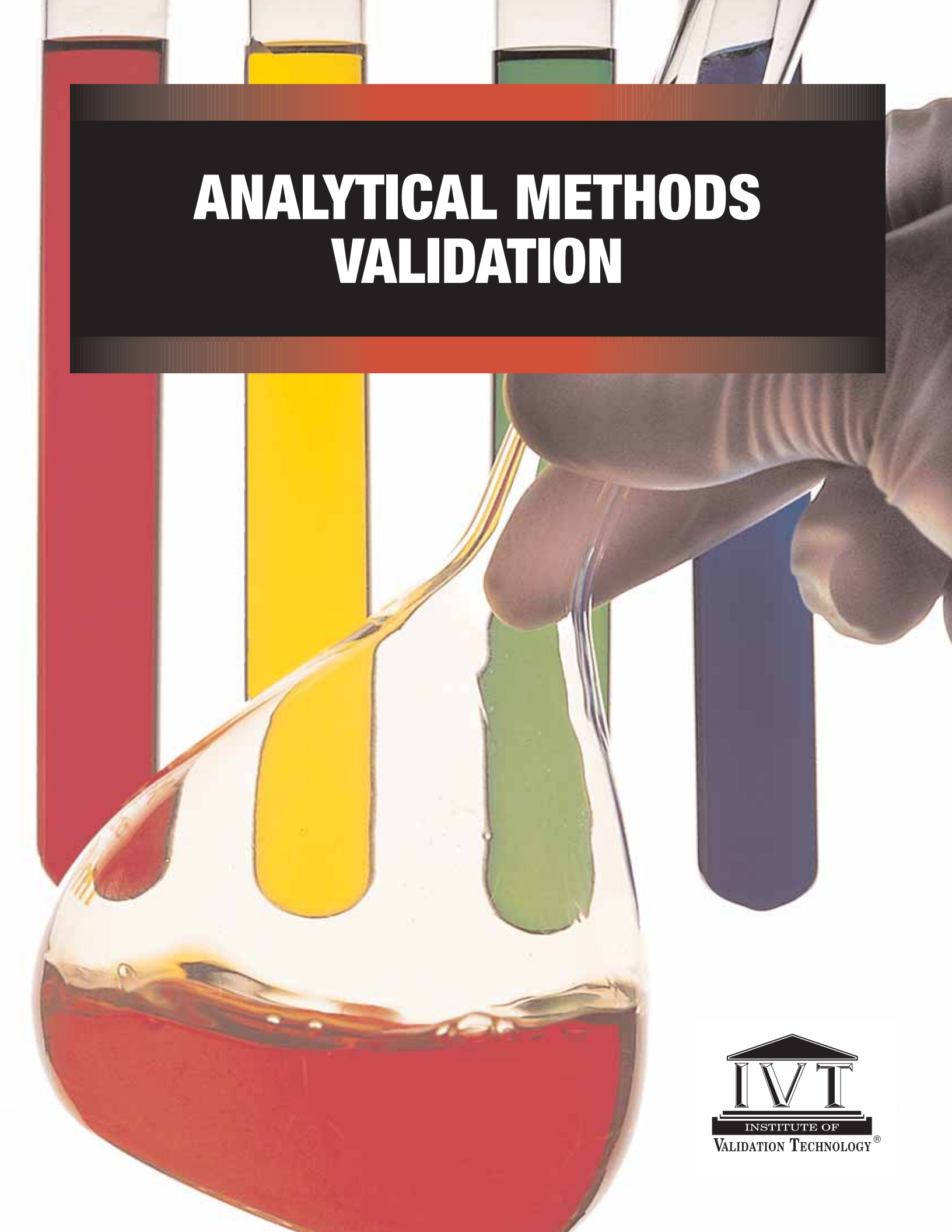
ANALYTICAL METHODS VALIDATION



IVT
INSTITUTE OF
VALIDATION TECHNOLOGY®

ANALYTICAL METHODS VALIDATION

Val•i•da•tion (văl´ĭ-da´shən) n. Establishing documented evidence that provides a high degree of assurance that a specific process will consistently produce a product meeting its predetermined specifications and quality attributes.

– USFDA

Don't miss these informative, information-packed Journals:

SUBSCRIBE TODAY!

Journal of GXP Compliance

The *Journal of GXP Compliance* is the complete guide to compliance practice and strategy. It's the only peer-reviewed journal dedicated exclusively to providing practical approaches and the news and information you need to make more informed decisions in your work every day.

	1 YEAR	2 YEARS
U.S. & Possessions	◯ $315.00	◯ $582.00
Canada/Mexico/All Others	◯ $412.00	◯ $752.00

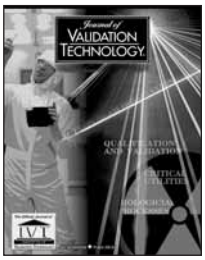
www.ivthome.com/journals

The **JOURNAL OF GXP COMPLIANCE** is a peer-reviewed publication providing an objective forum for expanding the global knowledge base for professionals in FDA-regulated industries, including Good Manufacturing Practice (GMP), Good Laboratory Practice (GLP), and Good Clinical Practice (GCP).

SAVE 3% when you subscribe to either JOURNAL **ONLINE!**

Journal of Validation Technology

- Four peer-reviewed issues annually
- Proposed validation standards
- Validation fundamentals written by Industry Experts
- Interviews with FDA officials
- In-depth Technical Guides on critical subject matter
- Roadmap articles on validation considerations



	1 YEAR	2 YEARS
U.S. & Possessions	◯ $315.00	◯ $582.00
Canada/Mexico/All Others	◯ $412.00	◯ $752.00

www.ivthome.com/journals

The **JOURNAL OF VALIDATION TECHNOLOGY** is a peer-reviewed publication that provides an objective forum for the dissemination of information to professionals in FDA-regulated industries. The **JOURNAL**'s Editorial Advisory Board reviews all submissions to ensure articles have been researched thoroughly, reflect current industry standards, and are not promotional in nature.

Signature _____ Date _____

Name (please print) _____

Job Title _____

Business Name _____

Address* _____

City_____ State/Province_____ ZIP/Postal Code _____

Country _____

*Is this your home address? ◯ Yes ◯ No

Phone Number (_____)_____ Fax Number (_____) _____

E-Mail Address _____

Payable in U.S. funds drawn on a U.S. bank.

◯ My check enclosed (made payable to Advanstar Communications Inc.)

◯ Charge my subscription(s) to: ◯ VISA ◯ MasterCard ◯ AMEX

Name on Card _____

Billing Address _____

City_____ State/Province_____ ZIP/Postal Code _____

Country _____

Card # _____ Exp.Date _____

Charges will appear on your credit card statement as Advanstar Communications Inc.

JGSE1

3 EASY WAYS TO SUBSCRIBE:

1) SAVE 3% - SUBSCRIBE ONLINE **IVThome.com/journals** (Enter Priority Code **JGSE1W**)

2) PHOTOCOPY & FAX THIS COMPLETED FORM TO **1-218-723-9417 or 9437**

3) PHOTOCOPY & MAIL COMPLETED FORM TO **Institute of Validation Technology P.O. Box 6004, Duluth, MN 55806-6004**

IVT INSTITUTE OF VALIDATION TECHNOLOGY®

TABLE OF
CONTENTS

ANALYTICAL METHODS VALIDATION

Step-by-Step Analytical Methods Validation and Protocol in the Quality System Compliance Industry

BY GHULAM A. SHABIR

Introduction

Methods Validation: Establishing documented evidence that provides a high degree of assurance that a specific method, and the ancillary instruments included in the method, will consistently yield results that accurately reflect the quality characteristics of the product tested.

Method validation is an important requirement for any package of information submitted to international regulatory agencies in support of new product marketing or clinical trials applications. Analytical methods should be validated, including methods published in the relevant pharmacopoeia or other recognized standard references. The suitability of all test methods used should always be verified under the actual conditions of use and should be well documented.

Methods should be validated to include consideration of characteristics included in the International Conference on Harmonization (ICH) guidlines[1, 2] addressing the validation of analytical methods. Analytical methods outside the scope of the ICH guidance should always be validated.

ICH is concerned with harmonization of technical requirements for the registration of products among the three major geographical markets of the European Community (EC), Japan, and the United States (U.S.) of America. The recent U.S. Food and Drug Administration (FDA) methods validation guidance document,[3-5] as well as the United States Pharmacopoeia (USP),[6] both refer to ICH guidelines.

The most widely applied typical validation characteristics for various types of tests are accuracy, precision (repeatability and intermediate precision), specificity, detection limit, quantitation limit, linearity, range, and robustness *(Figure 1)*. In addition, methods validation information should also include stability of analytical solutions and system suitability.[7]

Health Canada (HC) has also issued guidance on methods validation entitled *Acceptable Methods Guidance*.[8] HC has been an observer of ICH, and has adopted ICH guidelines subsequent to its reaching Step Four of the ICH process. An acceptable method predates ICH, and HC plans to revise this guidance to reflect current ICH terminology.

Figure 2 shows the data required for different types of analysis for method validation. Where areas of the *Acceptable Methods Guidance* are superseded by ICH Guidelines Q2A[1] and Q2B,[2] HC accepts the requirements of either the ICH or *Acceptable Methods Guidance*; however, for method validation, ICH acceptance criteria are preferred. HC's *Acceptable Methods Guidance* provides useful guidance on methods not covered by the ICH guidelines (e.g., dissolution, biological methods), and provides acceptance criteria for validation parameters and system suitability tests for all methods.

HC has also issued templates recommended as an approach for summarizing analytical methods and validation data ICH terminology was used when developing these templates.

This paper suggests one technique of validating methods. There are numerous other ways to validate methods, all

Figure 1

ICH, USP, and FDA Methods Validation Characteristics Requirements for Various Types of Tests

Validation Characteristics	Assay	Testing for Impurities		Identification
		Quantitative	Limit	
Accuracy	Yes	Yes	No	No
Precision - Repeatability	Yes	Yes	No	No
Precision - Intermediate Precision	Yes[1]	Yes*	No	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	No
Robustness	Yes	Yes	No	No

* In cases where reproducibility has been performed, intermediate precision is not needed.[7]

Figure 2

Health Canada Methods Validation Parameter Requirements for Various Types of Tests

Validation Parameters	Identity Tests	Active Ingredients		Impurities / Degradation Products		Physico-Chemical Tests
		Drug Substance	Drug Product	Quantitative	Limit Tests	
Precision (of the system)	No	Yes	Yes	Yes	1	Yes
Precision (of the method)	No	1	Yes	Yes	1	Yes
Linearity	No	Yes	Yes	Yes	No	Yes
Accuracy	No	Yes	Yes	Yes	1	Yes
Range	No	1	Yes	Yes	No	Yes
Specificity	Yes	1	Yes	Yes	Yes	*
Detection Limit	1	No	No	Yes	Yes	*
Quantitation Limit	No	No	No	Yes	No	*
Ruggedness	1	Yes	Yes	Yes	Yes	Yes

* May be required depending upon the nature of the test.

equally acceptable when scientifically justified.

Prepare a Protocol

The first step in method validation is to prepare a protocol, preferably written, with the instructions in a clear step-by-step format, and approved prior to their initiation. This approach is discussed in this paper. The suggested acceptance criteria may be modified depending on method used, required accuracy, and required sensitivity. (Note: Most of the acceptance criteria come from the characterization study.) Furthermore, some tests may be omitted, and the number of replicates may be reduced or increased based on scientifically sound judgment.

A test method is considered validated when it meets the acceptance criteria of a validation protocol. This paper is a step-by-step practical guide for preparing protocols and per-

forming test methods validation with reference to High Performance Liquid Chromatography (HPLC) (use similar criteria for all other instrumental test method validation) in the quality system compliance industry.

Analytical Methods Validation Protocol Approval Cover Page

Methods validation must have a written and approved protocol prior to its initiation. A project controller will select a validation Cross-Functional Team (CFT) from various related departments and functional areas. The project controller assigns responsibilities. The following tables illustrate one suggested way of documenting and preserving a record of the approvals granted at the various phases

Summary Information

Summary Information	
Organization name	
Site location	
Department performing validation	
Protocol title	
Validation number	
Equipment	
Revision number	

Project Controller

Project Controller	Name	Signature	Date

Document Approval

Document Approval			
Department / Functional Area	**Name**	**Signature**	**Date**
Technical Reviewer			
End Lab Management			
Health & Safety			
Quality Assurance			
Documentation Control (reviewed and archived by)			

Revision History

Revision History			
Revision No.	**Date**	**Description of change**	**Author**

of the validation:

Writing a Test Method Validation Protocol

Analytical method validations should contain the following information in detail:

Purpose: This section provides a short description of what is to be accomplished by the study.

Project scope: Identify the test methods and which products are within the scope of the validation.

Overview: This section contains the following: a general description of the test method, a summary of the characterization studies, identification of method type and validation approach, test method applications and validation protocol, the intended use of each test method application, and the analytical performance characteristics for each test method application.

Resources: This section identifies the following: end user laboratory where the method validation is to be performed; equipment to be used in the method validation; software to be used in the method validation; materials to be used in the method validation; special instructions on handling, stability, and storage for each material.

Appendices: This section contains references, signature, and a review worksheet for all personnel, their specific tasks, and the documentation of their training. Listings of all equipment and software necessary to perform the method validation should be found here along with document and materials worksheets used in the method validation and in the test method procedure(s).

1. Analytical Performance Characteristics Procedure

Before undertaking the task of methods validation, it is necessary that the analytical system itself be adequately designed, maintained, calibrated, and validated. All personnel who will perform the validation testing must be properly trained. Method validation protocol must be agreed upon by the CFT and approved before execution. For each of the previously stated validation characteristics *(Figure 1)*, this document defines the test procedure, documentation, and acceptance criteria. Specific values are taken from the ICH, U.S. FDA, USP, HC, and pertinent literature as references. (See the References section at the end of this article for further definitions and explanations.)

1.1. Specificity

1.1.1. Test procedure

The specificity of the assay method will be investigated by injecting of the extracted placebo to demonstrate the absence of interference with the elution of analyte.

1.1.2. Documentation

Print chromatograms.

1.1.3. Acceptance criteria

The excipient compounds must not interfere with the analysis of the targeted analyte.

1.2. Linearity

1.2.1. Test procedure

Standard solutions will be prepared at six concentrations, typically 25, 50, 75, 100, 150, and 200% of target concentration. Three individually prepared replicates at each concentration will be analyzed. The method of standard preparation and the number of injections will be same as used in the final procedure.

1.2.2. Documentation

Record results on a datasheet. Calculate the mean, standard deviation, and Relative Standard Deviation (RSD) for each concentration. Plot concentration (x-axis) versus mean response (y-axis) for each concentration. Calculate the regression equation and coefficient of determination (r2). Record these calculations on the datasheet.

1.2.3. Acceptance criteria

The correlation coefficient for six concentration levels will be ≥ 0.999 for the range of 80 to 120% of the target concentration. The y-intercept must ≤ 2% of the target concentration response. A plot of response factor versus concentration must show all values within 2.5% of the target level response factor, for concentrations between 80 and 120% of the target concentration.[9,10] HC states that the coefficient of determination for active ingredients should be ≥ 0.997, for impurities 0.98 and for biologics 0.95.[8]

1.3. Range

1.3.1. Test procedure

The data obtained during the linearity and accuracy studies will be used to assess the range of the method.

Linearity - Data Sheet		Electronic file name:	
Concentration (mg/ml)	**Concentration as % of Analyte Target**	**Peak Area (mean of three Injections)**	**Peak Area RSD (%)**
5 (e.g.)	25		
10	50		
15	75		
20	100		
30	150		
40	200		
Equation for regression line =		Correlation coefficient (r^2) =	

Range - Data Sheet	Electronic file name:
Record range:	

Accuracy - Data Sheet		Electronic file name:		
Sample	**Percent of Nominal (mean of three injections)**	**Amount of Standard (mg)**		**Recovery (%)**
		Spiked	**Found**	
1	75 (e.g.)			
2	100			
3	150			
Mean				
SD				
RSD%				

Repeatability - Data Sheet		Electronic file name:	
Injection No.	**Retention Time (min)**	**Peak Area**	**Peak Height**
Replicate 1			
Replicate 2			
Replicate 3			
Replicate 4			
Replicate 5			
Replicate 6			
Replicate 7			
Replicate 8			
Replicate 9			
Replicate 10			
Mean			
SD			
RSD%			

The precision data used for this assessment is the precision of the three replicate samples analyzed at each level in the accuracy studies.

1.3.2. Documentation

Record the range on the datasheet.

1.3.3. Acceptance criteria

The acceptable range will be defined as the concentration interval over which linearity and accuracy are obtained per the above criteria, and in addition, that yields a precision of ≤ 3% RSD.[9]

1.4. Accuracy

1.4.1. Test procedure

Spiked samples will be prepared at three concentrations over the range of 50 to 150% of the target concentration. Three individually prepared replicates at each concentration will be analyzed. When it is impossible or difficult to prepare known placebos, use a low concentration of a known standard.

1.4.2. Documentation

For each sample, report the theoretical value, assay value, and percent recovery. Calculate the mean, standard deviation, RSD, and percent recovery for all samples. Record results on the datasheet.

1.4.3. Acceptance criteria

The mean recovery will be within 90 to 110% of the theoretical value for non-regulated products. For the U.S. pharmaceutical industry, 100 ± 2% is typical for an assay of an active ingredient in a drug product over the range of 80 to 120% of the target concentration.[9] Lower percent recoveries may be acceptable based on the needs of the methods. HC states that the required accuracy is a bias of ≤ 2% for dosage forms and ≤ 1% for drug substance.[8]

1.5. Precision - Repeatability

1.5.1. Test procedure

One sample solution containing the target level of analyte will be prepared. Ten replicates will be made from this sample solution according to the final method procedure.

1.5.2. Documentation

Record the retention time, peak area, and peak height on the datasheet. Calculate the mean, standard deviation, and RSD.

1.5.3. Acceptance criteria

The FDA states that the typical RSD should be 1% for drug substances and drug products, ± 2% for bulk drugs and finished products. HC states that the RSD should be 1% for drug substances and 2% for drug products. For minor components, it should be ± 5% but may reach 10% at the limit of quantitation.[8]

1.6. Intermediate Precision

1.6.1. Test procedure

Intermediate precision (within-laboratory variation) will be demonstrated by two analysts, using two HPLC systems on different days and evaluating the relative percent purity data across the two HPLC systems at three concentration levels (50%, 100%, 150%) that cover the analyte assay method range 80 to 120%.

1.6.2. Documentation

Record the relative % purity (% area) of each concentration on the datasheet.

Calculate the mean, standard deviation, and RSD for the operators and instruments.

1.6.3. Acceptance criteria

The assay results obtained by two operators using two instruments on different days should have a statistical RSD ≤ 2%.[9, 10]

1.7. Limit of Detection

1.7.1. Test procedure

The lowest concentration of the standard solution will be determined by sequentially diluting the sample. Six replicates will be made from this sample solution.

1.7.2. Documentation

Print the chromatogram and record the lowest detectable concentration and RSD on the datasheet.

1.7.3. Acceptance criteria

The ICH references a signal-to-noise ratio of 3:1.[2] HC recommends a signal-to-noise ratio of 3:1. Some analysts calculate the standard deviation of the signal (or response)

Intermediate Precision - Datasheet	Electronic file name:					
	Relative % Purity (% area)					
	Instrument 1			Instrument 2		
Sample	S1 (50%)	S2 (100%)	S3 (150%)	S1 (50%)	S2 (100%)	S3 (150%)
Operator 1, day 1						
Operator 1, day 2						
Operator 2, day 1						
Operator 2, day 2						
Mean (Instrument)						
Mean (Operators)						
RSD%	S1 + S1	S2 + S2	S3 + S3			
Instruments						
Operators						

Limit of Detection - Data Sheet	Electronic file name:
Record sample data results: (e.g., concentration, S/N ratio, RSD%)	

Limit of Quantitation - Data Sheet	Electronic file name:
Record sample data results: (e.g., concentration, S/N ratio, RSD%)	

of a number of blank samples and then multiply this number by two to estimate the signal at the limit of detection.

1.8. Limit of Quantitation

1.8.1. Test procedure

Establish the lowest concentration at which an analyte in the sample matrix can be determined with the accuracy and precision required for the method in question. This value may be the lowest concentration in the standard curve. Make six replicates from this solution.

1.8.2. Documentation

Print the chromatogram and record the lowest quantified concentration and RSD on the datasheet. Provide data that demonstrates the accuracy and precision required in the acceptance criteria.

