consistent. This means that the manufacturing process itself must possess a degree of controlled flexibility – for if a truly fixed process is applied to variable input (i.e. variable raw materials), inevitably only variable output (finished product) will be achieved.

Such flexibility can only be established during the development of the manufacturing process, such that a full understanding of the complex inter-relationship between process operating parameters and the quality attributes of the finished product is obtained. In order to achieve this 'process understanding', it is necessary to have analytical measurement capability which allows information to be obtained within a very short time frame (relative to the process) such that the influence of changes in process operating parameters can be readily seen in terms of their impact on the finished product quality attributes.

This leads to the concept of 'real time' analytical measurement – that is, a measurement or set of measurements which is made during the operation of the process itself (probably within the production environment), rather than after completion of the process and on the output of the process (and probably within a laboratory environment). The 'real time' element of such measurements is the key to establishing the process understanding, and subsequently the process control, needed in order to implement a 'Quality by Design' development and manufacturing philosophy.

There are significant implications regarding the validation of such 'real time' analytical methods and, while it may be true that the principles and 'spirit' of published guidance [3, 4] remain applicable, there are nevertheless a number of further consid-

erations and points of detail which must be addressed as a result of these newer types of measurement and control approaches.

11.2
'Real Time' Analytical Methodologies

One of the first realisations, when recognising the need to introduce analytical assessment earlier into the pharmaceutical manufacturing process, is that it is seldom possible, or indeed appropriate, simply to transpose the laboratory-based measurement technique into the production environment. For example, although high pressure liquid chromatography (HPLC) is used extensively in the laboratory to confirm powder blend homogeneity during pharmaceutical blending processes, the equipment remains unsuitable for routine operation in the potentially dusty manufacturing environment. In addition, the difficulty in taking randomised, representative samples (which themselves may require fairly complex dissolution, dilution and/or filtration prior to the HPLC measurement step) is likely to limit the effectiveness of the technique in 'real time' process monitoring applications of the type indicated.

In order to overcome these shortcomings, alternative analytical methodology is generally used, where techniques more suitable for 'real time' process analysis can be incorporated. Near infra-red spectroscopy (NIR), with its information-rich data output, minimal sample pre-treatment and capability for rapid measurement, is particularly suitable, for example, for powder blend homogeneity testing [5]. In turn, the use of passive acoustics for the endpoint control of high shear tablet granulation processes [6], especially in respect of the physical attributes of the granule, is also well-established.

Both of these technologies share common features characteristic of successful 'real time' analytical methodology:

a) *Non-invasive*
   This means that no discrete sampling is required. Indeed, such technologies may be regarded as 'macro' techniques, where the data generated reflects an averaging of the many successive measurements on the entire available 'sample'. Spectroscopic techniques may achieve this using fibre optic probes, for example, or simply by introducing electromagnetic radiation into the system via external windows on the process equipment. Similarly, acoustic techniques depend on the use of transducers fixed to the outside of the process equipment and, while both types of methodology could be said to involve 'virtual' samples (the size of which are calculable based on the optical or acoustic configuration of the equipment, frequency of measurement, depth of penetration, etc.), the fact remains that no discrete sample needs to be removed from the system under investigation.
b) *Continuous output*
   Unlike laboratory-based methodology which generally gives a single result for each available sample (an HPLC assay on a batch of tablets, for example),

'real time' techniques allow virtually continuous measurements to be made. NIR spectra can be generated and displayed several times each second, while acoustic signals are usually averages of measurements made typically at rates up to 2000 times per second.

c) *Information-rich*
The types of data set produced by 'real time' techniques tend simultaneously to contain information about a number of quality attributes of the system under investigation. For example, NIR spectroscopic data includes information about chemical composition (and, hence, can be used for assay determination), but is also affected by some physical aspects of the sample (and may, therefore, be used to assess properties such as particle size, granularity etc). As a result, simple uni-variate treatment of the data generated may be inappropriate, suggesting that a multi-variate statistical (or 'chemometric') approach is likely to be required.

d) *Application to dynamic systems*
Most 'real time' measurements are made on dynamic processes – that is, the system which is being measured is itself changing during the measurement process. Unless a manufacturing processes is deliberately halted during the measurement step (an action which, if absolutely critical in obtaining meaningful analytical data, would bring into question the suitability of the so-called 'real time' technique), then the frequency of measurement must be sufficiently high that meaningful averages can be taken. This will help to overcome the transient effect of measurement parameters changing due to the dynamics of the system under investigation.