1.8.3. Acceptance criteria

The limit of quantitation for chromatographic methods has been described as the concentration that gives a signal-to-noise ratio (a peak with height at least ten times as high as the baseline noise level) of 10:1.[2] HC states that the quantitation limit is the best estimate of a low concentration that gives an RSD of approximately 10% for a minimum of six replicate determinations.[8]

1.9. System Suitability

1.9.1. Test procedure

System suitability tests will be performed on both HPLC systems to determine the accuracy and precision of the system by injecting six injections of a solution containing analyte at 100% of test concentration. The following parameters will be determined: plate count, tailing factors, resolution, and reproducibility (percent RSD of retention time, peak area, and height for six injections).

1.9.2. Documentation

Print the chromatogram and record the data on the datasheet

1.9.3. Acceptance criteria

Retention factor (k): the peak of interest should be well resolved from other peaks and the void volume; generally k should be ≥2.0. Resolution (Rs): Rs should be ≥2 between the peak of interest and the closest eluted peak,

System Suitability – Data Sheet		Electronic file name:		
System Suitability Parameter	Acceptance Criteria	Results		Criteria Met/ Not Met
		HPLC 1	HPLC 2	
Injection precision for retention time (min)	RSD ≤ 1%			
Injection precision for peak area (n = 6)	RSD ≤ 1%			
Injection precision for peak height	RSD ≤ 1%			
Resolution (R_s)	Rs = ≥ 2.0			
USP tailing factor (T)	T = ≤ 2.0			
Capacity factor (k)	K = ≥ 2.0			
Theoretical plates (N)	N = ≥ 2000			

Robustness - Data Sheet	Electronic file name:
Explain / record sample data:	

which is potentially interfering (impurity, excipient, and degradation product). Reproducibility: RSD for peak area, height, and retention time will be 1% for six injections. Tailing factor (T): T should be 2. Theoretical plates (N): ≥2000.[3]

1.10. Robustness

As defined by the USP, robustness measures the capacity of an analytical method to remain unaffected by small but deliberate variations in method parameters. Robustness provides some indication of the reliability of an analytical method during normal usage.

Parameters, which will be investigated, are percent organic content in the mobile phase or gradient ramp, pH of the mobile phase, buffer concentration, temperature, and injection volume. These parameters may be evaluated one factor at a time or simultaneously as part of a factorial experiment.

The chromatography obtained for a sample containing representative impurities, when using modified parameter(s), will be compared to the chromatography obtained using the target parameters. The effects of the following changes in chromatographic conditions will be determined: methanol content in mobile phase adjusted by ± 2%, mobile phase pH adjusted by ± 0.1 pH units, column temperature adjusted by ± 5°C. If these changes are within the limits that produce acceptable chromatography, they will be incorporated in the method procedure.[9, 10]

2. Appendices

List all appendices associated with this protocol. Each appendix needs to be labeled and paginated separately

Article Acronym Listing

CFT:	Cross-Functional Team
EC:	European Community
FDA:	Food and Drug Administration
HC:	Health Canada
HPLC:	High Performance Liquid Chromatography
ICH:	International Conference on Harmonization
RSD:	Relative Standard Deviation
U.S.:	United States
USP:	United States Pharmacopoeia

List of Appendices

Appendix No.	Document Title	Total Pages

Appendix 1

Method Validation Personnel Signature and Review Worksheet

Analyst Name	Dept.	Validation Activity	Protocol Reference	Analyst Signature	Date

Comments:

Completed By:	Signature:	Date:

Appendix 2

Equipment and Software Used in Method Validation Worksheet

Equipment Name/Module #	Last Calibration Date	Next Calibration Date	Software Name and Version	Validation Reference

Comments:

Completed By:	Signature:	Date:

Appendix 3

Document and Materials Used in Method Validation Worksheet

Complete Pre-protocol Execution				
Document Name/Ref. No.	**Edition/Version Number**	**Material Name**	**Supplier/ Lot Number**	**Expiration Date**
Comments:				
Completed By:	**Signature:**		**Date:**	

Appendix 4

Analytical Test Method Procedure

This procedure should include the entire testing method and all procedures associated with it. This appendix can appear in any format, but it should always be included in the documentation

from the body of the document. The following information must be found on every page of each appendix: validation protocol number; validation protocol title; appendix number (e.g., 1, 2, 3, … or A, B, C, …); and page X of Y. ❑

Acknowledgements

I thank Abbott Laboratories and MediSense for permission to publish this article. I also thank Dr. Alison Ingham (Health Canada) for his comments on the text.

About the Author

Ghulam Shabir is a Principal Scientist at Abbott Laboratories, MediSense UK. His group is responsible for materials characterization, analytical methods development, and validation and equipment qualification. Ghulam is a Fellow of the Institute of Quality Assurance and a Companion of the Institute of Manufacturing with 17 years of broad-based experience in the areas of production, quality control, quality assurance, in-process control, R&D, and validation in the pharmaceutical industry. Ghulam has received many technical excellence industrial awards as well as academic awards for authoring 'best scientific papers.' Ghulam's work has appeared in many publications. He has given several presentations at international conferences as well as having organized and moderated at symposiums. He holds a Master's Degree in Chemistry and Pharmaceutical Sciences. He can be reached by phone at 44-1993-863099, by fax at 44-1235-467737, or by e-mail at ghulam.shabir@abbott.com.

References

1. International Conference on Harmonization (ICH), Q2A: Text on Validation of Analytical Procedures, March 1995.
2. International Conference on Harmonization (ICH), Q2B: Validation of Analytical Procedures: Methodology, May 1997.
3. U.S. Center for Drug Evaluation and Research, Reviewer Guidance: Validation of Chromatographic Methods, November 1994.

4. U.S. FDA, Guidance for Submitting Samples and Analytical Data for Methods Validation, Rockville, Md., USA, Center for Drugs and Biologics, Department of Health and Human Services, February 1987.
5. U.S. FDA DHHS, 21 CFR Parts 210 and 211, Current Good Manufacturing Practice of Certain Requirements for Finished Pharmaceuticals, Proposed Rule, May 1996.
6. Validation of Compendial Methods, <1225>, U.S. Pharmacopoeia 26-National Formulary 21, United States Pharmacopeial Convention, Rockville MD, 2003.
7. U.S. FDA, Guidance for Industry: Analytical Procedures and Methods Validation: Chemistry, Manufacturing and Controls Documentation, August 2000.
8. Drugs Directorate Guidelines, Acceptable Methods, National Health and Welfare, Health Protection Branch, Canada, July 1994. (This guidance is available from HC as a print copy, but is soon to be released on the website http://www.hc-sc.gc.ca/hpfb-dgpsa/tpd-dpt/).
9. M.J. Green, Anal. Chem., Vol. 68, 1996. p. 305A.
10. G.A. Shabir, J. Chromatogra. A, Vol. 987, 2003. p. 57.

Validation of Analytical Methods Used in Cleaning Validation

BY HERBERT J. KAISER, PH.D. & BRUCE RITTS, M.S.

The information used to establish a positive cleaning validation is based on the result of validated analytical measurements. There must be a high degree of confidence in these results, as human safety depends on the lack of residues remaining on equipment. This article will describe various aspects regarding the validation of analytical methods used in cleaning validations. The validation elements are explored from both a theoretical point of view and through examples. References are provided to guide the reader to more in-depth information.

An analytical method is one of the deciding factors in establishing the cleanliness of pharmaceutical manufacturing equipment. It is, therefore, important that there be a high level of confidence in the results obtained using the method. This high level of confidence is established by testing and defining the usefulness of the analytical method. A properly developed cleaning validation strategy includes the analytical method validation, which defines the method parameters necessary in providing a high level of confidence in the cleaning results. The analytical method validation study demonstrates to scientific staffs, manufacturing personnel, and regulatory agencies, that the method performs as required, and that the results are reliable. There are many articles available that address analytical method validation within and outside of the pharmaceutical industry, both domestically and worldwide.[1,2,3,4,5,6,7,8]

Personnel other than analytical chemists may not understand the need for analytical method validation, let alone the extent to which these methods need to be evaluated. They may not understand that analytical method validation, as well as cleaning validation, has an important impact on everyday pharmaceutical manufacturing.[9] As is the case with cleaning validation, analytical methods may need to be revalidated.[10] This revalidation may arise from changes in instrumentation, analytes or manufacturing methods, or cleaning processes that affect the ability of the analytical method to determine the correct analyte level.

Measuring cleanliness is a difficult task. Essentially, trace residues on surfaces are the target analytes. The residue must first be extracted from a surface, recovered from the extraction medium, and then suitably quantitated. Residue analysis is quite different from analyzing bulk or formulated drug actives, as obtainable precisions and accuracies may be larger than the analyst is accustomed. Sensitivity levels of the techniques employed need to be considered for linearity, precision, and accuracy. The first decision to be made is the decision as to which residue will be measured. This residue could be the active drug, formulation excipients, or a component of the cleaner.[11] In most cases, the residue being analyzed has the potential to be a combination of all of these. The next step is to decide on the allowed residue limit,[12,13] followed by the choice of whether to use a specific or nonspecific technique. It is only after these decisions have been made, that the analytical technique can be selected.

Analytical method validation is the analysis of reproducibility of the method developed. There should not be any surprises in a validation study. All of the parameters required in a validation study need to be preliminarily evaluated during method development. Method development is the process by which the analytical chemist obtains the initial information to establish the limits and goals that are listed in a validation protocol, e.g., precision of 5% between analysts, linearity, accuracy, repeatability, etc. Understanding the required parameters during method development is a requirement for successful analytical method validation.

Regulations

The requirement for analytical method validation is identified in the Good Manufacturing Practice (GMP) regulations (21 CFR 211). The United States Pharmacopoeia (USP) provides a widely used standard for analytical method validation, and is probably the most often used reference regarding the subject.[14] The Food and Drug Administration (FDA) submitted guidelines for analytical method validation[15] in 1995 that correlate with the recommendations of the International Conference on Harmonization (ICH).[16] The ICH then issued a document describing different approaches that can be used in analytical method validation.[17] The FDA has also issued a guide for the validation of cleaning processes that state the need for validated analytical methods[18] in cleaning validations.

The ICH documents, along with the USP document, describe validation guidelines for methods used in different applications. The USP describes four different categories of methods. Category I methods involve the quantitation of major components of bulk drug substances or active ingredients in finished pharmaceutical products. Category II methods involve the determination of impurities in bulk drug substances, or degradation compounds in finished pharmaceutical products. These include quantitative and limit tests. Category III methods are used for the evaluation of performance characteristics, and Category IV methods are identification tests. The ICH guidelines define the same categories as the USP, except for USP Category III. *Table 1* lists the ICH categories and required parameters.

What parameters apply to analytical methods used in cleaning validations? The residues being determined are potential impurities. Therefore, the parameters that should be evaluated are most closely associated with Category II requirements (quantitative analysis of residues). It could be argued that the most important parameters are the limits of quantitation and detection, because these are the measures of sensitivity of the analytical method.

Chromatographic versus non-Chromatographic Methods

While various chromatographic methods, specifically High Performance Liquid Chromatography (HPLC), may be the more common methods used in analytical laboratories, there are certainly other applicable methods.[19,20] Some examples are Total Organic Carbon (TOC),[21] capillary electrophoresis[22], Atomic Absorption (AA), Inductively Coupled Plasma (ICP), titrations, ultraviolet spectroscopy, near infrared,[23] enzymatic,[24] etc. These "other" methods also require

Table 1

ICH method parameters.

Parameter	Identity	Impurities		Assay
		Quantitative	Limit	
Accuracy	-	+	-	+
Precision				
Repeatability	-	+	-	+
Intermediate	-	+	-	+
Specificity	+	+	+	+
LOD	-	-*	+	-
LOQ	-	+	-	-
Linearity	-	+	-	+
Range	-	+	-	+

* May be needed for some applications

validation, and while the USP and ICH parameters may appear more suitable for chromatographic methods, they are certainly applicable to non-chromatographic methods.[25] A good understanding of how to adapt and measure the required parameters using the specified analytical technique, is all that is required. In fact, various methods can be used to validate each other,[26] i.e., a mass spectroscopic technique could be used to determine the specificity of another technique.

Specificity

Specificity is the ability of a method to measure the analyte in the presence of components, which may be expected to be present. For cleaning validation methods, the potential presence of drug actives, formulation excipients, impurities, known degradation products, and cleaner components (if any)[27] should be anticipated. Experiments must be conducted that demonstrate the absence of interferences when the analyte is in a typical matrix.

Co-elution of components in chromatographic methods is typically the primary concern here. If HPLC with diode array capabilities is utilized, peak purity can be evaluated by examining the spectra across the peak. Most HPLC software programs will automatically calculate peak purity. Poor peak purity may be an indication of the presence of excipients, degradants, or cleaning components within the peak of interest.

Cleaning validation methods should be carefully evaluated for interferences. Studies should be conducted involving the analyte in the presence of the cleaning agent. If the cleaning agent is being quantitated, the effect of the drug active and formulation components should be evaluated. If the drug active or a formulation component is being analyzed, it must be shown that the cleaning process does not affect the analyte. This means that the effect of the cleaning process does not change the analyte in such a manner that it is no longer analyzable using the method being validated. One approach is to perform a recovery study of the analyte, by exposing it to the cleaning agent at use concentrations, time, and temperature. If suitable, recovery can be obtained, then the method can be used. If not, the method must be modified, or a new method developed.

Table 2

Data used for linearity evaluations

[Active X], ppm	Response	Standard Deviation	% RSD
0.10	613	72.5	11.80
0.40	900	100.0	11.00
0.70	998	33.3	3.33
1.1	1472	51.1	3.50
1.7	2398	53.0	2.21
2.5	3398	55.3	1.63
5.0	6800	100.0	1.50
10.0	13467	153.0	1.10
20.0	27133	208.0	0.77
30.0	40417	284.0	0.70
40.0	49500	500.0	1.00

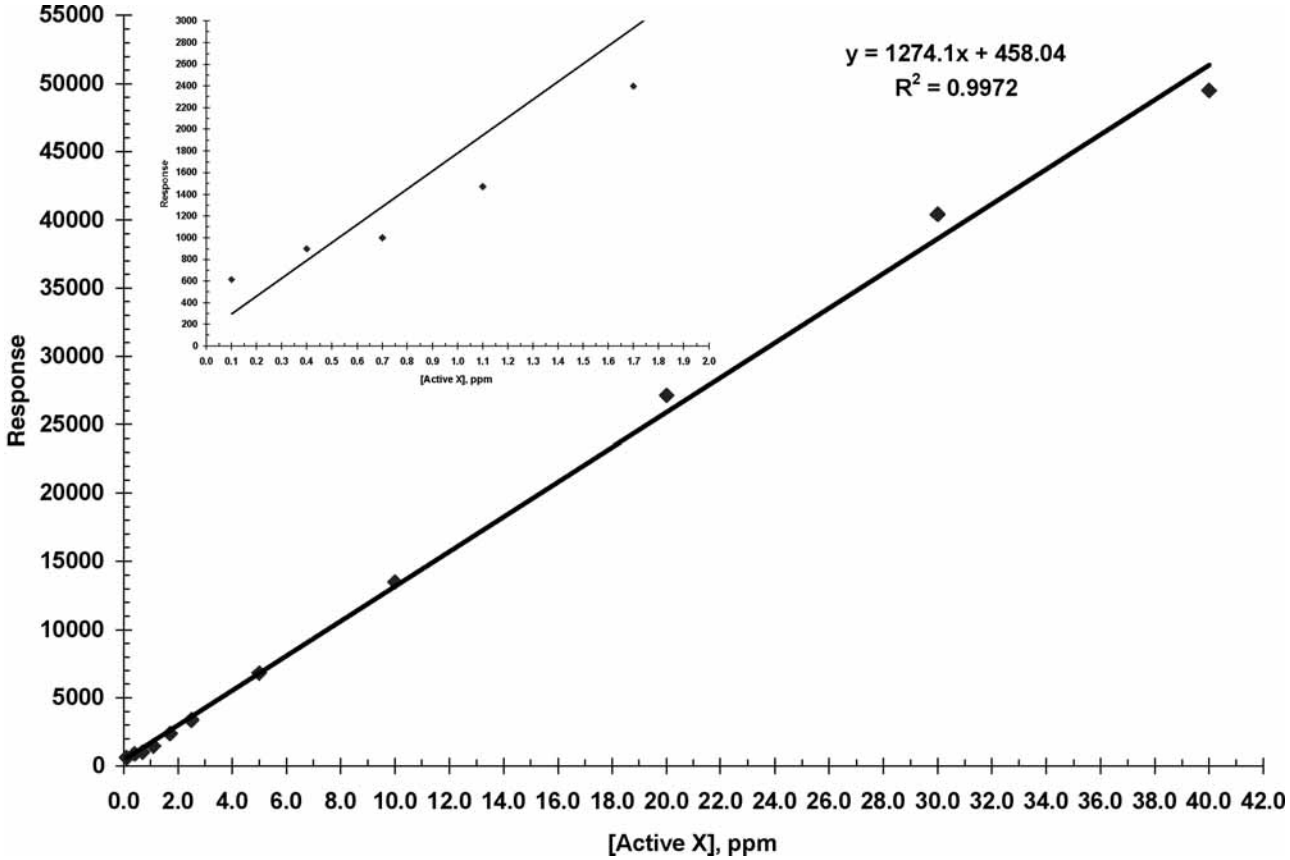
Linearity and Range

The linearity of a method is the ability of an assay to elicit a direct and proportional response to changes in analyte concentration. There are some detectors that produce, or have the ability to produce, non-linear responses (e.g., gas chromatography with flame photometric detectors and others, such as evaporative light scattering or mass spectrometers may have limited linear ranges when compared to flame ionization or UV-visible detectors). However, specific ranges could be found within the non-linear response that approach near linearity. If a non-linear curve must be used, a suitable number of points need to be utilized that will accurately describe the curve.

The range of a method is the interval between the upper and lower concentrations of analyte for which the method has been shown to have suitable precision, linearity, and accuracy. ICH recommends a range of 80 to 120% of the test concentration for finished drug products, 70 to 130% of the test concentration for content uniformity, and up to 120% of the reporting level for impurities, ensuring that the detection and quantitation limits are lower than the controlled level. For cleaning validation analysis, the range will potentially be much greater than what ICH recommends for drug actives and impurities. Generally, the range will extend from the limit of quantitation to perhaps 200% or greater for the amount of allowable residue in the sample. The wide range is important

Figure 1

Linear evaluation of all data points



in order to allow monitoring of the residue. If the low end of the range was equal to the allowable residue limit, process monitoring would not allow for early warning of potential problems. Again, the key requirement for the range is that it is suitable with regard to precision, linearity, and accuracy.

There are several methods described in industry literature to determine the linearity and the range of a method.[28, 29, 30] These methods range from simple observation to comprehensive statistical treatments. Statistical methods should be used to verify observed results. The method of choice will normally be dictated by a pharmaceutical company's policies. The specifics should be described in the validation, protocol if not in Standard Operating Procedures (SOPs).

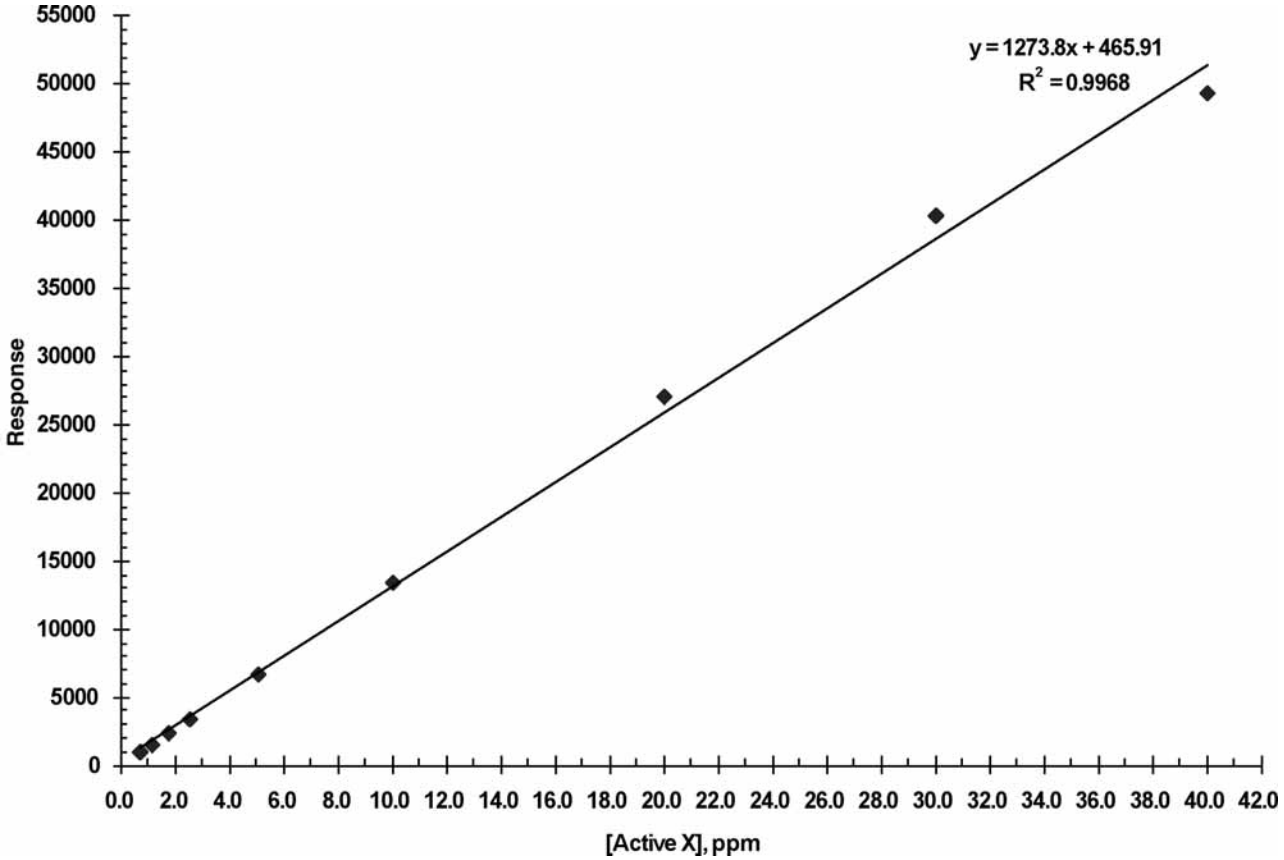
The linearity criteria are to be set prior to the validation. For example, for an HPLC analysis, the criteria for linearity may be that injections at each level must have an RSD < 3%, and the regression analysis yield an R^2 value of greater than 0.999 over the defined range. It should be noted that these requirements will vary greatly by method type and instrumentation in use. Acceptable %RSD values will also vary greatly, depending on the level of analyte. This will be covered in detail later.

The linearity experiment should include a minimum of five concentration levels. Each concentration level should be analyzed minimally in duplicate. Three analysis for each point are generally utilized, but more are preferred. The %RSD is then calculated for each injected level. This data is the base information set that will be utilized in establishing the linearity and range of the method.

Two simple methods for determining linearity will be described here. Both methods involve preparing a series of solutions containing known concentrations of the analyte of interest. This series encompasses the range of results expected from the analysis of actual samples. Usually, these solutions are free of expected matrix components. For the following examples, the pre-established (and somewhat arbitrary) criteria that was applied, is that all points must have a %RSD < 5 and $R^2 > 0.999$. These solutions would then be analyzed by

Figure 2

Linear evaluation less the first two data points



the method of choice. *Table 2, (see prior page)* presents the data that will be used for both examples. These data are only presented by way of example. A graph of the response versus the concentration shown earlier in *Figure 1*.

The first step would be to examine the %RSD values for each point. The very low concentrations of 0.10 and 0.40 ppm produce a %RSD of 12 and 11, respectively. Since these %RSDs are greater than the 5% RSD required, these data would be eliminated from the curve. The rest of the data meet the criteria. It should be noted that data cannot be arbitrarily eliminated. For example, if the 0.10 ppm data point produced an acceptable %RSD, but the 0.40 ppm point did not, the 0.40 ppm alone could not be eliminated. All data below 0.40 ppm would have to be eliminated. There may also be a need to investigate the discrepancy in RSD values.

Figure 2 shows the resulting graph after the lowest concentration levels have been eliminated. The R^2 value has decreased, due to the shortening of the range. Since the lower concentration levels have been eliminated, it is not surprising that the y-intercept has also increased. The y-intercept value is a good indication of bias. If the y-intercept is 0, no bias exists. Bias exists if the y-intercept deviates from 0. Statistical significance testing of the regression can provide evidence of bias. Examination of the y-intercept can be used to justify single point calibration criteria. If there is no statistical evidence of bias, or the bias is judged to be small, then single point calibration can be successfully used. In making this evaluation of the statistical evidence, it is the P-value and the upper and lower 95% confidence intervals for the y-intercept that are important to examine. These are typically values obtained from automated regression analysis as is shown in *Figure 3 (see next page)*. If the P-value is large and 0 is included in the confidence interval, there is no evidence that the y-intercept is anything other than 0.

Close examination of this calibration curve reveals that the highest concentration point (40 ppm) may be deviating from linearity. This concentration level is eliminated from the data set as a result of the fact that the inclusion of this data

Figure 3

Typical output from a regression

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.999930643
R Square	0.999861291
Adjusted R Square	0.999843952
Standard Error	339.9654599
Observations	10

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	6664892272	6664892272	57666.49335	1.01229E-16
Residual	8	924612.1115	115576.5139		
Total	9	6665816884			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	324.5939254	133.9655524	2.422965603	0.041656901	15.66860789	633.519243
X Variable 1	2684.578938	11.17929476	240.1384879	1.01229E-16	2658.799422	2710.358455

Figure 4

Linear evaluation less first two and the last data point. This produced quite acceptable results.

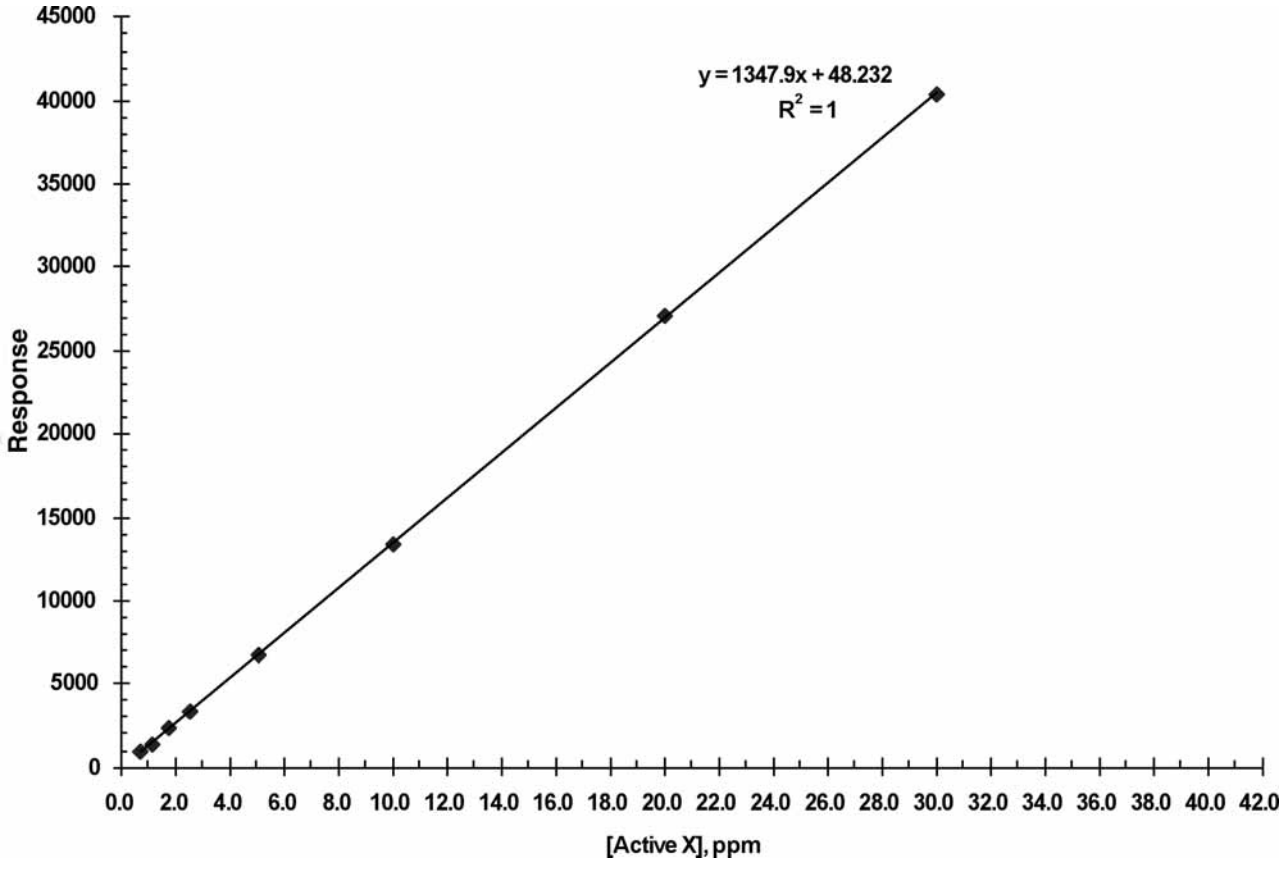
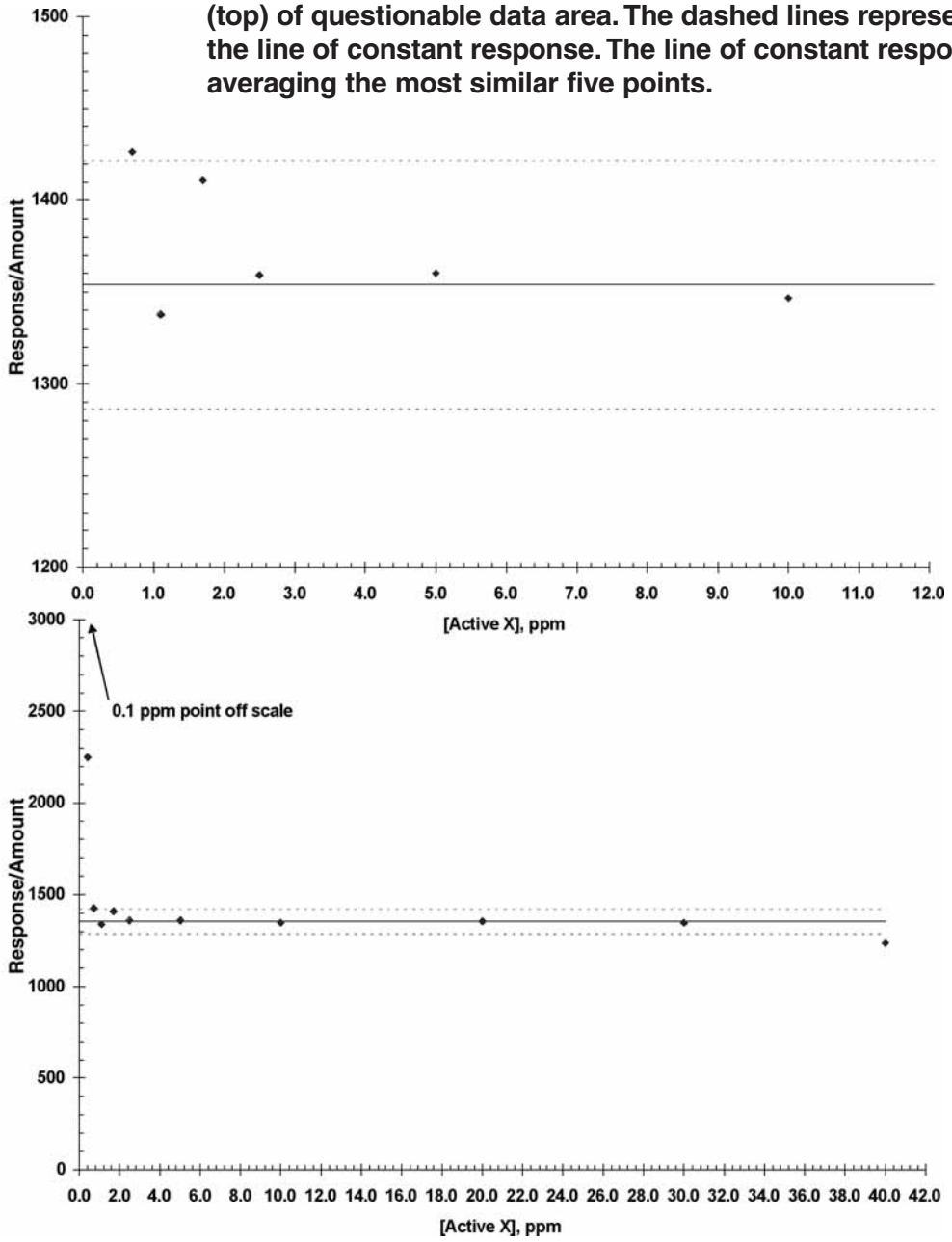


Figure 5

Graphical representation of linearity evaluation (bottom) with an exploded view (top) of questionable data area. The dashed lines represent 95 and 105% values of the line of constant response. The line of constant response was calculated by averaging the most similar five points.



point results in a failure to meet the regression acceptance criteria (R^2 = 0.999). *Figure 4* shows the resulting graph after eliminating the lowest two concentration points, along with the highest concentration point. The R^2 value now equals 1.000. This meets the pre-established criteria, and the linearity experiment has been completed. Note that the y-intercept has now also dramatically dropped. The use of a single point calibration for future analysis may be justified after checking for significance of the intercept. Based on the results of this example analysis, the linearity of the method has been established for the range of 0.70 ppm up to 30 ppm.

The responses of analytes across a linear range should be relatively constant.[28] The plot of the quotients of the responses and their respective concentrations versus the concentrations is shown in *Figure 5*. A line of constant response is drawn that will fit the more similar points. Two lines representing 95% and 105% of the constant response line are then drawn. Points that lie above 105% or below 95% are

considered to be outside of the linear range. As in the previous technique, the lowest two data points are obviously not linear, and are eliminated from the data set. The highest concentration point is also outside of this range, and can be eliminated. The third lowest concentration is borderline acceptable, as is shown in the top graph of *Figure 5*. The inclusion or elimination of this point is insignificant to the line. The elimination of this point does not change the slope, and only decreases the y-intercept from 48.232 to 46.439. The results of this second technique are in close agreement with the first technique for determining linearity.

Accuracy

The accuracy of measurements must be established. Accuracy is also referred to as "recovery" or "trueness." Accuracy is a measurement of how close the analytical result is to the actual value. This parameter is obtained through the evaluation of known analyte concentrations, or results from another accepted method. The result is typically expressed as a percentage of the calculated concentration. The accuracy needs to be determined across the analysis range. *Table 4* presents example data for a recovery or accuracy experiment.

Recovery studies provide assurances that an analyte can be analyzed in a matrix. The matrix can be a process sample containing a variety of reactants, formulation, swab, rinse sample, surface, body fluid, etc. The goal is to accurately define the amount of the analyte that can be recovered from a sample. In some cases, this amount will need to be as close to 100% as possible. In other cases, an acceptable recovery may be as low as 50%.[31]

In cleaning validation studies, both specific and nonspecific methods are commonly used. If a nonspecific method is being used (i.e., TOC) then the total response is usually assumed to be that of the analyte. Sampling is sometimes done by indirect methods (i.e., rinse solutions[32]), but direct methods (swabbing) are usually preferred. The choice is usually dependent on the accessibility of the surfaces to be sampled and the solubility of the residues.

When using analyte specific methods, i.e., HPLC, IMS for rinse samples (indirect sampling) it is generally sufficient to show that the presence of matrix components does not interfere with the determination of the analyte over the range required. Matrix components can be defined as excipients used in the formulation of a formulated drug, or impurities/solvents present in the bulk drug, depending on what is being manufactured. If a cleaning agent is used, the components of the cleaning agent should be considered. One approach for evaluating recovery is to spike solutions containing common or perhaps maximum possible levels of drug impurities, drug excipients, and cleaning formulation components with the analyte of choice, and demonstrate that the method results are close to the amounts calculated. The solvents used in these experiments should be representative of those used in the final rinse of the equipment being cleaned. Usually, these are aqueous-based solutions, but can be aqueous/solvent (such as alcohol) combinations or other solvents used in the manufacturing process. Several solutions are prepared having the appropriate concentrations of impurities, excipients, and cleaner components. The analyte is then spiked into these solutions at various concentrations. These concentrations should bracket the concentration that represents the allowable limit of the analyte. The range used is typically 25 to 150% of the allowable limit of analyte, although values ranging from 10 to 500% of the allowable limit have been used. Again, the stability of the analyte in these solutions should be documented.

For swab samples, the chore becomes a little more tedious, as the ability of the swab to remove the analyte from the surface being cleaned, ability of the sample preparation procedure to remove the analyte from the swab, and the potential of interferences from impurities, formulation excipients, cleaning agents, and the swabbing material itself need to be considered. Assuming that the analytical method is thought to have sufficient accuracy, all of these parameters can be studied as a whole. If poor results are obtained, then separate testing of each step cannot only identify the step re-

Table 4

Recovery data

Found [Active X], ppm	Theoretical [Active X], ppm	% Recovered
0.691	0.701	98.6
0.710	0.701	101.0
0.693	0.701	98.9
9.970	9.990	99.8
9.950	9.990	99.6
10.100	9.990	101.0
29.800	30.100	99.0
30.500	30.100	101.0
29.600	30.100	98.3
	Average:	99.7
	Standard Deviation:	1.17

sponsible for the poor results, but also allow the testing of alternate solvents or swab materials that lead to improved recovery. Ideally, the accuracy of these different steps is considered during method development so that the analyst is not surprised during the validation process. As mentioned before, there should not be any surprises during validation. If there are, then sufficient method development was not done.

In a typical experiment, the surface area that is to be swabbed during the cleaning validation is established. This is usually based on the levels of analyte expected and/or the sensitivity of the method. Coupons representing the composition and size of this surface(s) are obtained (e.g., 5x5 inch sections of 304 stainless steel, 316 stainless steel, Teflon®, Tygon®, or other product contact materials). Solutions of the analyte and matrix components are prepared, and aliquots evenly applied to the coupon surface. The coupon is allowed to dry for a specified period (e.g., overnight). As an example, assume that the active ingredient of a formulation or a cleaner is the target analyte with an allowable limit in the final prepared sample of 20 µg/mL. Solutions containing 1500, 1000, 250, and 0 µg/mL of the analyte, as well as appropriate amounts of cleaner components, formulation excipients, and/or impurities, are prepared. Since the solvents are evaporated, the solvent composition of this solution is usually less important than in rinse samples, but should be similar to those used in the cleaning process. Typically, a 100 µL aliquot of each solution is applied as evenly as possible to separate coupons. This can be accomplished by slowly dragging the pipette tip across the coupon surface in a uniform row pattern. The residue is allowed to dry for several hours, typically overnight. Care should be taken to ensure the coupons dry on a flat surface, and are protected from contamination during the drying process. If the equipment being cleaned is heated, then the coupons can be placed in an oven to simulate the temperatures seen during the cleaning process. Once dry, the analyte is recovered for analysis by rubbing a swab uniformly across the surface of the coupon. Cotton and synthetic swabs have been cited in industry literature. The type of swab used will depend on the method of analysis.[21] Prior to swabbing the surface, the swab is usually dipped into the solvent that will be used to extract the residue. The excess solvent is removed by squeezing the swab against the inside of the vial containing the extraction solvent. The swab is usually rubbed along the coupon surface side-to-side from both a vertical and horizontal perspective. The swab is periodically rotated, so that both sides of the swab surface are used. In some instances, a second swab can be used to pick up any remaining analyte from the coupon surface. Oftentimes, the second swab is dry and picks up excess solvent and residue left by the first swab. Sometimes, the second swab is wetted with the same extraction solvent to help pick up stubborn residuals still dried to the coupon surface. Any swabbing pattern can be used, as long as the same pattern is used, both in the recovery study from the coupon and in the actual sampling of the equipment. The important consideration is that it is done consistently each and every time. Insufficient training is often a source of errors. Appropriate diagrams should be placed in SOPs, accompanied by specific training and qualification of the person(s) performing the swabbing.

The analyte can then be extracted from the swab by placing the swab in the same extraction vial used to wet the swab. This vial contains the appropriate amount of extraction solvent. Ideally, this solvent is appropriate for direct use in the analysis. For TOC, ultra-pure water is used as the solvent. With an HPLC method, water, mobile phase or a solvent (weaker than the mobile phase) can be used. If a stronger solvent is used, dilution or other modifications may be necessary prior to analysis. In this example, 5mL of water or mobile phase would be used. Assuming 100% recovery, these final solutions would represent 150, 100, 25, and 0% of the allowable limit. It is common practice to cut the handle from the swab (swabs are available that break upon bending), and allow the portion of the swab that came in contact with the coupon to be fully immersed in a minimum amount of extraction solvent. The swab is often allowed to sit immersed for long periods or sonicated to produce the highest recoveries. Studies should be conducted to be certain that samples are stable if long periods are expected between the sampling time and the time of analysis. The extraction solvent can then be filtered (if necessary) and analyzed. TOC samples are generally not filtered. Results are expressed as % recovery, by comparing the experimentally found results with the theoretical values. The 0 concentration (blank) experiment can be used to determine whether any interferences are present in the system. Ideally, a system is found where the blank has no effect on the results. It is acceptable, however, for the blank values to be subtracted from the other results to adjust for possible interferences.