11.3
Validation Consequences of 'Real Time' Analytical Methodologies

Based on the characteristics described above, there are a number of consequences when the issue of method validation is addressed.

If no discrete sample is taken, what does the 'real time', non-invasive measurement actually relate to? For a dynamic system, is it fair to say that the true sample is the entire batch or bulk material contained within the processing equipment? Clearly the answer is only in the affirmative when a sufficiently large number of replicate measurements on a sufficiently sized 'virtual' sample have been made.

Thus, a key aspect of the validation of any non-invasive analytical method must involve an understanding of the required measurement frequency and duration (i.e. the overall number of replicate measurements) as well as the calculation (or, at least, estimation) of the 'virtual' sample size.

This latter point is particularly true when attempting to establish chemical homogeneity in a bulk system, for example, as statistical criteria for acceptance are inevitably based on the variation seen in individual dosage units of defined size and/or mass. At the regulatory level, there is clear published guidance [7] regarding the lev-

el at which powder blend uniformity needs to be demonstrated during tablet manufacture, for example. Hence, validation of 'real time' methods used to demonstrate compliance with such requirements will need to show that an appropriate 'virtual' sample size has been achieved. Such considerations will also allow the data processing algorithms regarding signal averaging to be established – in turn reflecting the likely rate of change which the measurement technique experiences due to the dynamics of the system under investigation. All in all, this is a critical aspect in terms of demonstrating method suitability, depending as it clearly does on a number of dynamic factors.

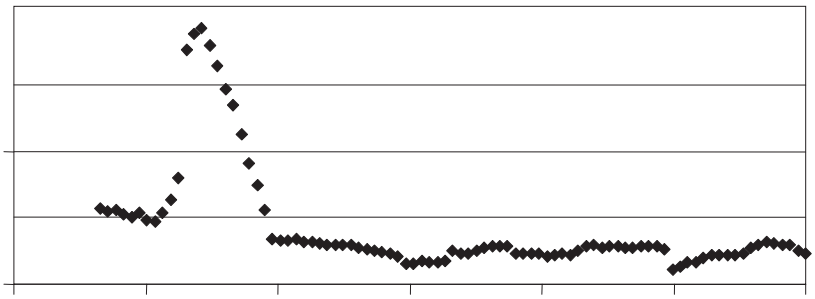


Figure 11-1 Variation in replicate NIR spectra (n = 12) versus time of blending.

As an example of a typical application of the approach described, Figure 11-1 represents a powder blend homogeneity determination where, based on the 'virtual' sample size obtained with the particular optical configuration used, the endpoint of the blending process was established by calculating the variation in successive sets of twelve NIR spectra. Part of the validation for this method involves the justification for basing the conclusions on sets of twelve spectra.

The term 'non-invasive' also implies a physical separation between the measurement device and the system under investigation. Even when a fibre optic probe is introduced into a pharmaceutical manufacturing process, but especially when sensors or detectors are placed on the outside of process container walls, questions arise concerning the reproducibility of location of these sensors or probes.

Are the sensors positioned correctly and are they positioned in exactly the same locations as before? If not, the quality of the data obtained will inevitably be reduced and, in the worst case, may not truly relate to the system under investigation. Also, if the sensors need to be mechanically fixed to the outside of the process container, how critical and how reproducible is the fixing process? The quality of an acoustic signal is heavily dependent on the acoustic coupling between the sensor and the process container and needs to be consistently achieved if data are to be compared from one set of measurements to another.

Finally, in considering the validation aspects of non-invasive, sensor-based methodology, the question of the number of sensors must also be addressed. It would be wrong to assume that one sensor alone will always be sufficient to monitor and con-

trol a given pharmaceutical process. On the other hand, multiple sensors may also be unnecessary, providing redundant or, in the worst case, conflicting information. It becomes necessary, therefore, to establish the optimal number of sensors during the development of the 'real time' method, but to continue to validate the suitability of the chosen number during subsequent usage and routine application.