Acceptable levels of recovery for analytical methods used in cleaning validation are a source of debate. The FDA's guideline on cleaning validation[18] states that the % recovery should be determined, and lists 50% and 90% recoveries as examples. The recovery value should either be used in calculating the found value, or factored into the acceptance criteria, but not included as a factor in both. Acceptance crite-

ria based on the results obtained with a particular designated method, is a logical approach.

For example, the recovery for a specific sampling method was determined to be 75%. The acceptable limit in an analytical sample had been previously calculated to be 5 ppm in the analytical sample (independent of recovery). A sample is analyzed and found to contain 4 ppm of the target analyte. Utilizing the recovery value, the actual value for the sample is 4 ppm / 0.75 = 5.33 ppm. This is above the acceptance criteria, and the sample fails. The recovery factor can be applied to the acceptance criteria: 5 ppm x 0.75 = 3.75 ppm. This would then be the new limit per analytical sample. Again, be certain not to apply the factor to both the analytical result and the limit criteria.

If the % recoveries are lower than desired, additional recovery experiments should be conducted in such a way that the major source of the reduced recovery can be ascertained. A common experiment is to spike the swab material with the solutions that were used to prepare the coupons. If the spiking solution is different than the extraction solution, the swabs are allowed to dry, and then placed directly into the extraction solvent. If this experiment yields low recoveries, stronger extraction solvents or alternate swabbing materials can be tried. If the recoveries are acceptable, the problem can be assumed to arise from removing the residue from the coupon itself. Again, alternate swabbing materials, stronger swabbing/extraction solvents, or a variation in the swabbing technique (i.e., more swabs, wet/dry variations, more forceful swabbing) can be tried. A reexamination of the potential volatility of the analyte should not be overlooked. This is especially true if a heating or vacuum oven is used in the drying process.

Precision

Precision is a measurement of how close a series of numbers are to each other with no regard to the actual value (accuracy). Precision is broken down into three subgroups: repeatability, intermediate precision, and reproducibility. Precision is typically expressed as a percentage of the relative standard deviation.

Repeatability refers to the precision under the same operating conditions over a short interval of time. For example, the same sample may be injected six consecutive times and the results evaluated for each injection with respect to one another. This is a measurement of instrument precision, such as the injector and integrator performance. *Table 5* presents example data obtained for a repeatability experiment using six analysis. If an auto-injector is utilized, care must be taken to assure that the septa are not damaged from repeated penetrations. This may introduce errors into the precision experiments. It is wise to limit the number of injections from a single vial. For non-chromatographic methods that utilize a sampling device, such as a cell, the repeatability can be measured by loading the cell, scanning, emptying the cell, and repeating this procedure for up to six replicates. Concentration may not be the only parameter measured for precision. The wavelength, retention time, etc. may be important parameters to characterize.

Table 5

Repeatability data

Injection	Result
1	0.691
2	0.705
3	0.699
4	0.705
5	0.715
6	0.690
Average:	0.701
Standard Deviation:	0.00952
%RSD:	1.36

Intermediate precision evaluates intra-laboratory variations. For example, samples may be analyzed on different days by various analysts using different equipment (or at least a different column if a chromatographic method is being evaluated). Intermediate precision evaluates the precision between analysts and instruments. It has the potential to account for often overlooked environmental conditions. Fluctuations in lab temperature, or even humidity, may add (or subtract) to the intermediate precision. Intermediate precision can be determined by having a second analyst repeat the accuracy and/or repeatability study, performing three replicates at each level being determined, or six replicates at the target value. If duplicate equipment is available, the analyst should use a different piece of equipment (i.e., a second HPLC or TOC). If a chromatographic method is employed, a second column, preferably from a different lot or even a different supplier, should be used. For swab studies, alternate sources or lots of swabs may be included. *Table 6 (see next page)* presents example data for an intermediate precision experiment.

Reproducibility evaluates the precision between laboratories. This is utilized when methods are transferred, or when collaborative method standardizations are being made. One such example is the transfer of an analytical method from a Research and Development (R&D) laboratory to a quality control laboratory.

In most instances, since materials of construction and cleaning procedures are generally site specific, validations of analytical methods used in cleaning validations will be concerned with repeatability

Table 3

Simplistic evaluation of precision as a function of analyte level

	Level 1	Level 2	Level 3	Level 4	Level 5
Injection 1	100	75	50	25	5
Injection 2	101	76	51	26	6
Injection 3	99	74	49	24	4
Average:	100	75	50	25	5
Standard Deviation:	1.0	1.0	1.0	1.0	1.0
%RSD:	1.0	1.3	2.0	4.0	20

and intermediate precision. If another facility within an organization does not have the exact same procedures and equipment, reproducibility may be of concern. Including other facilities in the validation or performing method transfer after validation should be considered.

ICH guidelines recommend that a minimum of nine determinations throughout the range of interest or six determinations at the test concentration be used to determine repeatability. The limits of quantitation are oftentimes being approached during validation of the analytical method. Repeating the accuracy testing outlined above, so that three data points are collected for each level, allows separate estimates of %RSD to be calculated at each of the three levels. Additional data points at each level are a benefit, and additional statistical testing can be used to determine if the different levels have different %RSDs.

Table 6

Intermediate precision data

Sample	Analyst A Result	Analyst B Result	Combined
1	**0.715**	**0.702**	
2	**0.725**	**0.720**	
3	**0.730**	**0.699**	
4	**0.715**	**0.691**	
5	**0.709**	**0.698**	
6	**0.720**	**0.695**	
Average:	**0.719**	**0.701**	**0.710**
Standard Deviation:	**0.00762**	**0.0101**	**0.0128**
%RSD:	**1.06**	**1.44**	**1.80**

It is not uncommon for the accuracy and repeatability of the method to change with the amount of analyte being determined. It is also common for the %RSD to be a function of the accuracy. If there are significant changes in accuracy, as the amount of analyte determined decreases, it is not unusual to also see the repeatability of the method decrease with decreasing amounts of analyte.[31, 33]

A simple example of this is shown in Table 3. If five levels of an analyte are determined, and in each case, the differences in three injection values differ by only 1 ppm from the average, the precision can vary significantly. In this case, the repeatability ranged from 1% for a 100 ppm sample, up to 20% for a 5 ppm sample. This is something that is also often overlooked when establishing acceptance criteria for trace levels of analyte. It is a rare case indeed that identical precision will be obtained for both bulk analysis and trace analysis of the same analyte.

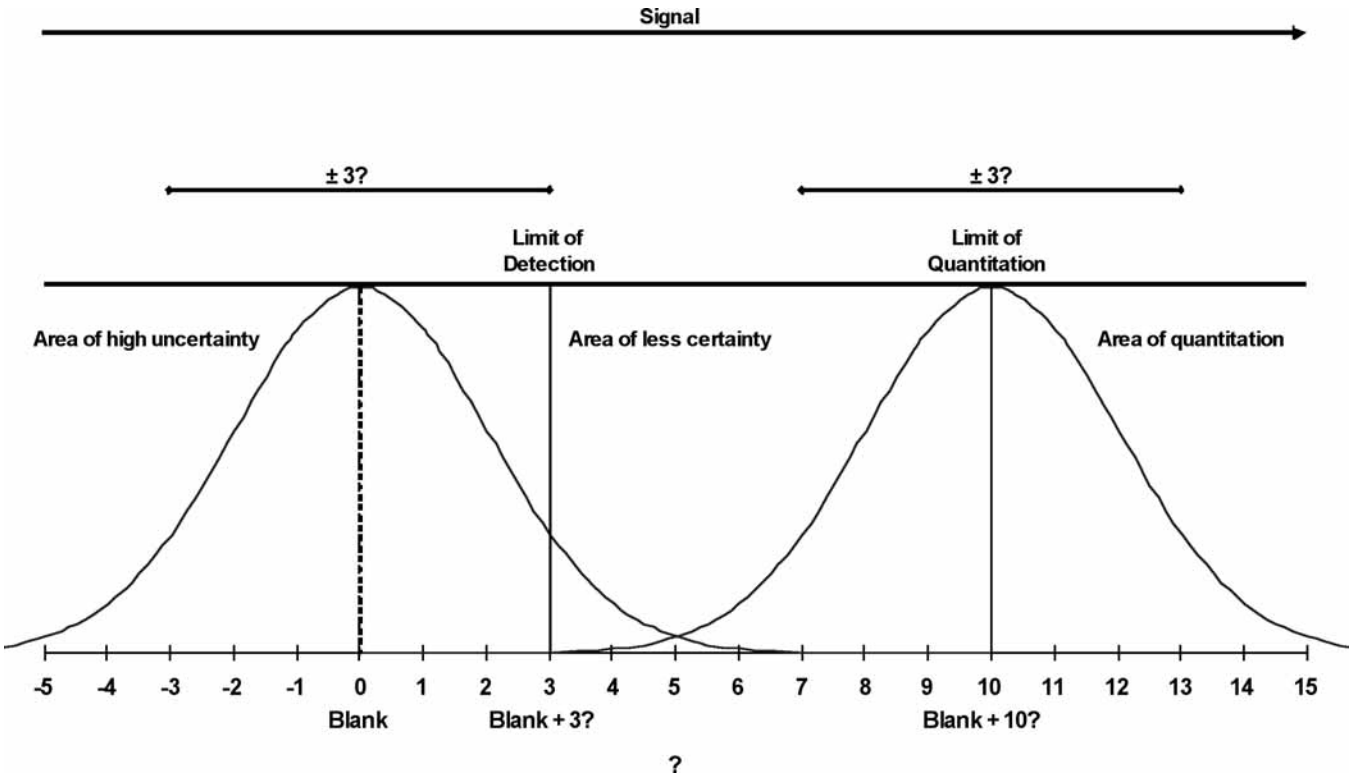
Limits of Detection and Quantitation

The American Chemical Society (ACS) has extensively addressed the concept of Limits Of Detection (LOD) and Limits Of Quantitation (LOQ).[34,35] *Figure 6 (see next page)* presents a graphical representation of LOD and LOQ. The LOD of an individual analytical procedure is the lowest amount of analyte in a sample that can be detected, but not necessarily quantitated as an exact value. The LOQ of an individual analytical procedure is the lowest amount of analyte in a sample that can be quantitatively determined with suitable precision and accuracy.

In routine assays of bulk or formulated analytes, the LOD and LOQ values are generally not of great concern. This is because the levels of analytes are generally large, and the methods are applied well above the LOQ value. In cleaning validation analysis, this is not the case, because the analysis

Figure 6

Graphical comparison of the limits of detection and quantitation relative to the signal from a blank.



are typically looking for trace residues, and the results are oftentimes "bumping" against these values. These values need to be defined during the validation of the analytical methods used for cleaning validation.

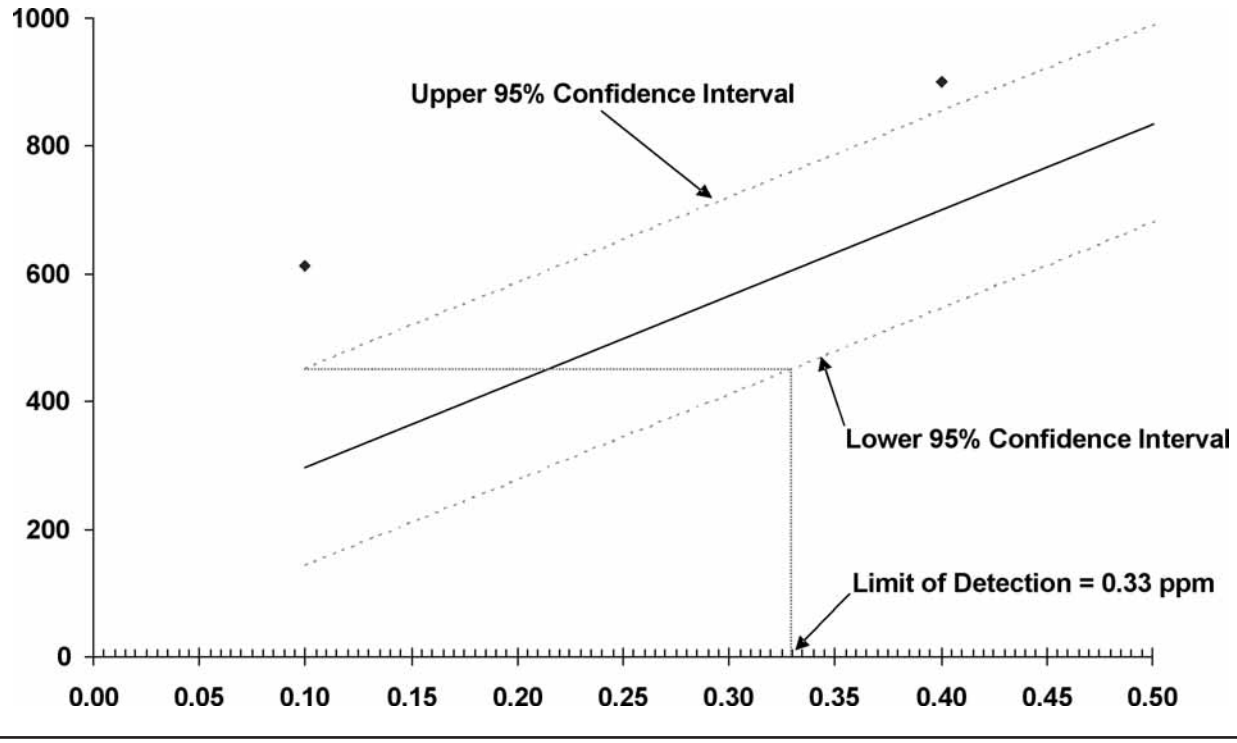
The traditional method is based on signal-to-noise ratio. The majority of chromatographic and spectroscopic data analysis software available today incorporate a means of measuring the noise. The LOD would then be the concentration equivalent to three times the signal-to-noise ratio. The LOQ would be the concentration equivalent to ten times the signal-to-noise ratio. A common method for establishing LOD and LOQ is the same as that used for establishing the lower limit for the range of the method. The LOQ is the concentration that meets a certain pre-established precision and satisfies the linearity requirement. In the example used for the linearity and range experiment, the LOQ was found to be 0.70 ppm. The LOD would then be (3/10) x LOQ or 0.21 ppm.

There are numerous approaches to calculating the LOD and LOQ values.[17,36,37,38] There is, however, no specific method recommended in any regulatory document. The establishment of these values is up to the analyst. The method used for determining LOD and LOQ values needs to be written in an official document, such as the analytical method or an SOP. Residue limits should not be based on LOD values, since values measured below the LOQ have an inherently high degree of uncertainty. Likewise, if a process is to be monitored with alert and action limits in place, residue limits should not be based on the LOQ, if at all possible. A single approach or methodology of LOD and LOQ determination should never be blindly followed. The good judgment of a well-trained, experienced analyst should be regarded as a valuable asset. The discussion of several methodologies found in the literature follows.

Figure 7 on the facing page, shows an example of one such method for the determination of LOQ and LOD using confidence intervals.[39] Data are collected approaching the LOD, plotted and fitted with a regression line. The upper and lower 95% Confidence Intervals (CI) for the regression are then plotted. This example shows the CIs to be linear, but they typically are not. A line is then drawn horizontally from the end of the upper CI to the lower CI parallel to the x-axis. A line is then dropped vertically to the x-axis. The intersection point on the x-axis is regarded as the LOD. The

Figure 7

Graphical determination of the limit of detection (LOD). A horizontal line is drawn from the lower end of the upper 95% CI line to the lower 95% CI. A vertical line is then dropped from this point to the x-axis. The point on the x-axis is the LOD.



LOQ is then 3.3 x LOD. For this example, the LOD was found to be 0.33 ppm and the LOQ was 1.1 ppm.

Another approach makes use of the slope and standard error of the regression line.[17,37] *Figure 8* provides an example using the same data as *Figure 7*. The LOD and LOQ values of 0.84 and 2.5 ppm, respectively, are defined by Equation 2 and Equation 1, where σ is the standard error of the regression, and S is the slope of the regression line.

All of the methods used here *(See Equations 1 and 2)* for the LOD and LOQ calculations were based on the same set of hypothetical data. These calculations produced different results. The statistical significance of these differences can be argued, but in actuality they are quite similar. In most real-world cases, the experimentally determined LOQ will be higher than the values that are calculated.

The LOD value, being free of stringent precision requirements, will be primarily based on instrumental or technique considerations, whereas LOQ values will not only be based on the instrument or technique, but also on all of the steps and manipulations included in the method. In addition, all of these literature LOD and LOQ techniques assume that there is a constant matrix that contributes to a constant noise level. In cleaning validations, where residues are being recovered from the surfaces of manufacturing equipment, a constant matrix cannot be assumed (this is one reason why residue limits based on the LOD or LOQ of a method are not advisable).

Equation 1

LOD calculation from linear fit

$$LOD = \frac{3.3\sigma}{S} = \frac{3.3(339.97)}{1342.3} = 0.84\ ppm$$

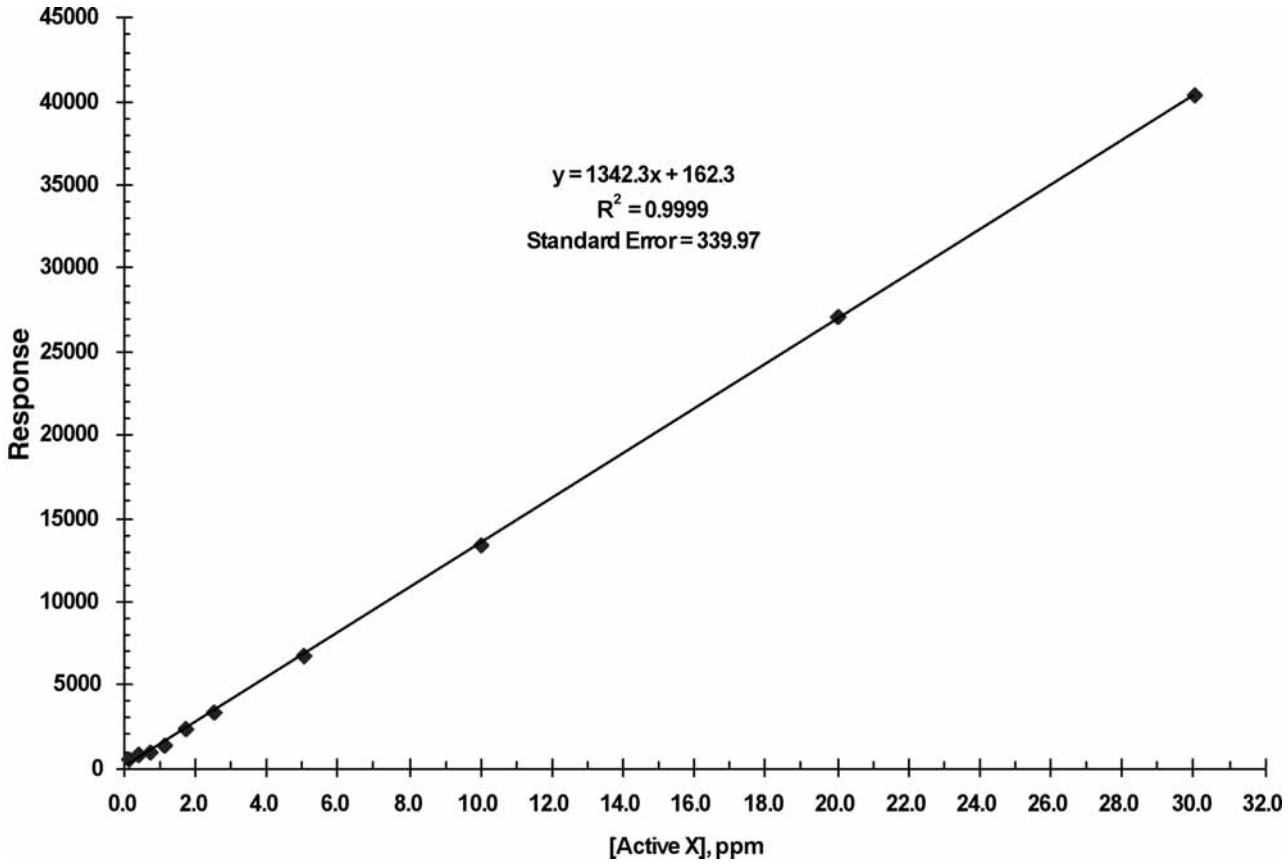
Equation 2

LOQ calculation from linear fit

$$LOQ = \frac{10\sigma}{S} = \frac{10(339.97)}{1342.3} = 2.5\ ppm$$

Figure 8

Data plotted and fitted with a linear regression line. The slope and standard error obtained are used in the calculations for LOD and LOQ.



Ruggedness

All analytical methods that are used routinely by a variety of personnel should be rugged. That is, a method should not be affected by small changes that may occur on a day-to-day basis. For example, a chromatographic method that ordinarily does not use a column heater should not be affected by changes of temperature in a laboratory. If it is, then this should be evaluated and defined.

There are no regulatory requirements for ruggedness and ruggedness studies, and do not need to be included in a validation report. However, it is assumed by regulatory agencies that the effects of changes to various parameters of the method have been explored during method development, that those parameters that do affect the method are controlled, and that these controls are included in the method.

Again, the ruggedness of a method is usually explored during method development. Ruggedness can be explored through experimental designs[40] or by simply varying parameters in a stepwise process. Examples of parameters that could be studied are column temperature, injection size, flow rates, buffer concentrations, pH, solvent concentrations, etc.

Summary

The output of an analytical method validation is an analytical method validation report. It is suggested (but not required) that the report follow the same format as other documents in the validation process. The similar formatting will make it easier for both internal and external agencies to understand and follow. The report should contain all of the parameters studied, along with whether or not the pre-established criteria were met.

The validation of an analytical method is the last step in the process prior to implementation. The various parameters required in an analytical method validation are first explored during method development, and then rigorously defined

during the method validation process. These parameters are dictated by various regulatory bodies around the world, and necessitated by the need to understand the process. The validation of analytical methods should not only be understood by the chemists who perform them, but also by the personnel involved in cleaning validation, as it is an integral part of the process. It is important that all personnel involved in a cleaning validation understand that changes in the overall process may affect the method and lead to the necessity for revalidation. Industry literature is full of information, resources, and examples that should be called upon and utilized. A chemist is only as good as her/his resources. ❑

About the Author

Herbert J. Kaiser, Ph.D., Senior Manager – Analytical Services and Development, at STERIS Corporation, has twenty-two years experience in cleaning and surface technologies. Dr. Kaiser received his B.A. degree from St. Mary's University in San Antonio, Texas, his M.S.(R) from St. Louis University, and his Ph.D. from the University of Missouri. He is a member of the American Chemical Society, and the editorial advisory board for the Journal of Validation Technology.

Bruce Ritts, M.S., Senior Scientist – Analytical Services and Development, at STERIS Corporation, has twenty years of expedience in the development and validation of analytical methods. Mr. Ritts received both his B.S. and M.S. degrees from the University of Missouri. He is a member of the American Chemical Society.

References

1. "Guidelines for Single-Laboratory Validation of Analytical Methods for Trace-Level Concentrations of Organic Chemicals" Special Publication - Royal Society of Chemistry: Principles and Practices of Method Validation, 2000, 256, 179-252.
2. Chowdary, K. P. R.; Rao, G. D.; Himabindu, G. "Validation of Analytical Methods" Eastern Pharmacist, 1999, 42(497), 39-41.
3. Hsu, H.; Chien, C. "Validation of Analytical Methods: A Simple Model for HPLC Assay Methods" Yaowu Shipin Fenxi, 1994, 2(3), 161-76.
4. Collard, J.; Kishi, Y.; Costanza, C. "Analytical Method Validation in the Semiconductor Industry Applied to High Purity Hydrogen Peroxide" 2001 Semiconductor Pure Water and Chemicals Conference, 20th, 190-213.
5. Seno, S.; Ohtake, S.; Kohno, H. "Analytical Validation in Practice at a Quality Control Laboratory in the Japanese Pharmaceutical Industry" Accreditation and Quality Assurance, 1997, 2(3), 140-145.
6. Clarke, G. S. "The Validation of Analytical Methods for Drug Substances and Drug Products in UK Pharmaceutical Laboratories" Journal of Pharmaceutical and Biomedical Analysis, 1994, 12(5), 643-52.
7. Kirsch, R. B. "Validation of Analytical Methods Used in Pharmaceutical Cleaning Assessment and Validation" Pharmaceutical Technology (Supplement), 1998, 40-46.
8. Frewein, E.; Geisshuesler, S.; Sibler, E. "Pharmaceutical Quality Assurance. Validation of Analytical Methods in Vitamin Assays" Swiss Chem, 1996, 18(7/8), 10-16.
9. Hokanson, G.C. "A Life Cycle Aproach to the Validation of Analytical Methods During Pharmaceutical Product Development. Part I: The Initial Method Validation Process" Pharmaceutical Technology, 1994, 18(9), 118-130.
10. Hokanson, G.C. "A Life Cycle Approach to the Validation of Analytical Methods During Pharmaceutical Product Development. Part II: Changes and the Need for Additional Validation" Pharmaceutical Technology, 1994, 18(10), 92-100.
11. Kaiser, H. J.; Tirey, J. F.; LeBlanc, D. A. "Measurement of Organic and Inorganic Residues Recovered from Surfaces" Journal of Validation Technology 1999, 6(1), 424-436.
12. Fourman, G. L.; Mullen, M. V. "Determining Cleaning Validation Acceptance Limits for Pharmaceutical Manufacturing

Article Acronym Listing

AA:	Atomic Absorption
ACS:	American Chemical Society
CI:	Confidence Interval
FDA:	Food and Drug Administration
GMP:	Good Manufacturing Practice
HPLC:	High Performance Liquid Chromatography
ICH:	International Conference on Harmonization
ICP:	Inductively Coupled Plasma
LOD:	Limits of Detection
LOQ:	Limits of Quantitation
R&D:	Research and Development
RSD:	Relative Standard Deviation
SOP:	Standard Operating Procedure
TOC:	Total Organic Carbon
USP:	United States Pharmacopoeia
UV:	Ultra Violet

Operations" Pharmaceutical Technology, 1993, 17(4), 54-60.
13. LeBlanc, D. A. "Establishing Scientifically Justified Acceptance Criteria for the Cleaning Validation of APIs" Pharmaceutical Technology, 2000, 24(10), 160-168.
14. <1225> Validation of Compendial Methods" United States Pharmacopoeia XXVI, 2003.
15. Food and Drug Administration "International Conference on Harmonization; Guideline on Validation of Analytical Procedures: Definition and Terminology; Availability" Federal Register, 1995, 60 (40), 11260-11262.
16. International Conference on Harmonization of Technical Requirements for the Registration of Pharmaceuticals for Human Use "Validation of Analytical Procedures" ICH-Q2A, Geneva, 1994.
17. International Conference on Harmonization of Technical Requirements for the Registration of Pharmaceuticals for Human Use "Validation of Analytical Procedures: Methodology" ICH-Q2B, Geneva, 1996.
18. Food and Drug Administration "Guide to Inspections, Validation of Cleaning Processes" Office of Regulatory Affairs, FDA, Rockville, MD, 1993.
19. Kaiser, H. J.; Minowitz, M. "Analyzing Cleaning Validation Samples: What Method?" Journal of Validation Technology 2001, 7(3), 226-236.
20. Kaiser, H. J. "Methods for Pharmaceutical Cleaning Validations" in Surface Contamination and Cleaning; Mittal, K.L., Ed.; VSP BV: Utrecht, The Netherlands, 2003; Vol. 1, pp 75-84.
21. Jenkins, K. M., Vanderwielen, A. J., Armstrong, J. A., Leonard, L. M., Merphy, G. P., Piros, N. A., "Application of Total Organic Carbon Analysis to Cleaning Validation" PDA Journal of Pharmaceutical Science & Technology, 1996, 50, 6-15.
22. Altria, K. D.; Chanter, Y. L. "Validation of a Capillary Electrophoresis Method for the Determination of a Quinolone Antibiotic and Its Related Impurities" Journal of Chromatography A, 1993, 652, 459-463.
23. Ciurczak, E. "Validation of Spectroscopic Methods in Pharmaceutical Analysis" Pharmaceutical Technology, 1998, 22(3), 92-102.
24. Goez, C. E.; Luu, T. D.; Meier, D. J. "The Luciferin-Luciferase Assay. An Example of Validating an Enzymic Analytical Method" Swiss Chem, 1996, 18(6), 9-11.
25. Brittain, H. G. "Validation of Nonchromatographic Analytical Methodology" Pharmaceutical Technology, 1998, 22(3), 82-90.
26. Van Zoonen, P.; Van 't Klooster, H.; Hoogerbrugge, R.; Gort, S. M.; Van De Wiel, H. J. "Validation of Analytical Methods and Laboratory Procedures for Chemical Measurements" Arhiv za Higijenu Rada i Toksikologiju 1998, 49(4), 355-370
27. Vessman, J. "Selectivity or specificity? Validation of Analytical Methods from the Perspective of an Analytical Chemist in the Pharmaceutical Industry" Journal of Pharmaceutical and Biomedical Analysis, 1996, 14(8-10), 867-869.
28. Huber, L. "Validation of Analytical Methods: Review and Strategy" LC-GC International, February 1998, 96-105.
29. Boque, R.; Maroto, A.; Riu, J.; Rius, F. X. "Validation of Analytical Methods" Grasas y Aceites 2002, 53(1), 128-143.
30. Kirsch, R. "Validation of Methods Used in Pharmaceutical Cleaning Validation" Pharmaceutical Technology, 1998 (Supplement), 40-46.
31. Ellis, R. L., "Validation and Harmonization of Analytical Methods for Residue Detection at the International Level" in Residues of Veterinary Drugs and Mycotoxins in Animal Products; Enne, G.; Kuiper, H. A.; Valentini, A., Eds.; Wageningen Pers: Wageningen, The Netherlands, 1996; pp 52-62.
32. LeBlanc, D. A. "Rinse Sampling for Cleaning Validation Studies" Pharmaceutical Technology, 1998, 22(5), 66-74.
33. AOAC Peer Verified Methods Program "Manual on Policies and Procedures" Arlington, Virginia, 1993.
34. American Chemical Society Committee Report "Guidelines for Data Acquisition and Data Quality Evaluation in Environmental Chemistry" Analytical Chemistry, 1980, 52(14), 2242-2249.
35. American Chemical Society Committee Report "Principles of Environmental Analysis" Analytical Chemistry, 1983, 55(14), 2242-2249.
36. Paine, T. C.; Moore, A. D. "Determination of the LOD and LOQ of an HPLC Method Using Four Different Techniques" Pharmaceutical Technology, 1999, 23(10), 86-90.
37. Krull, I.; Swartz, M. "Determining Limits of Detection and Quantitation" LC-GC, 1998, 16(10), 922-924.
38. Thomsen, V.; Schatzlein, D.; Mercuro, D. "Limits of Detection in Spectroscopy" Spectroscopy, 2003, 18(12), 112-114.
39. Hubaux, A.; Vos, G. "Decision and Detection Limits for Linear Calibration Curves" Analytical Chemistry, 1970, 42(8), 849-855.
40. Virlichie, J. L.; Ayache, A. "A Ruggedness Test Model and Its Application for HPLC Method Validation" S.T.P. Pharma Pratiques, 1995, 5(1), 49-60.

Good Analytical Method Validation Practice

Deriving Acceptance Criteria for the AMV Protocol: Part II

Stephan O. Krause, Ph.D.

Bayer HealthCare Corporation

To avoid potential inspection observations for test method validations by the regulatory agencies, it has become critical for pharmaceutical companies to derive reasonable acceptance criteria for the Analytical Method Validation (AMV) protocol. Part I of Good Analytical Method Validation Practice (GAMVP) (November 2002 issue, *Journal of Validation Technology*) mostly emphasized ground rules for the AMV department to be compliant and efficient within the Quality Control (QC) unit. The scope of this article provides more detail on how to systematically derive reasonable acceptance criteria for AMVs, and how to integrate those into the AMV protocol. One specific example to describe the process of deriving AMV acceptance criteria is provided. This example summarizes most aspects to be considered in order to generate an AMV protocol that can readily be executed, and lead to a solid AMV report.

"The scope of this article provides more detail on how to systematically derive reasonable acceptance criteria for AMVs, and how to integrate those into the AMV protocol."

For successful AMVs, available data and other supporting information for the test method to be validated must be carefully reviewed against current in-process or product specifications. This process takes time and requires a certain expertise, since acceptance criteria should balance method performance expectations with method requirements (from product specifications) and AMV execution conditions (conducted by QC analysts under routine QC testing). In general, the time required to prepare the AMV protocol should account for about 50% of the total time allocated to the complete (approved) validation. Less time spent on the protocol may result in time-consuming discrepancy reports, and validation retesting when acceptance criteria failed during execution. Or, the acceptance criteria do not sufficiently challenge the test system suitability, so this validation failed to demonstrate that this method will yield accurate and reliable results under normal testing conditions. In addition, invalid and potential Out-Of-Specification (OOS) results may be obtained when test system suitability is not properly demonstrated. Management should keep in mind that a rigorous AMV program, employing reasonable acceptance criteria, may prevent discrepancy reports, OOS results, and potential product loss, since unsuitable test methods should not be used for routine QC testing.

Selecting Assay Categories

When an AMV is generated to demonstrate test system suitability to bring a routine testing procedure into compliance, an assay category must be selected. Guidelines published by the International Conference on Harmonization (ICH), United States Pharmocopeia (USP), and the Food and Drug Administration (FDA) are similar in content and terminology used. Following ICH guidelines is advisable when product is distributed worldwide. This article will only focus on following ICH guidelines. The FDA accepts these guidelines as long as those are consistently followed, as intended by the ICH. The ICH Q2A guidelines list four assay categories:

- Category I: Identification Tests
- Category II: Quantitation of Impurities
- Category III: Qualitative Limit Test for Impurities
- Category IV: Quantitation of Active Ingredients

Once an assay category is appropriately chosen, all validation parameters required for that category must be included in the AMV protocol. All product or in-process specifications can be classified within five specification "codes."

1. Match/No Match (Yes/ No)
2. No More Than (NMT; i.e., ≤ 1.0%)
3. Less Than (i.e., < 1%)
4. No Less Than (NLT; i.e., ≥ 85%)
5. Range (i.e., 80 – 120 units/mL)

Specification code no. 1 (Match/No Match) will require validation as per ICH category I. Specification code no. 2 (≤ 1.0%) will require ICH category II validation, because results are numerically reported (quantitated). Code no. 3 requires ICH category III, since results are reported as "less than" (< 1%). Codes no. 4 and 5 (≥ 85% and 80 – 120 units/mL) require validation per ICH category IV. The relevant required validation parameters (i.e., Accuracy) for each product specification code and ICH category are listed in *Figure 1*.

Three out of five specification codes (nos. 2, 4, and 5) require numerical (quantitative) results. Those are graphically illustrated in *Figure 2*. In this figure, product specifications are related to ICH Q2B and method capability expectations. All factors should be evaluated and integrated to derive acceptance criteria. Product specifications for qualitative assays are generally "coded" as Match/No Match (or pass/fail, present/absent, etc.), and should be qualified or validated on a case-by-case basis. Many microbiological assays have abnormal (non-gaussian) data distributions (usually well-described by Poisson statistics), and are more difficult to generally classify for validation.

Figure 1

Required Validation Parameters for ICH Assay Categories and Specification Codes

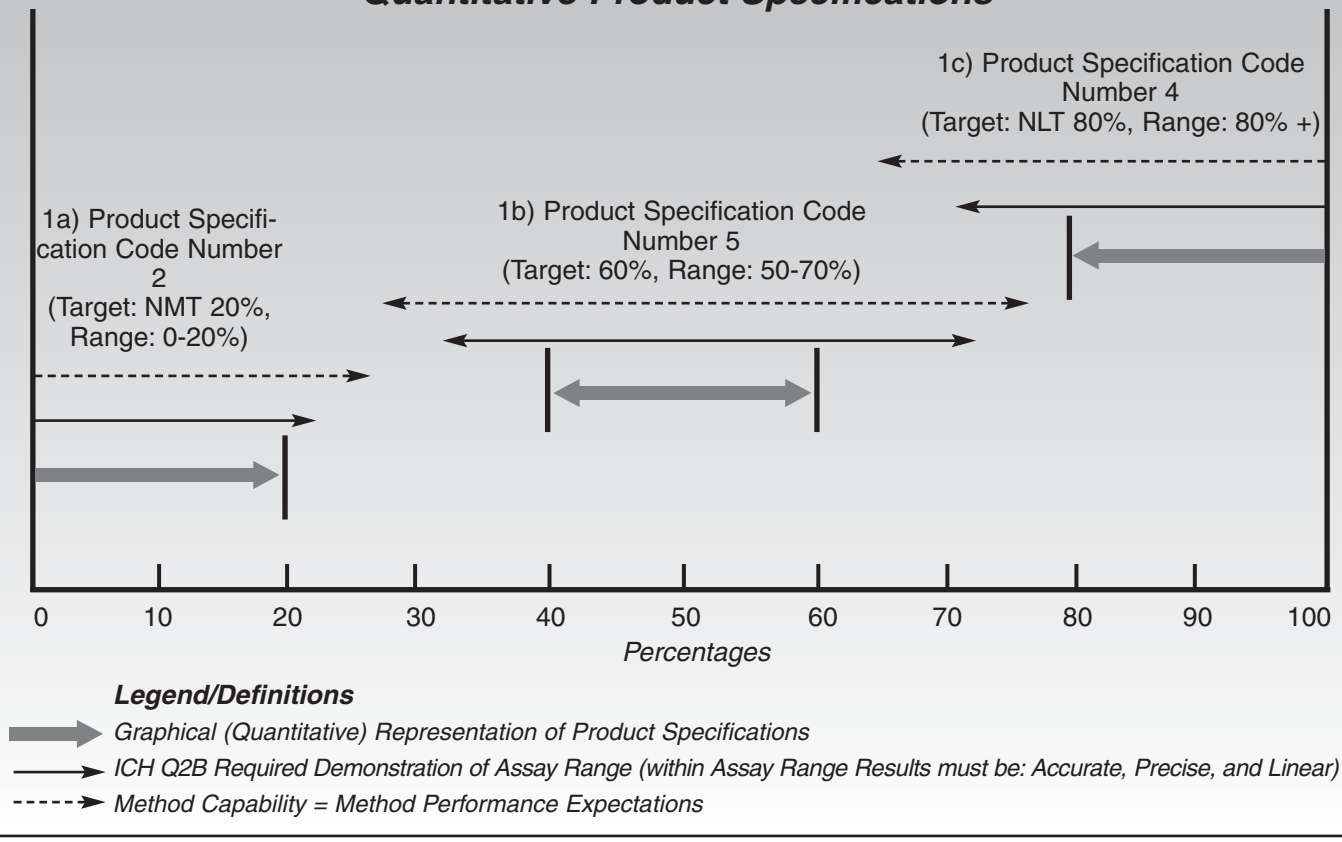
	Required Validation Parameters for ICH Category (I – IV)			
Specification Code	**1**	**2**	**3**	**4 and 5**
ICH Category	**I**	**II**	**III**	**IV**
Accuracy	No	Yes	No	Yes
Repeatability Precision	No	Yes	No	Yes
Intermediate Precision	No	Yes	No	Yes
Specificity	Yes	Yes	Yes	Yes
Linearity	No	Yes	No	Yes
Assay Range	No	Yes	No	Yes
Limit of Detection	No	No	Yes	No
Limit of Quantitation	No	Yes	No	No

ICH Validation Parameters

When an AMV protocol is generated, the assay category must be selected first. Then, the scientific approach to demonstrate assay suitability for each required validation parameter must be described in detail in the protocol. General guidance and considerations are described for each validation parameter. These should be followed when acceptance criteria are derived. Additional information can be found in the specific example used in this article.

Accuracy is usually demonstrated by spiking an accepted reference standard into the product matrix. Percent recovery (observed/expected x 100%) should ideally be demonstrated over the entire assay range by using multiple data points for each selected analyte concentration. In practice, the demonstration of accuracy is mostly affected by how well systematic errors can be controlled. When deriving acceptance criteria, one must keep in mind that in addition to ideal accu-

Figure 2



racy expectations (assuming expected equals true), potential systematic error (i.e., different response factor of spiked reference material) must be evaluated and factored into the acceptance criteria, unless the AMV protocol permits 'normalization,' if required. To keep systematic error at a minimum, common scientific sense should be used when describing spike sample preparation in the protocol (i.e., large volumes for spiked stock solutions, only calibrated equipment).