The notion of 'continuous output' also poses some interesting questions at the regulatory level. With measurements being taken at an extremely high frequency (especially compared to those obtained during conventional, laboratory-based testing) and over a relatively long period of time (i.e. the duration of the manufacturing process), the probability of obtaining a high number of spurious or atypical results is increased (see Chapter 10). Although such results will be heavily outweighed by 'correct' data, the issue of how to deal with atypical or 'out-of'specification' results must clearly be addressed.

Part of the validation process must, therefore, incorporate a thorough review of atypical data generated during the development of the 'real time' method, so that a meaningful specification and acceptance criteria may also be developed to reflect the suitability of the method for control and routine application. It would be naive to assume that atypical data will not be generated during routine application – especially in the relatively 'analytically hostile' production environment – so a statistically-based performance specification for the 'real time' method, using knowledge gained during its development, will allow the issue of atypical data to be addressed.

As previously discussed, many 'real time' techniques generate data sets which contain information relating simultaneously to a number of quality attributes. The processing of such data sets, therefore, requires the use of multi-variate statistical techniques in order to understand the relationship between the analytical measurements and the quality attributes of the system under investigation.

Generally, although well-established statistical methods may be used, the validation of 'real time' analytical methodology will involve demonstration that appropriate multi-variate techniques are being employed. Guidance is available for NIR methodology in particular [8], showing how multi-variate calibration models can be developed successfully but, in general, an empirical, but scientifically justifiable approach may need to be taken for less mature technologies (for example, acoustics where, for pharmaceutical applications, the complexity of the signals obtained currently preclude trivial description). Indeed, for acoustic methodology, in particular, there may be more value in using data treatment techniques which reveal distinctive *features* of the signals obtained rather than trying to condense the relevant process information into oversimplified single numbers or quantities.

11.4
Additional Validation Factors

11.4.1
To Calibrate or not to Calibrate?

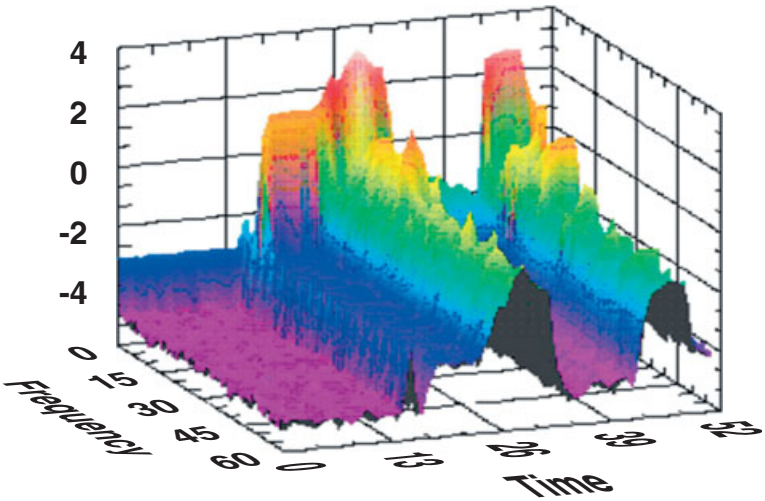


Figure 11-2 Typical passive acoustic signal obtained during a high shear granulation process and revealing the distinctive features attributable to the dry mixing, wet massing and wet granulation phases of the process.

It is important to realise that absolute calibration of many 'real time' techniques may be unnecessary and that such techniques may more usefully be regarded as 'indicators of change'. If we argue that process endpoints are often revealed by answering the question "has the process reached a point of equilibrium?", and if the position of that equilibrium is of lesser importance, then calibration is no longer required.

Powder blend homogeneity is a useful example of this situation as homogeneity is revealed when replicate spectra do not differ significantly from one another. While such results may fail to convey the final chemical composition of the powder blend system, the key question – "is the system homogeneous?" – is clearly answerable using validated, yet uncalibrated methodology.