Many quantitative assays have ranges for their product specifications (code no. 5). The midpoint of this range is the target concentration that was either set historically from testing results, or as a manufacturing process target. When deriving acceptance criteria, one should consider that test system suitability must be demonstrated for this target range, which is exactly half of the specification range (target range = target concentration ± 0.5 x specification range). During QC routine testing, the test system must be capable to readily meeting this target range, and must be demonstrated in the AMV. It must therefore be demonstrated that the combined effects of lack of accuracy and reliability (precision) within the assay range can routinely be limited in order to support results within and outside product specifications (OOS). In other words, the acceptance criteria for accuracy and precision, combined within the assay range, should not be wider than half of the product specifications range, (at maximum) because one would otherwise fail to demonstrate test system suitability for this product. Intermediate precision should ideally be used here, since all routine testing samples could be tested by any trained operator on any qualified instrument on any given day. Repeatability precision (less variability) simply would not reflect this overall assay variability. The derivation of acceptance criteria for the remaining quantitative assays (code nos. 2 and 4) should also be dealt with in a similar matter.

Given what was mentioned above, there are several ways to derive acceptance criteria for accuracy. One way is: intermediate precision acceptance criteria could be derived first from historical data (Analytical Method Development [AMD] or QC testing). The numerical limits for intermediate precision are then

subtracted from the target range, and the remaining difference will set the maximum permissible acceptance criteria range for accuracy. This is illustrated in the AMV acceptance criteria example (*Figure 6*).

It may be more advisable not to use statistical approaches to demonstrate accuracy, such as t-statistics (comparing means of observed versus expected percent recoveries of various spike concentrations). The reason is that a potential systematic error is not accounted for in the expected recovery (mean = 100%, variance = 0). The expected recovery will then be compared to the observed recovery (mean ≠100%, variance ≠ 0), so that a statistical difference (i.e., t-test at 95% confidence) is likely to occur, although this difference may not be significant when compared to a numerical limit (percent or units). It may therefore be more practical to give numerical limits for accurate acceptance criteria.

Data generated for accuracy may be used to cover required data for other validation parameters, such as, repeatability precision, linearity, assay range, and Limit of Quantitation (LOQ).

Repeatability Precision indicates how precise the test results are under ideal conditions (same sample, operator, instrument, and day). Repeatability precision should be demonstrated over the entire assay range, just like accuracy and data generated for this parameter may be used. This has the advantage that fewer samples will have to be run. Even more important, when acceptance criteria are derived and connected, only one data set will be used, therefore, decreasing potential random error introduced by multiple sample preparations. The demonstration of repeatability precision is mostly affected by how well random errors in sample preparation can be controlled. Random experimental errors can only be controlled to some degree, since the Standard Operating Procedure (SOP) and AMV protocol should be followed as written by operators routinely generating QC testing results.

When using AMD data, the actual generation conditions of this data must be evaluated and put into perspective to set AMV acceptance criteria. When using QC routine testing data, data for the assay control can be summarized and used as a worse-case scenario for the AMV protocol. The Standard Deviation (SD) of this historical data can be expressed as confidence limits (i.e., 95% confidence ≅ 2 x SD), units, or percent (coefficient of variation, CV = SD/Mean x 100%), and should be used as the absolute limit for the AMV data, since historical data (several operators, instruments, days) should have less precision (greater CV) than AMV data.

Intermediate Precision should be demonstrated by generating a sufficiently large data set that includes replicate measurements of 100% product (analyte) concentration. This data should ideally be generated by three operators on each of three days, on each of three instruments. Different analyte concentrations to demonstrate intermediate precision over the entire assay range could be used, but results must be converted to percent recoveries before those can be compared. A data matrix where the total amount of samples can be limited, but differences among or between variability factors, such as operators and days, can still be differentiated, is illustrated in *Figure 3*.

The complete data set should then be statistically evaluated by an Analysis of Variance (ANOVA), where results are grouped by each operator, day, and instrument, but analyzed in one large table. Acceptance criteria state no significant difference at 95% confidence ($p > 0.05$) of data sets evaluated by ANOVA. It is advisable to include a numerical limit (or percentage) because the likelihood of observing statistical differences increases with the precision of the test method. In addition, some differences among various instruments, operator performances, and days (i.e., sample stability or different sample preparations for each day) are normal. The overall intermediate precision allowed should be relative to the expected accuracy, and must be within the combined limits for accuracy and intermediate pre-

Figure 3

Intermediate Precision Sample Matrix

Sample	Day Number	Operator Number	Instrument Number
3x 100% Conc.	1	1	1
3x 100% Conc.	1	2	2
3x 100% Conc.	1	3	3
3x 100% Conc.	2	1	2
3x 100% Conc.	2	2	3
3x 100% Conc.	2	3	1
3x 100% Conc.	3	1	3
3x 100% Conc.	3	2	1
3x 100% Conc.	3	3	2

cision. Additional F-tests and T-tests should be performed if overall p-value is less than 0.05 to evaluate the differences among factors and within factors. More detail will be provided in Part III of GAMVP: Data Analysis and the AMV Report.

Specificity of an assay is usually ensured by demonstrating none or insignificant matrix and analyte interference. The matrix may interfere with assay results by increasing the background signal (noise). Or, matrix components may bind to the analyte of interest, therefore potentially decreasing the assay signal. Spiking of the analyte of interest into the product (liquid), and comparing the net assay response increase versus the same spike in a neutral liquid (i.e., water or buffer), provides information on potential matrix interference. Reasonable acceptance criteria are: No observed statistical difference (t-test at 95% confidence) between assay responses of spiked samples of product matrix, versus those of buffer matrix. If the assay precision is relatively high, it is advisable to also include a numerical limit, in case $p < 0.05$, which should be similar to the limit stated under the validation parameter repeatability precision. This has the advantage that in case a statistical difference is observed, a reasonably derived numerical limit should be able to compensate for differences in sample preparation.

Other analytes potentially present in the product matrix should be spiked in proportional concentrations into the product matrix (keeping final analyte concentrations constant). Results of unspiked versus spiked product should also be compared by a t-test, and the acceptance criteria should be the same as those for matrix interference.

Linearity of the assay response demonstrates proportionality of assay results to analyte concentration. Data from accuracy may be used to evaluate this parameter. Linearity should be evaluated through a linear regression analysis, plotting individual results of either analyte concentration versus assay results, or observed versus expected results. The later approach should ideally yield a linear regression line slope of one (1). A slope smaller than one indicates a decreasing assay response with increasing analyte concentrations and vice versa. A y-intercept significantly greater or less than 0 with a slope of one, suggests a systematic error (i.e., sample preparation or spiked sample response factor $\neq 1$). A correlation coefficient less than one may reflect a lack of linearity, inaccuracy, imprecision, or all of the above. ICH Q2B requires reporting the regression line y-intercept, slope, correlation coefficient, and the residual sum of squares. Only acceptance criteria for the slope and the correlation coefficient need to be reported for linearity. Deriving these from accuracy and precision expectations is rather complex, and may not be practical. Depending on the sample preparation and the method capabilities for accuracy and precision, reasonable acceptance criteria should be stated in the AMV protocol. Reasonable criteria are: $r \geq 0.98$ (98% curve fit) and the 95% confidence interval of the regression line slope should contain 1.

The *Assay Range* of a method must bracket the product specifications. By definition, the LOQ constitutes the lowest point of the assay range, and is the lowest analyte concentration that can be quantitated with accuracy and precision. In addition to the required accuracy and precision for all analyte concentration points within the assay range, the assay response must also be linear, as indicated by the regression line coefficient. Data for the assay range may be generated in the AMV protocol under accuracy. Again, the advantages are a limited sample size to be run and evaluated, and the ability to evaluate this and other validation parameters from one set of prepared samples. Acceptance criteria for assay range should therefore be identical to those of accuracy, repeatability precision, and linearity.

Limit of Detection (LOD) of an analyte may be described as that concentration giving a signal significantly different from the blank or background signal. ICH Q2B suggests three different approaches to determine the LOD. Other approaches may also be acceptable when these can be justified. Per ICH, the LOD may be determined by visual inspection (A), signal-to-noise ratio (B), or the SD of the response and the slope (C).

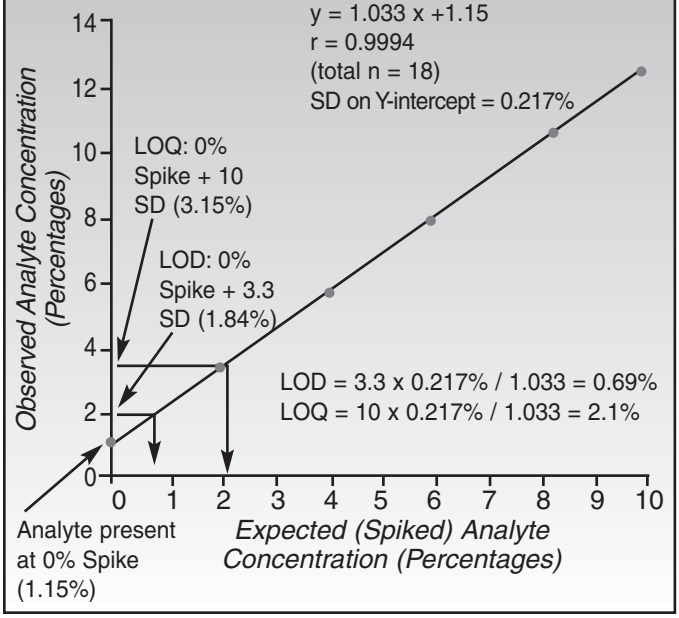
Visual inspection should only be used for qualitative assays where no numerical results are reported. The signal-to-noise approach (B) may be used whenever analyte-free product matrix is available. The analyte should then be spiked at low concentrations in small increasing increments into the product matrix. The LOD is then determined as the signal-to-noise ratio that falls between 2:1 and 3:1. This is the simplest and most straightforward quantitative approach.

Acceptance criteria derived for approach B should be similar to those based on repeatability precision. Criteria could be, for a desired signal-to-noise ratio of 3:1, three times the SD of repeatability precision.

Approach C uses the following formula: LOD = 3.3 s/m , where s is the SD of the response, and m is the slope of the calibration or spiked-product regression line. An estimate of the LOD is then obtained by the principle of the method of standard additions. This is graphically represented in *Figure 4*. If an assay simultaneously quantitates the active product and the impurity, data generated in the accuracy section and evaluated in linearity may be used to estimate the LOD using the regression line approach. Sufficient low analyte (impurity) concentrations must be included in the initial data set for accuracy to evaluate the LOD from one sample preparation set. The LOD acceptance criteria for approach C should be identical to those based on repeatability precision if the identical data set was used. When linearity data is evaluated by regression analysis, the LOD must not exceed the repeatability precision criteria when the predicted SD regression line y-intercept is multiplied by 3.3, and divided by the regression line slope (slope ≅ 1).

For approach A, B, or C, and any other justified approaches, the LOD acceptance criteria must be significantly lower than the product specifications and the LOQ. Selecting and justifying a particular approach should be done with a knowledge of method capabilities, in particular the level of precision. One cannot expect to determine a relatively low LOD, as the variance within low analyte concentrations is relatively high.

Figure 4



Limit of Quantitation (LOQ) is by definition the lowest analyte concentration that can be quantitated with accuracy and precision. Since the LOQ constitutes the beginning of the assay range, the assay range criteria for linearity must be passed for the particular analyte concentration determined to be the LOQ. The determination of the LOQ involves the same approaches (A, B, and C) as those for LOD. The only difference is the extension of the required signal-to-noise ratio to 10:1 (approach B), or the change in the formula (approach C) to: LOQ = 10 s/m. The acceptance criteria for LOQ should therefore be set proportionally similar to those indicated for LOD. In addition, the LOQ acceptance criteria should contain the same limits for accuracy, repeatability precision, and linearity, as set for each of these validation parameters.

Two reasons of caution should be considered when following ICH approach C. One, the determination of particular analyte concentrations for LOD and LOQ are independent of sample size, but sample size should be ≥ 6. Individual results plotted for each analyte concentration tested (instead of averages) generally yield higher SDs, and therefore higher LODs and LOQs. Two, approach C only delivers acceptable LODs and LOQs when the assay response is highly linear, precise, and accurate over the plotted concentration range. In addition, the spiked sample preparation must be accurately performed to prevent further random deviations from the regression line. If any of these raised issues may be a real concern, a different justified approach should be chosen.

Robustness should be addressed during method development. The main reason is that a method and its governing SOP are not to be changed for routine testing and the validation of that SOP. The SOP controls operational limits within the overall system suitability criteria that are set during AMD. Deliberate small changes to the test system should be done during development, because significant differences in the AMV results may not be easily explained in the AMV report.

System Suitability should be demonstrated by showing

that a complete test system is capable of delivering accurate and reliable results over time when used under routine QC testing conditions. All materials to be tested or used in testing should be stable in excess of the duration of the test procedure. Appropriate reference material (standards and/or controls) should be used to establish and control system suitability. Standards and controls should have reasonable acceptance limits properly derived from historical data. These limits should be regularly monitored and adjusted to account for minor changes, such as those potentially expected from switching reagents.

Overall test system suitability is generally demonstrated by passing the acceptance criteria of all AMV parameters evaluated. During the AMV execution, all invalids, repeats, and OOS results generated should be evaluated in the AMV report. More detail will be provided in Part III of GAMVP.

AMV Acceptance Criteria Example

Once it has been decided that a test method must be validated, as per standard practice instructions (see also GAMVP, Part I, November 2002 issue, *Journal of Validation Technology*), a successful AMV approach should be thoroughly planned. Provided below is an example how to select the appropriate assay categories (therefore the required validation parameters), develop and describe a validation strategy, and systematically derive reasonable acceptance criteria for the AMV protocol.

Hypothetical Scenario

The formulation of a therapeutic protein will be changed (minor) at a late stage of the purification process. Several final container test methods require complete revalidations (current method) or validations (new method), while some will require only partial revalidations, depending on the formulation change impact on each test method. It was decided that the purity test requires a complete revalidation. Quantitative Capillary Zone Electrophoresis (CZE) is used to simultaneously provide results for the active protein and the impurities present in low, but reported concentrations. All protein components present are quantitated as Relative Percent Area (RPA) out of all components present (100%). Final container product specifications are NLT 90% for active protein, NMT 5% of protein impurity A, NMT 10% of protein impurity B.

Approach

The CZE test method must be validated for content/potency (major component) and for quantitation of impurities. From the information listed in *Figure 1*, the CZE test method must be validated simultaneously for ICH category I and II. The required validation parameters are accuracy, repeatability precision, intermediate precision, specificity, linearity, assay range, LOD, and LOQ.

The next step is to analyze product specifications, and compare those to the historical assay performance. In general, the historical assay performance can be evaluated from AMD data, previous validation data, historical product final container QC testing data, and historical assay control data. Since we are revalidating this CZE test procedure without having changed test method system parameters besides our minor product reformulation, there is no need to evaluate AMD and previous validation data. Assuming that there were no recent minor changes (i.e., change in reagent manufacturer) that could have shifted historical results for the assay control (and product), historical QC data for final containers of product, and the assay control of the last several months (n ≥30) should be evaluated. Historical product results will contain lot-to-lot variation due to an expected lack of complete product uniformity. These results are therefore expected to have a greater variation than those of the assay control. The historical QC testing data for the control and product are listed in *Figure 5*.

Figure 5

Historical Testing Data for the Assay Control and Product Over the Last Six Months

Sample/Statistic	Percent Purity		Percent Impurity A		Percent Impurity B	
Sample	**Prod.**	**Cont.**	**Prod.**	**Cont.**	**Prod.**	**Cont.**
Product Specifications	90%		5%		10%	
n	90	90	90	90	90	90
Mean (in percentages)	94.1	91.4	2.0	2.8	3.9	5.8
Standard Deviation (in percentages)	1.32	1.14	0.43	0.31	0.55	0.39
CV (in percentages)	1.41	1.25	28.6	11.1	13.8	6.72

KEY: *Prod. (Product)* *Cont. (Control)*

The data of *Figure 5* may then be used to generate the acceptance criteria for all required validation parameters. *Figure 6* lists each validation parameter with the relevant AMV design, brief sample preparation, reported results, acceptance criteria, and a rationale for acceptance criteria for those areas.

The Validation Protocol

The AMV protocol may consist of sections listed in *Figure 7*. In general, the protocol should have sufficient detail to be executed by the operators routinely performing the test procedure to be validated. The SOP (or draft version) must be followed as written, unless specified and reasoned in the protocol. This is important because the SOP, which includes sample preparation and instructions as to how results are generated and reported, should be validated as a complete test system.

Following a signature page and a list of content sections, reasons, and scope of the AMV, as well as previous or supporting validations, should be mentioned in the introduction section. A brief description of the principle of the test methodology should be given in the principle section. Materials, equipment, and instrumentation to be used must be described in detail, including Certificates of Analysis (CoA) for all reference materials, instrument ID numbers, and all products or in-process material to be tested. Historical assay performance should be summarized from analytical method development data (new method) or routine testing results (revalidation), and integrated into the acceptance criteria. The selected assay classification (i.e., category IV assay validation to be used for the quantitation of the main drug component) should be clearly given in the beginning of the section on validation parameters and design. The validation approach used to demonstrate system suitability for each validation parameter should be described and justified, and reported results and their acceptance criteria should be provided. In addition, detailed instructions for sample preparation, AMV execution, and validation result generation should be included. A section, data analysis, should indicate which software (validated) should be used to statistically evaluate results versus acceptance criteria.

A table (validation execution matrix) should be included in the protocol, listing which validation parameter will be executed by which operator, on which day, and on which instrument. This table will demonstrate to the reader of this document that the proposed validation is well-planned, and should furthermore prevent execution deviations by the operators. A validation execution matrix example is given in *Figure 8*.

A list of references to the governing Standard Practice (SP) and supporting documents ensures the reader that all relevant procedures are followed, and that relevant supporting documents (CoA, product specifications, historical data, and supporting reports) were consulted. All supporting documents should be attached (list of attachments) and filed with the protocol. A final section, AMV matrix and acceptance criteria, in which the reader can refer to a table where each validation parameter's validation approach, reported results, and acceptance criteria are summarized, will be helpful. Information can be copied from the validation parameter section.

Significant Digits of Reported Results

Final container and in-process product specifications should report test results with the appropriate number of significant digits. AMVs should generate this number by consistently following a designated SP. Test results must be reported reflecting the uncertainty in these results. This uncertainty can be expressed by using the appropriate number of significant digits based on assay precision. How exactly this is to be done, depends on definitions and instructions within the SP(s). One relatively simple way of dealing with this issue is to use a widely accepted SP, such as E 29-02, published by the American Society for Testing and Materials (ASTM E 29-02).[1] This practice gives clear instructions on how to generate significant digits from repeatability precision, which is required of quantitative AMVs, as per ICH, USP, and FDA guidelines. The reason that AMVs should deliver the appropriate reported uncertainty for test results lies mostly in the fact that by the time an AMV is executed, at a minimum, a draft version of the SOP is already in place on which QC operators have been trained. Following this ASTM E 29-02 practice, in which the definition for repeatability precision matches those of ICH, USP, and FDA, provides the advantage of having reference to an accepted federal document.

Figure 6

Summary of Overall AMV Design and Acceptance Criteria

Validation Parameter	AMV Design	Sample Preparation	Minimum ICH Q2B Requirements	Reported Results	Acceptance Criteria	Rationale for Acceptance Criteria
Pre-requirement (1)	Identification and purity of commercially purchased protein impurity A and B must be determined using complimentary tests (other methods such as SDS-PAGE, HLPC, HPSEC, MS, Western Blot). Run in triplicates.	Follow corresponding SOPs for other tests.	N/A	Mean purity (n=3) in %, identification (n=3): Yes/no	Identification of commercially purchased proteins must match impurity protein A and B, respectively.	We cannot expect 100% purity of commercial proteins. Less than 100% purity can be normalized for percent recovery calculations. Identification(s) must match because response factors for impurity protein A and B (Accuracy) can otherwise not be validated.
Pre-requirement (2)	Potential response factor differences for protein impurity A and B must be determined. Differences in purity and/or response factors must be 'normalized' for percent recovery calculations. Run in triplicates.	Follow SOP for CZE. Ideally, protein impurity A and B should be tested individually at product specification concentration, and final container product lot (A) should be tested at 100%.	N/A	Mean area counts for each of impurity A and B. Response factors.	None	(Caution must be exerted here because we are currently using the CZE test (validated for final product release testing).
Accuracy	Percent recoveries of commercially purchased reference material for protein impurity A and B will be determined from increasing spike concentrations by using Relative Percent Area (RPA). RPAs for each protein impurity and corresponding therapeutic protein will be determined using individual response factors (if required). All spike concentrations will be run in triplicates by Operator 1 on Day 1 using Instrument 1. Percent Recovery = (Observed RPA/Expected RPA) x 100%.	Spike commercially purchased protein impurity A and B each into reformulated final container product (lot A) with increasing concentrations (0.0, 0.5, 1.0, 2.0, 5.0, 10.0, 15.0, 20.0 %) keeping final protein concentration constant.	Data: three replicates over three concentrations covering the Assay Range.	Mean percent recoveries (n=3) for each spiked concentration (n=7) for impurity A, impurity B, and the corresponding percent recoveries for the therapeutic protein will be tabulated.	Mean spike recoveries for impurity A and impurity B for each spike concentration (n=7) must fall within 100±40% and 100+ -20%, respectively. Each corresponding mean spike recovery (n=2x7) for the therapeutic protein must fall within 98-102%.	The combination (worst-case scenario) of assigned limits for Intermediate Precision and Accuracy must be no greater than the difference between historical mean product results (n=3, see Table 3) and their corresponding product specifications (n=3). A worst-case limit of historically recorded 2 SDs (of assay control, see Intermediate Precision) has been assigned to Intermediate Precision. This limit is then subtracted from the product specifications, and constitutes the maximum value for the acceptance criteria for Accuracy. An example for the therapeutic protein

Continued

Figure 6 *(Continued)*

Summary of Overall AMV Design and Acceptance Criteria

Validation Parameter	AMV Design	Sample Preparation	Minimum ICH Q2B Requirements	Reported Results	Acceptance Criteria	Rationale for Acceptance Criteria
Accuracy						recovery is given here: {[(94.1% - 90.0%) - (2 x 1.14%)] / 90.0%} x 100% = 2.02%. Therefore, percent recovery = 100±2%.
Repeatability Precision	Data will be generated in Accuracy to demonstrate precision over the entire Assay Range. In addition, Operator 1 on Day 1 using Instrument 1 will generate n=15 data points using one final product container lot. This extensive data set for Repeatability Precision will be used to generate the appropriate number of significant digits to be reported for test results.	Follow SOP for CZE and test one final product container lot (A) at 100%.	Data: Nine determinations over Assay Range (e.g., three replicates over three concentrations). six determinations at 100% test concentration. Report: Standard Deviation (SD), Coefficient of Variation (CV), Confidence Interval (CI).	From Accuracy data: CVs (in %), means (n=3), SDs, CIs (p=0.05) for means, for % therapeutic protein, protein impurity A, and protein impurity B. From Repeatability data: CV (in %), mean (n=15), SD, CI (p=0.05) for mean, for % therapeutic protein, protein impurity A, and protein impurity B.	Mean CVs (n=8) from Accuracy data must be within the following limits (in RPA): % therapeutic protein: NMT 2.5, % impurity A: NMT 22. % impurity B: NMT 13. CVs (n=3) from 15 data points must be within the following limits (in RPA): % therapeutic protein: NMT 1.3, % impurity A: NMT 11. % impurity B: NMT 6.7.	CVs may differ over the studied assay range, and we have very limited data points (n=3) for each test concentration. Therefore, we must keep mean CVs as wide as possible to avoid failing acceptance criteria. CVs from samples at 100% test concentrations (n=15 data points) shall be no greater than those of the historical assay control because these data were generated over six months by different operators on different instruments.
Intermediate Precision	One unspiked final product container lot (A) will be tested in triplicates on each of three days by each of three operators on each of three instruments. Intermediate Precision will be determined for each purity and integrity characteristic by using an Analysis of Variance (ANOVA). Any statistical differences (at the 95% confidence level) between and within factors (operators, days, instruments) will	Follow SOP for CZE and test one final product container lot (A) at 100%.	Data/Report: No specific requirements. Variations (factors) to be studied (in a matrix) are days, operators, and equipment.	Overall and individual P-values of factors (operators etc.) from ANOVA. Overall and factor CV(s) and SD(s) for % therapeutic protein, protein impurity A, and protein impurity B.	P-value of ANOVA must be NLT 0.05. If $p < 0.05$, additional F-tests and T-tests will be performed to isolate factors with statistically different means and/or variations. An investigation must demonstrate that each different factor mean (at p=0.05) will not affect assay performance and overall system suitability.	The means and precision variabilities among and between factors should not be statistically different at 95% confidence. The likelihood of observing statistical difference(s) increases with assay precision, and may not impact system suitability. It is therefore advisable to set an "escape clause" by generating numerical limits for overall CV (2 SDs of assay control) and factor CVs (1 SD of assay control) from the historical data. It is

Continued

Figure 6 *(Continued)*

Summary of Overall AMV Design and Acceptance Criteria

Validation Parameter	AMV Design	Sample Preparation	Minimum ICH Q2B Requirements	Reported Results	Acceptance Criteria	Rationale for Acceptance Criteria
Intermediate Precision	be investigated. (A matrix for Intermediate Precision is illustrated in Table 2 of this article)				Overall CV must comply with the following limits: % therapeutic protein (in RPA): NMT 2.5, % impurity A: NMT 22. % impurity B: NMT 13. Factor CVs must comply with the following limits: % therapeutic protein (in RPA): NMT 1.3, % impurity A: NMT 11.% impurity B: NMT 6.7.	more meaningful to use the historical assay control data (see Table 3) here because product data includes normal variation among different product lots.
Specificity	Matrix interference: Matrix interference will be evaluated by comparing results for each impurity-spiked (A and B) sample, spiked into final product container (lot A), to those of spiked assay control, and spiked current final product (lot B). Percent recoveries will be compared by ANOVA and, if required, by t-tests to evaluate potential differences between product lot (lot A), the assay control, and current final product (lot B). One operator will run all samples on one day on one instrument. The following samples will be prepared: Three spiked sample preparations of each impurity (n=2) for each sample	Matrix interference: All samples (constant final concentrations) will each be spiked with 5% of protein impurity A and B.	No specific requirements.	Individual and mean (n=3) RPAs and corresponding percent recoveries for spiked samples (n=6) will be reported. An ANOVA table will be presented.	No statistical significant difference (at 95% confidence level) shall be obtained ($p > 0.05$) in ANOVA. If $p < 0.05$, additional F-tests and t-tests will be performed to isolate spiked samples with statistically different means and/or variations. An investigation must demonstrate that each different factor mean (at $p=0.05$) will not affect assay performance and overall system suitability. The difference(s) among spiked	The means and precision variabilities among and between factors should not be statistically different at 95% confidence. Similar to Intermediate Precision, the likelihood of observing statistical difference(s) increases with assay precision, and may not impact system suitability. In addition, we should account for potential differences in results due to sample preparations. It is therefore advisable to set an "escape clause" by generating numerical limits for difference limit (1 SD of assay control) from the historical data. It is more meaningful to use the historical assay control data (see Table 3) here

Continued

Figure 6 *(Continued)*

Summary of Overall AMV Design and Acceptance Criteria

Validation Parameter	AMV Design	Sample Preparation	Minimum ICH Q2B Requirements	Reported Results	Acceptance Criteria	Rationale for Acceptance Criteria
Specificity	(n=3). All samples will be run three times (total runs: n=3x2x3x3=54). Analyte interference: Analyte interference can be inferred from the matrix interference studies.				matrices (lots A and B, and assay control) for each spiked impurity (n=2), must be no greater than the following limits (in RPA): NMT 1.3, % impurity A: NMT 11. % impurity B: NMT 6.7.	because product data includes normal variation among different product lots. A reason of caution is how well differences in sample preparation can be controlled.
Linearity	Linearity will be determined at the low percentage range (approx. 0-20 RPA) to cover a potential impurity range (NMT 5% impurity A; NMT 10% impurity B), and at the high percentage range (approx. 75 to 95 RPA) to cover the product specifications for the therapeutic protein (NLT 90 %). Three regression lines will then be generated, one each for the two low (impurity A and B), and one for the high (therapeutic protein) percentage ranges. Individual RPA results (n=3 for each spiked concentration) for each spiked concentration (0.0, 0.5, 1.0, 2.0, 5.0, 10.0, 15.0, 20.0%) will be plotted against actual spike concentrations (in RPA) present.	See Accuracy.	Correlation coefficient(s), y-intercept(s), slope(s) of regression line(s), and Residual Sum(s) of Squares (RSS) should be reported. A plot of the data (regression line) to be provided. NLT 5 concentrations to be tested.	Regression line slopes, intercepts, correlation coefficients, RSS for each regression line. Plots (n=3) of the regression lines of individual RPA results (n=3 for each spiked concentration) for each spiked concentration (0.0, 0.5, 1.0, 2.0, 5.0, 10.0, 15.0, 20.0%) versus actual spike concentrations (in RPA) present will be provided.	Correlation coefficient ≥ 0.98 for each of three regression lines. All three CIs (at 95% confidence) for each regression line slope must contain 1.	Because lack of Accuracy, Repeatability Precision, and differences in sample preparation(s) may contribute to a decrease in regression line fit (lower correlation coefficient), a generally acceptable correlation coefficient (≥ 0.98) should be used here. The confidence limits of the slope should contain 1 since otherwise assay response may not be sufficiently proportional to support quantitative results over the entire assay range.
Assay Range	Assay Range will be determined at the low percentage range (approx. 0-20 RPA) to	See Accuracy.	For therapeutic protein: 80 to 120% of test concentration.	Regression line slopes, intercepts, correlation coefficients, RSS	Correlation coefficients for each of three regression lines. All three	All results generated within the determined Assay Range must be accurate and precise. The

Continued

Figure 6 *(Continued)*

Summary of Overall AMV Design and Acceptance Criteria

Validation Parameter	AMV Design	Sample Preparation	Minimum ICH Q2B Requirements	Reported Results	Acceptance Criteria	Rationale for Acceptance Criteria
Assay Range	cover a potential impurity range (NMT 5% impurity A; NMT 10% impurity B), and at the high percentage range (approx. 75 to 95%) to cover the product specifications for the therapeutic protein (NLT 90 %). For details, see Linearity section.		For impurity A and B: From reporting level to 120% of specification	for each regression line will be reported. All coefficients of variation (CV) for RPA for each spiked concentration will be reported. An overall CV for each of the three spiked samples series (impurity A, B, and therapeutic protein) will be reported.	regression line slope CIs (95% confidence) must contain 1. All acceptance criteria for Accuracy, Repeatability Precision, and Linearity must be passed.	assay response within the Assay Range must be linear. For further details, see sections Accuracy, Repeatability Precision, and Linearity.
Limit of Detection	The LOD will be determined for each impurity (A and B) concentration from data generated in the Accuracy section and evaluated in the Linearity section. For details, refer to the Linearity section. Since final product container lot (A) may contain significant levels of each of impurity A and B (> 1%), the LOD will be determined from the regression lines generated for impurity A and B in the Linearity section as per section VII.C.1 of ICH Guidance to Industry document Q2B. $LOD = (3.3 \times \sigma) / S$ The slopes (S) will be determined from the linear regression data for each impurity (A and B). The standard deviation (σ) of the response will be determined from	See Accuracy and Repeatability Precision.	Approach C (see section LOD of this article): $LOD = (3.3 \times \sigma) / S$, where σ = SD of response and S = regression line slope.	All concentrations and results (in RPA) will be tabulated. The apparent LODs (in RPA) for each impurity (n=2) will be reported.	The LODs for impurity A and B must be NMT 0.4% and 0.9%, respectively.	In general, this ICH recommended approach to determine LOD may yield relatively high values for LOD (and LOQ) versus some alternative approaches. The level of Accuracy, Repeatability Precision, and Linearity in results generated by this test system will be reflected in the LOD (and LOQ). The LOD should be less (33%) than the LOQ, which in turn must be significantly less than the historical product impurity means. See also LOQ.

Continued

Figure 6 *(Continued)*

Summary of Overall AMV Design and Acceptance Criteria

Validation Parameter	AMV Design	Sample Preparation	Minimum ICH Q2B Requirements	Reported Results	Acceptance Criteria	Rationale for Acceptance Criteria
Limit of Detection	the RPA results for each impurity (A and B) in the Repeatability Precision section.					
Limit of Quantitation	The LOQ will be determined for each impurity (A and B) concentration from data generated in the Accuracy section, and evaluated in the Linearity section. For details, refer to the Linearity section. Since final product container lot (A) may contain significant levels each of impurity A and B (> 1%), the LOQ will be determined from the regression lines generated for impurity A and B in the Linearity section, as per section VIII.C.1 of ICH Guidance to Industry document Q2B. LOQ = (10 x σ) / S The slopes (S) will be determined from the linear regression data for each impurity (A and B). The standard deviation (σ) of the response will be determined from the RPA results for each impurity (A and B) in the Repeatability Precision section.	See Accuracy and Repeatability Precision.	Approach C (see section LOQ of this article): LOQ = (10 x σ) / S, where (= SD of response and S = regression line slope.	All concentrations and results (in RPA) will be tabulated. The apparent LOQs (in RPA) for each impurity (n=2) will be reported.	The LOQs for impurity A and B must be NMT 1.1% and 2.8%, respectively.	The LOQ should be significantly less than the historical mean impurity results (2.0% and 3.9% for impurity A and B, respectively, see Table 3). We can determine the LOQ (and therefore the LOD) by subtracting 2SDs for product impurity results from the historical mean impurity results (e.g., impurity A: 2.0% - 2 x 0.43% = 1.14%). See also rationale under LOD.
System Suitability	All current criteria for system suitability (per SOP) must be satisfied in order for each test to be considered valid. Each failing test will be	See all sections.	No specific requirements.	Number of valid and invalid tests. Appropriate number of significant digits to be used for final result reporting.	As per SOP. No acceptance criteria for number of invalids and appropriate number of significant digits.	System suitability will be demonstrated by passing all acceptance criteria. System suitability criteria of the SOP may change, depending on the number

Continued

Figure 6 *(Continued)*

Summary of Overall AMV Design and Acceptance Criteria

Validation Parameter	AMV Design	Sample Preparation	Minimum ICH Q2B Requirements	Reported Results	Acceptance Criteria	Rationale for Acceptance Criteria
System Suitability	repeated per SOP until the current criteria are met. System suitability will be evaluated by listing invalid tests. The appropriate number of significant digits in reported results will be determined following ASTM E-29-02.					of valids/invalids generated.

Acceptance Criteria System

When acceptance criteria are determined for each validation parameter, the fact that these are connected may often be overlooked. Each quantitative test system has certain capabilities to yield accurate, precise, and analyte-specific results over the desired assay range. Since every test system has certain limits on its capabilities, the acceptance criteria that ideally should define these limits should be connected. Test performance expectations should be reflected in an Acceptance Criteria System (ACS) where all acceptance criteria for the required validation parameters (as per assay classification) are meaningful, and will focus on permissible worst-case conditions.

Like most concepts, the ACS has several drawbacks. One, it takes time and experience to evaluate and integrate all assay performance expectations into one system for all validation parameters, especially when validation data will be generated under QC routine testing

Figure 7

Suggested AMV Protocol Sections

Section Number	Section Number	Subsections
N/A	Protocol Approval	Protocol Title; Signatures with Job Titles
N/A	List of Protocol Sections	Table of Content; List of Figures (if applicable); List of Tables
1	Introduction	N/A
2	Principle	N/A
3	Materials, Equipment, and Instrumentation	Materials; Equipment; Instrumentation
4	Historical Assay Performance	Historical Data for Assay Control; Historical Data for Samples (if available); Product Specifications
5	Validation Parameters and Design	Test Method Description (summarizes SOP); Validation Pre-Requirements (if applicable); Validation Parameters
6	Validation Execution Matrix	See Table 5
7	Data Analysis	Calculation Samples; Statistical Software
8	List of References	N/A
9	List of Attachments	N/A
10	AMV Matrix and Acceptance Criteria	Table with Column Headings: Validation Parameters, Validation Approach, Sample Preparation, Reported Results, Acceptance Criteria

Figure 8

Validation Execution Matrix

Validation Parameter	Op. Number	Day Number	Ins. Number	Run Number	Sample (Spike Conc.)
Accuracy	1	1	1	1	(3x): 5, 10, 20, 40, 60, 80, 100, 120%
Repeatability	1	1	1	1	As Accuracy
Int. Precision	1	2	1	2	3x 100% Conc.
Int. Precision	2	2	2	3	3x 100% Conc.
Int. Precision	3	2	3	4	3x 100% Conc.
Int. Precision	1	3	2	5	3x 100% Conc.
Int. Precision	2	3	3	6	3x 100% Conc.
Int. Precision	3	3	1	7	3x 100% Conc.
Int. Precision	1	4	3	8	3x 100% Conc.
Int. Precision	2	4	1	9	3x 100% Conc.
Int. Precision	3	4	2	10	3x 100% Conc.
Specificity	1	5	1	11	Matrix Interference
Specificity	1	5	1	12	Analyte Interference
Linearity	1	1	1	1	As Accuracy
Assay Range	1	1	1	1	As Accuracy
LOD	1	1	1	1	As Accuracy
LOQ	1	1	1	1	As Accuracy

conditions. Two, systematic errors introduced during sample preparation for spiking studies (initially, small errors could also be magnified at the end of a dilution series) to determine accuracy (percent recovery) may not be accounted for when the ACS is solely developed using historical data and method capabilities. Three, when one validation parameter will fail its acceptance criteria, in general, all validation parameters will fail, leading to potential complete failure to demonstrate test system suitability. On the other hand, the opposite must then also be true, meaning that all criteria within the complete ACS will be passed when one acceptance criterion will be passed.

Although ACS may only be a concept at this point, and may not be applicable for all AMVs, the potential advantages of a well-developed ACS should outweigh the drawbacks, because the ACS is solid as a system, and can easily be justified and defended. Each individual acceptance criterion is now meaningful, related to all others, and reflects the test system performance capabilities. The concept of ACS should be considered for accuracy, precision (repeatability and intermediate), assay range, LOQ, and specificity. However, deriving acceptance criteria for the linearity parameter will be difficult, since an estimation of the potential worst-case combination(s) of regression line slope, y-intercept, and regression coefficient becomes very complex.

With a well-developed ACS, the auditors can no longer criticize acceptance criteria. Acceptance criteria are now derived as part of the ACS, which in turn, demonstrates method capabilities in respect to product specifications, historical data, and method capabilities. Furthermore, the ACS is a dynamic system that can be readily adapted as a unit to changes to the system, or for other reasons for revalidation. With experience, it will become easier and faster to set up an ACS, even for the AMV of a new test method.

Conclusion

Deriving reasonable acceptance criteria requires experience and a deep understanding of the method capabilities, product specifications, and historical data. This article provides a detailed approach to derive these criteria, which can now be justified and easily defended in an audit. The AMV can now accurately demonstrate that the test system is suitable for its intended use. ❑

About the Author

Stephan O. Krause, Ph.D. is managing the QC Analytical Validation department within the Biological Products division of Bayer HealthCare Corporation. He received a doctorate degree in bioanalytical chemistry from the University of Southern California. Dr. Krause can be reached by phone 510-705-4191, and by e-mail at stephan.krause.b@bayer.com.

Acknowledgement

I would like to thank my colleague, Christopher Fisher, for his helpful comments and critical review of this article.

Reference

1. As per ASTM E 29-02 Section 7.4, the following instructions are given: "A suggested rule relates the significant digits of the test result to the precision of the measurement expressed as the standard deviation σ. The applicable standard deviation is the repeatability standard deviation (see Terminology E 456). Test results should be rounded to not greater than 0.5 σ or not less than 0.05 σ, provided that this value is not greater than the unit specified in the specification (see 6.2). When only an estimate, s, is available for σ, s, may be used in place of σ in the preceding sentence. Example: A test result is calculated as 1.45729. The standard deviation of the test method is estimated to be, 0.0052. Rounded to 1.457 since this rounding unit, 0.001, is between 0.05 σ = 0.00026 and 0.5 σ = 0.0026."

   For the rationale for deriving this rule, refer to ASTM E 29-02. For definitions refer to ASTM E 456.

Suggested Reading

1. Krause, S. O., "Good Analytical Method Validation Practice, Part I: Setting-Up for Compliance and Efficiency." *Journal of Validation Technology*. Vol. 9 No. 1. November, 2002. pp 23-32.
2. International Conference on Harmonization (ICH), Q2A, "Validation of Analytical Procedures." *Federal Register*. Vol. 60. 1995.
3. ICH, Q2B, "Validation of Analytical Procedures: Methodology." *Federal Register*. Vol. 62. 1996.
4. United States Pharmacopoeia. USP 25 <1225>. "Validation of Compendial Methods."
5. American Society for Testing and Materials (ASTM) E 29-02. "Standard Practice for Using Significant Digits in Test Data to Determine Conformance with Specifications." July, 2002.
6. ASTM E 456 – 96. "Standard Terminology for Relating to Quality and Statistics." September, 1996.
7. Miller, J. C. and Miller, J. N. "Statistics for Analytical Chemistry." (2nd ed.). Ellis Horwood Ltd., England. 1988.