Nevertheless, it is important to be able to demonstrate the required degree of discrimination when a 'real time' technique is used in this way. If a method is used as an 'indicator of change', how much or how little change is detectable? How much or how little *needs* to be detected? Are these levels of sensitivity and level of detection reproducible?

There is also the issue of calibration of the sensors themselves. Whatever the application of the 'real time' method, validation of this method needs to demonstrate that the sensors are performing consistently and accurately, where appropriate. This is especially true if multiple sensors are being used for comparisons across a process and where decisions are likely to be made on any differences detected between these sensors. Such differences need to be attributable to the system under investigation, rather than being due to variable performance levels of the sensors themselves.

This leads to the conclusion that appropriate reference standards are needed for calibration of sensors or probes used in 'real time' methods. While many suitable reference standards exist for available 'real time' technologies, it is worth noting that the suitability and performance of some 'accepted' reference standards may not be wholly appropriate when certain sensors are used for pharmaceutical applications. For example, the Hsu Nielsen source [9], widely used for the calibration of acoustic transducers and based on the acoustic emission generated from the transient breakage of a standard pencil lead under standard conditions, bears little relationship to the emissions produced during a pharmaceutical granulation process. As a result, work is on-going to develop a more appropriate reference standard for the extended applications of acoustic monitoring into the pharmaceutical field.

11.5
Validation of Analytically-based Control Systems

The move towards process-based or 'real time' measurement provides great opportunity within pharmaceutical manufacture to develop automated feed-back (and feed-forward) control systems. Indeed, the true benefits of a process-based approach (or 'Quality by Design' approach) are likely only to be realised when the analytical measurement step is used in conjunction with appropriate decision-making algorithms and process control and automation systems. This allows 'real time' modification of the process to ensure that it remains 'in control' (as adjudged by the analytical monitoring technology), or to bring it back 'in control' if it seems to deviating outside a pre-defined acceptable operating range.

Such steps represent progress towards the ideas of continuous manufacturing processes for routine pharmaceutical production and the capability of sanctioning product quality based on process data rather laboratory data or end-product testing. In order to achieve these outcomes, however, a number of validation aspects must be addressed:

11.5.1
What is the Basis for the Decision-Making Process?

Often this is straightforward. For example, a drying process may be described as 'complete' when the water or solvent content of the bulk material reaches or falls below a pre-determined set point. At this stage, the drying operation may be terminated. In this case the decision-making algorithm is very clear (i.e. when the solvent

content is less than or equal to x, switch off the drier), stemming from the uni-variate nature of the relationship between the measurement and the critical quality attribute of the product.

However, if a more complex set of relationships exist (for example, a granulation endpoint may be dependent on a number of measurements complying with pre-determined windows of operation), then the decision-making algorithm is necessarily more involved (i.e. multi-variate) and may even incorporate elements of fuzzy logic (i.e. when the measured parameter is 'about x') or empirical relationships (derived using artificial neural networks, for example). Under these circumstances, demonstration of the validity, accuracy and reproducibility of the algorithm whereby product quality is ensured (and, perhaps, sanctioned) is vitally important and represents a significant development on current method validation practices.

11.5.2
What are the Acceptable Operating Ranges?

Clearly these operating ranges need to be established during process development and will relate to the acceptable quality of the finished product. Validation of this relationship (or more exactly the relationship between control measurements and the finished product quality) and its continuation through changes of scale, equipment, source or raw materials etc forms a major part of this new generation of validation procedures.

11.5.3
Robustness of Process Signature

The multi-variate description of the process endpoint (rather than the multi-variate decision-making algorithm itself) alluded to earlier may more easily be viewed as a 'process signature', compliance with which will ensure good quality product. Conversely, deviations from this process signature will suggest that product of different quality is being produced. Again, demonstration of the robustness (or otherwise) of this process signature under extremes of change (scale, equipment, raw materials etc) needs to be addressed in any validation programme.

11.6
Continuous Validation

As a final consideration of the consequences of new developments in the application of analytical measurement to pharmaceutical systems, the idea of 'continuous validation' is proposed.

Although re-validation of methods is recognised as a necessary activity under particular circumstances (for example, as part of technology transfer or when significant modifications are made to parts of a given method, see Section 9.6), there is still

considerable reliance on the fact that the information generated during the formal validation of a method will apply throughout the lifetime of the method.

Some confirmation of method performance is often carried out on a periodic basis (for example, HPLC analyses conducted over several hours may incorporate re-calibration of the system using standard solutions to overcome the possibility of detector drift due to environmental temperature changes). However, where measurement techniques are being used in a more or less continuous fashion, confirmation of all critical aspects of method performance (not just accuracy as in the HPLC example given) should be considered – and also on a continuous basis.

Thus, a validation program needs to be established which identifies the critical aspects of method performance when used for continuous measurement. Then a mechanism needs to be identified where these key performance aspects may be verified. This might involve exposure of a probe or sensor to a standard or reference material at a frequency commensurate with the method application. As an example, if a liquid suspension system is being monitored spectroscopically for concentration over a period of one hour, verification of method performance at least every 15 to 20 minutes seems appropriate. At the very least, however, verification of key performance aspects should be carried before and after the measurement application, but with additional intermediate confirmation wherever possible, especially when the timescale of the application exceeds the likely stable period of the measurement system.

11.7 Conclusion

Significant developments are occurring in the way in which pharmaceutical analytical methodology is being applied. These developments result in the need to consider additional aspects of method validation which have not previously applied to laboratory-based measurements.

Published guidance [3, 4] provides useful information on the principles and philosophy of analytical method validation, without being over-prescriptive in terms of how such validation programmes should be conducted. It is recommended that these principles (incorporating the principles of good science, rigour and honesty) be adopted when considering how to extend current validation concepts into new application areas.

11.8 References

[1] United States Food and Drug Administration: Process Analytical Technology (PAT) initiative (http://www.fda.gov/cder/OPS/PAT.htm).

[2] L. Abboud, S.Hensley: New prescription for drug makers: update the plants. The Wall Street Journal, 3rd September 2003.

[3] International Conference on Harmonisation, Guideline for Industry, Text on validation of analytical procedures, ICH-Q2A (March 1995) (http://www.fda.gov/cder/guidance/ichq2a.pdf).
[4] International Conference on Harmonisation, Guideline for Industry, Q2B Validation of analytical procedures: Methodology (November 1996) (http://www.fda.gov/cder/guidance/1320fnl.pdf).
[5] M. Blanco, J. Coello, H. Iturriaga, S. Maspoch, C. de la Pezuela: Near-infrared spectroscopy in the pharmaceutical industry. Analyst, August 1998, Volume 123 (135R-150R).
[6] M. Whitaker, G .R. Baker, J. Westrup, P. A. Goulding, D. R. Rudd, R. M. Belchamber, M. P. Collins: Application of acoustic emission to the monitoring and end point determination of a high shear granulation process. Int. J. Pharm. 205 (1–2), **2000**, 79–92.
[7] United States Pharmacopiea USP27-NF22, USP general test chapter <905> Uniformity of Dosage Units (http://www.uspnf.com/uspnf/usp27-nf22/default.htm).
[8] N. Broad, P. Graham, P. Hailey, A. Hardy, S. Holland, S. Hughes, D. Lee, K. Prebble, N. Salton, P. Warren: Guidelines for the Development and Validation of Near-infrared Spectroscopic Methods in the Pharmaceutical Industry. In: Handbook of Vibrational Spectroscopy, Eds. J. M. Chalmers and P. R. Griffiths, John Wiley and Sons Ltd, Chichester, 2002 (http://www.spectroscopynow.com/Spy/pdfs/hvs_nirpharma.pdf).
[9] R. Diederichs, E. Ginzel: Hsu-Nielsen source, Nondestructive Testing Encyclopedia – Acoustic Emission, NDT Internet Publishing (NDTnet) (http://www.ndt.net/article/az/ae/hsunielsensource.htm)

Index

Method Validation in Pharmaceutical Analysis. A Guide to Best Practice. Joachim Ermer, John H. McB. Miller (Eds.)

ISBN: 3-527-31255-2