Article Acronym Listing

ACS:	Acceptance Criteria System
AMD:	Analytical Method Development
ANOVA:	Analysis of Variance
AMV:	Analytical Method Validation
ASTM:	American Society for Testing and Materials
CI:	Confidence Interval
CoA:	Certificates of Analysis
CV:	Coefficient of Variation
CZE:	Capillary Zone Electrophoresis
FDA:	Food and Drug Administration
GAMVP:	Good Analytical Method Validation Practice
HPLC:	High Performance Liquid Chromatography
HPSC:	High Performance Size Exclusion Chromatography
ICH:	International Conference on Harmonization
LOD:	Limit of Detection
LOQ:	Limit of Quantitation
MS:	Mass Spectometry
NLT:	No Less Than
OOS:	Out-Of-Specification
QC:	Quality Control
RPA:	Relative Percent Area
RSS:	Residual Sum(s) of Squares
SD:	Standard Deviation
SDS-PAGE:	Sodium Dodecyl Sulphate-Polyacrylamide Gel Electrophoresis
SOP:	Standard Operating Procedure
SP:	Standard Practice
USP:	United States Pharmocopeia

Good Analytical Method Validation Practice

Setting Up for Compliance and Efficiency: Part I

By Stephan O. Krause, Ph.D.
Bayer HealthCare Corporation

In recent years, Analytical Method Validation (AMV) has become an increasing concern and focus of regulatory agencies. Agencies expect companies to continuously modify test methods along the long path from drug development to final release testing of the licensed product. The level of detail and validity within test method qualification and validation documents increases along this path. Although directed towards Quality Control (QC), departments of mid- to large-sized manufacturers for biological and pharmaceutical products, most of the suggestions presented here may also be applied to other industries. The scope of this article is to illustrate a rigorous AMV program, that when consistently applied, should maintain compliance with regulatory agencies in the future. This article will provide recommendations on how to incorporate compliance requirements into a practical and efficient approach to produce analytical validations.

"Once the FDA has audited AMVs and found general compliance, the Agency will certainly be more confident and trusting in the overall testing results."

Good Analytical Method Validation Practice (GAMVP) ensures accurate and reliable test results, and therefore the safety and quality of the product. How many validation resources a company decides to devote to the AMV program is a business decision like the purchase of insurance for the shipped licensed product. Companies can invest a minimum level of resources, and wait to be eventually hit by the U.S. Food and Drug Administration (FDA) with 483 observations and potential Warning Letters. On the other hand, companies may invest more resources into the AMV program and supporting departments, and this would then pay off later in time when FD-483 observations and warning letters could be avoided. Once the FDA has audited AMVs and found general compliance, the Agency will certainly be more confident and trusting in the overall testing results. This impression may carry on for years to come, and will furthermore ensure product safety to the patient.

Efforts towards a compliant and efficient AMV program include the generation of a method validation master plan, which aligns the timelines of the regulatory and manufacturing units with validation project completion. To be efficient, the AMV department must be well-organized, and staffed with expe-

rienced scientists who are sufficiently trained to consistently plan and execute validation projects. It is important to realize that the continuously increasing number of required validations must be counterbalanced also by an equivalent staffing increase of Quality Assurance (QA) personnel, delegated for protocol and report review and approval. In order to cope with increasing expectations, the AMV department must be integrated effectively within the QC unit. However, the set-up of the AMV program starts with the generation of the Standard Practice/Procedure (SP) document, which governs the process of AMVs. The SP clearly defines all required AMV procedures and responsibilities.

The AMV Standard Practice/Procedure

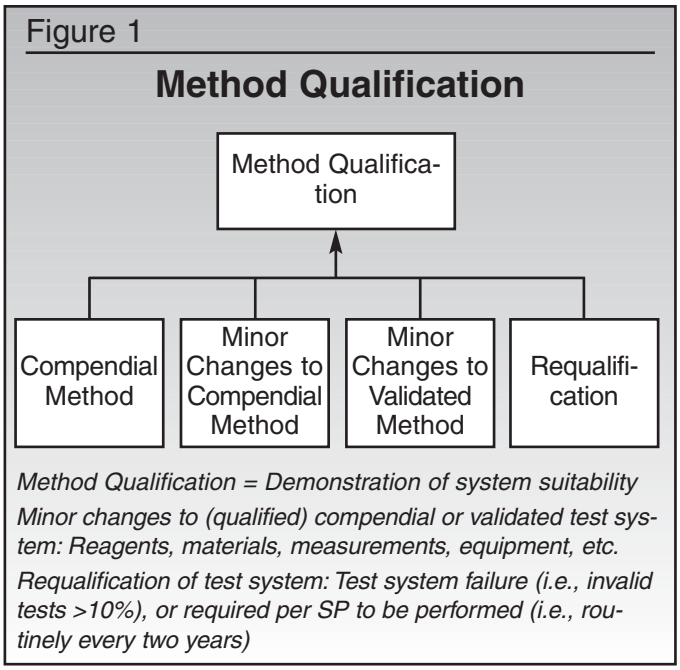
Qualification versus Validation

Differences in the definitions for analytical method qualifications versus validations among regulatory agencies, and often even within a company's functional units, have somehow lead to confusion for the user. Unfortunately, there is currently no guidance available why and when methods must be qualified or validated, with the exception of non-compendial methods used for release testing, which must be validated. It is important that these terms are clearly defined in the governing SP. This document should state at which drug development or process stage methods must be qualified or validated. Since regulatory agencies have still no harmonized classification for the terminology, it is not important how companies define these terms, as long as these are well-defined in the SP, and these definitions are understood and followed by all users.

Method qualification is demonstrating test system suitability. The method qualification report should contain an assessment of the accuracy and precision of the test results, and the appropriate number of samples to be tested. Method qualifications should be performed for all compendial methods used. Pre-set acceptance criteria for system suitability (accuracy, precision, and sample size) should be used in the qualification protocol, but may be "wider" than those in the validation protocol, since method qualification is not intended to challenge a compendial or otherwise validated test method.

Method validation is also demonstrating test system suitability, but requires a more rigorous approach. When following regulatory expectations set forth in the SP, a series of validation performance characteristics (see section on Validation Parameters) must be strictly followed. All non-compendial test methods must be validated in addition to those previously qualified, or validated methods, which have been automated or otherwise changed. Method validation must challenge system suitability, and should therefore have "tighter" pre-set acceptance criteria for each validation parameter in the validation protocol. Suggested applications for test method qualifications and validations

Figure 1

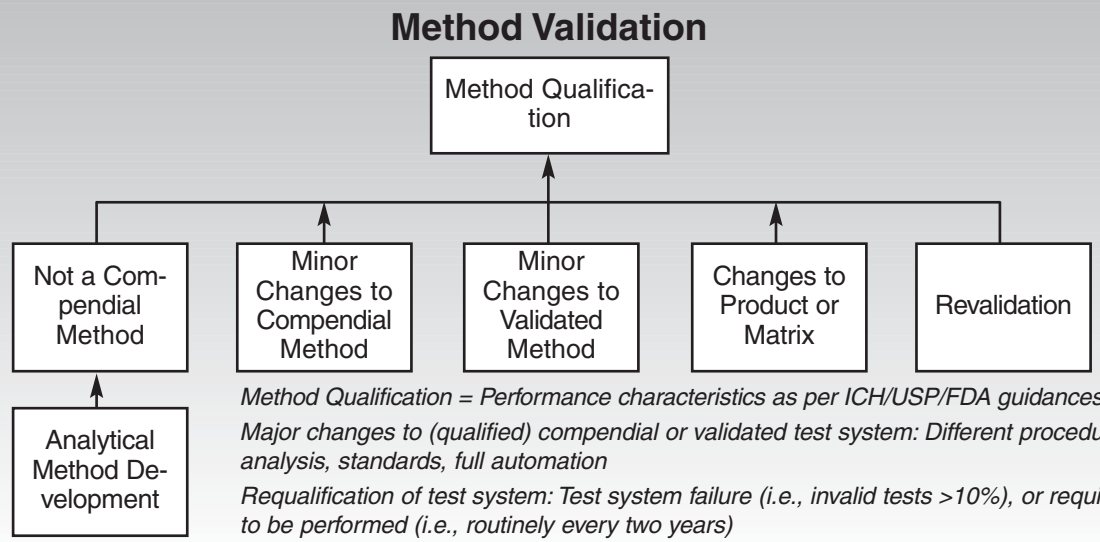


are given in *Figure 1* and *Figure 2*, respectively.

Test methods could also be defined as qualified when developed by the Analytical Method Development (AMD) group. This definition would then also correlate with the equipment and instrument terminology (qualification), and the AMD report could then be used as the Performance Qualification (PQ) for the relevant instrument(s) after the Design Qualification (DQ), Installation Qualification (IQ), and Operation Qualification (OQ) are executed and approved.

All test method qualification and validation protocols should contain acceptance criteria for each studied validation parameter. Not only validation, but also, qualification protocols should be generated with clear expectations what criteria an instrument or method must routinely produce. Qualifications may

Figure 2



include "wider" acceptance criteria than validations because, in general, less history of system suitability and overall assay performance may be available.

Assigning Responsibilities

Recent trends of auditors that criticize companies for the lack of assigned responsibility and accountability should be considered when these areas are defined in the SP. The method development reports should be reviewed by the AMV department and approved by QA management. The AMV protocol and report should be prepared by experienced AMV scientists, reviewed by the AMV manager, QC manager or supervisor, and by the leading QC operator. Review by the leading QC operator ensures that all technical details of the protocol are corresponding to the Standard Operating Procedure (SOP), and are followed by the operator(s), and that all data and deviations are properly documented. QA management must approve all documents required by the U.S. Code of Federal Regulations (CFR) Title 21. The AMV department is responsible for the timely completion (i.e., six months) of the validation, data, information integrity, and compiling and archiving of the AMV package. *Figure 3* illustrates the assigned department responsibilities, interfaces, and the overall process flow of a new analytical test method.

AMV Documentation

Nothing demonstrates control and overall integrity more than serialized (in red) pages for the AMV protocols and reports. This is truly the ultimate document control, ensuring that acceptance criteria and results cannot be changed during or after the validations are performed. Although not currently expected from regulatory agencies, this cumbersome effort will be rewarded with an overall positive impression on the regulatory agency auditor that AMVs are performed at the highest quality level.

AMV protocols and reports should have unique identification numbers, such as AMVP-02-001 for Analytical Method Validation Protocol generated chronologically as the first one in the year 2002 (AMVR-02-001 would be the corresponding report). An addition to the protocol or report is defined as an addendum, and should be denoted as such (i.e., AD = Addendum, therefore: AMVP-02-001-AD). A correction or change to the protocol or report is defined as an amendment, and should also be denoted as such (i.e., AM = Amendment so AMVP-02-001-AM). All identification numbers used for AMV documentation should be defined in the SP and correlate with general company document numbering rules. Ideally, the numbering could start with the analytical method development report (AMDR-02-001), but this may not be practical, since the AMV and AMD departments may report into different departments (i.e., QC versus Research & Development [R&D]).

Validation data must be generated using only acceptable data recording means, such as validated software applications and serial numbered laboratory

notebooks. Validation data should be, if possible, real-time reviewed (same day) by the operator and the relevant QC supervisor. Only original data must be included in the AMV package for archiving. Recommended documents to be included in the AMV package are listed below:

- Original Validation Data
- SOP
- Change Request for SOP (if required)
- Certificates of Analysis (C of A)
- Other Relevant Reagent and Reference Information
- Relevant Sample Information
- In-process and/or Product Specifications
- Operator Training Records
- Instrument Calibration Records
- Instrument User Logs
- Historical Data
- Statistical Analysis
- Operator Notebook Copies (if used)
- Analytical Method Development Report
- Analytical Method Validation Protocol
- Analytical Method Validation Report
- Addenda and/or Amendments (if done)

Once all documents have been reviewed by the AMV study originator, and are organized in a binder, the AMV package should then be archived in a limited access and secure location to avoid any losses of these critical documents. Clearly, nothing indicates disorganization and incompetence more than lost validation documents, which are a company's proof for the validity of all test results. Auditors will use the SP to compare it to the actual AMV documents, and will search for any lack of detail and clear guidance, and deviations or inconsistencies between the SP and AMV documents. The SP, AMV protocol, and AMV report should be well-structured, detailed documents that should be concisely written, well-formatted, and without grammatical or spelling errors.

A limit should be set for the maximum time allowed between the approvals of the AMV protocol and report (i.e., six months). Clearly, an exceeded time limit suggests problems with either the protocol, execution, actual results obtained (AMV is responsible), the lack of resources or subject matter expertise of delegated personnel from QA management. In the worst case, an excessive time delay may indicate problems with the method itself (AMD is responsible). The efficiency in AMV project completions is just as important as the overall quality, since even a poorly executed AMV is still better than a non-existing AMV.

Agency Guidance Documents

Depending on where the licensed product will be sold, published guidance from the FDA, United States Pharmacopoeia (USP), and/or International Conference on Harmonization (ICH) should be referred to in the SP, and the minimum requirements in the relevant guidelines should be integrated. Management should keep in mind that regulatory agencies will audit the AMV department based upon current guidance documents and the company's governing SP. It is beyond the scope of this article to discuss regulatory guidelines (Part II of this article will contain recommendations on how to integrate current guidelines into the AMV protocol). The reader can readily find these guidelines over the Internet (www.fda.gov/cber/guidelines.htm). Regulatory agencies will become more experienced in GAMVP, and may expect more solid and stringent AMVs in the future. It may therefore be advisable, in order to achieve long-term compliance, to deliver AMV packages now that can stand up to future expectations.

Although listed in the ICH Q2B guidelines for method validation and USP 25 <1225>, the robustness of a test method should be demonstrated and documented during method development. The validation parameter 'Ruggedness,' in USP 25 <1225> is equivalent to ICH Q2B's 'Reproducibility,' in that they describe overall precision among different laboratories. Both uses of terminology are different from 'Robustness' (deliberate changes to test conditions) and 'Inter-assay Precision' or 'Intermediate Precision' (overall precision within one laboratory). Unfortunately, there is still some confusion among industry experts when and how exactly to apply which terminology. Anyhow, 'Robustness' should be covered during development. This is clearly more economical, and makes more sense, since the validation scientist must know the level of robustness before a method is validated. A method should not be modified once the method development work is completed (including robustness), and the development report is approved. An unfinished method is unlikely to be ro-

bust enough to yield consistent test system suitability over time, and may cause test system failures and potential cumbersome and expensive efforts to fix the system. The function of the AMV department should clearly be to qualify and validate, not to modify and develop test methods.

Biological versus Biochemical/Chemical Methods

Biological assays are generally less quantitative than biochemical or chemical assays. In addition, less industry guidance on how to validate biological test methods is currently available. Although Agency expectations may therefore be less stringent, this however, does not make the task of biological method validation easier or more economical. Due to this general lack of experience by industry experts, the demonstration of test system suitability is often far from trivial.

The SP should differentiate between biological method qualifications (compendial) or validations (non-compendial), and biochemical or chemical methods (generally more quantitative). Protocol acceptance criteria must be thoroughly determined to incorporate expected results from qualitative versus quantitative tests, and normal versus non-normal (i.e., Poisson statistics) result distributions. Commonly used validation guidance documents (FDA, USP, ICH) should only be used with appropriate test methods. It may not be appropriate to incorporate the validation parameters, quantitation limit, linearity, and assay range when qualifying or validating a microbiological test procedure.

Validation Parameters

Assay performance criteria, such as accuracy and precision, are defined for chemical and biochemical methods in current guidelines, and should be covered according to assay classification (i.e., quantitative test, limit test). The validation parameters below are listed per ICH guidelines (Q2B):

- Specificity
- Linearity
- Assay Range
- Accuracy
- Repeatability Precision
- Intermediate Precision
- Reproducibility
- Limit of Detection
- Limit of Quantitation
- Robustness (if not done during AMD)
- System Suitability

Attention to assay classifications should be paid when one assay simultaneously tests for the main product and impurities (i.e., an electrophoresis assay may yield concentrations for the desired therapeutic protein, and also for other proteins not completely separated during purification). The SP should state when and how each validation parameter is to be executed. Guidance should be given how to derive acceptance criteria and use statistics (i.e., Analysis Of Variance [ANOVA], t-tests) to express confidence in each validation parameter. In-process or product specifications, historical data (assay development and/or routine testing), and guidelines set forth in the SP should be considered and integrated into each acceptance criterion. Other factors, as listed under *Figure 4*, should also be considered, since different methodologies and instruments may yield different assay performance expectations. More detail will be provided in Part II: Deriving Acceptance Criteria for the AMV Protocol.

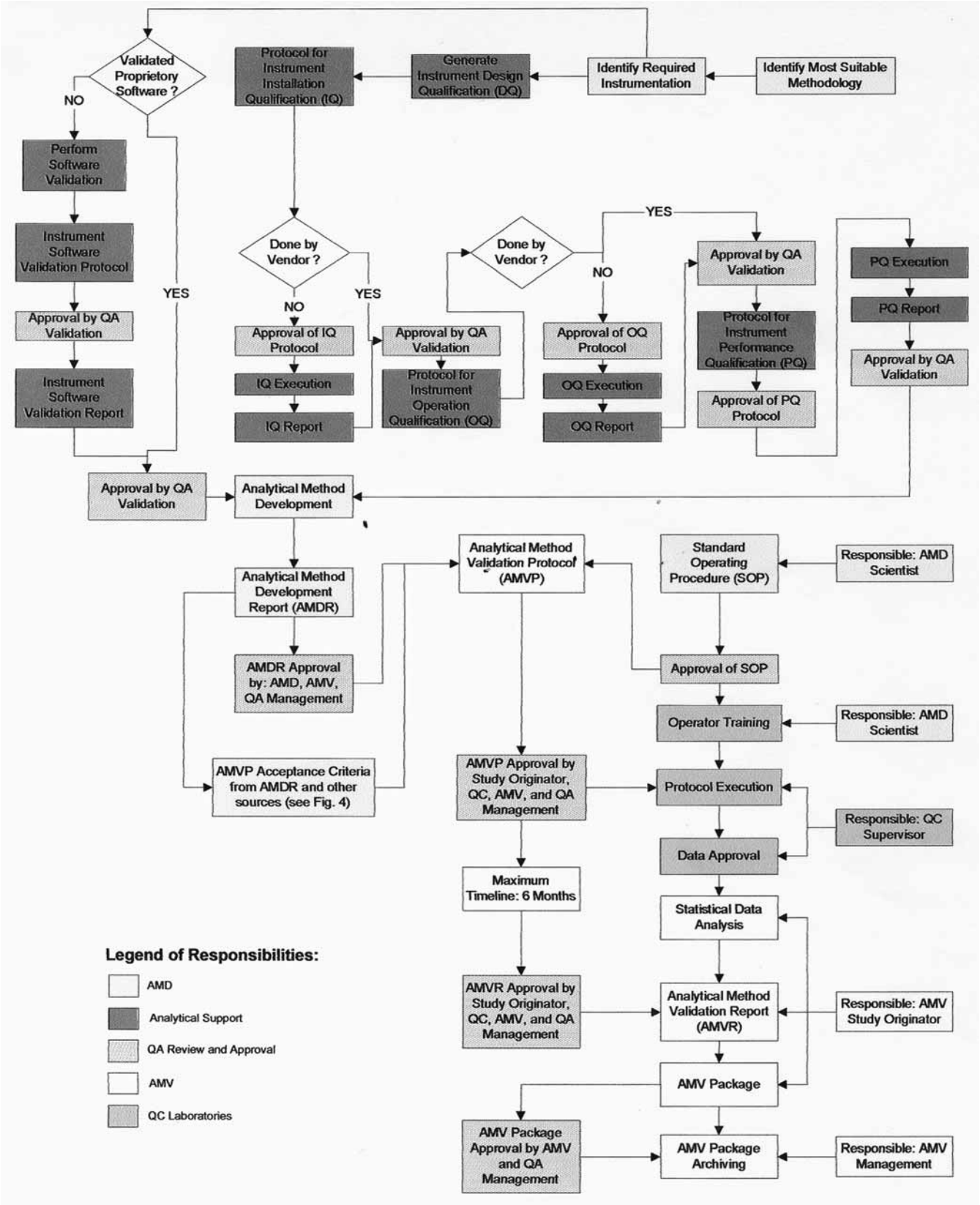
Acceptance Criteria and Statistics

The SP should state guidelines when and how descriptive statistics (i.e., mean, standard deviation) and comparative statistics (i.e., F-test, T-test) are to be used for AMVs, and how they are to be incorporated into acceptance criteria. For example, when using assay variability factors, such as multiple instruments, operators, and days to demonstrate intermediate assay precision (within one laboratory), it is recommended to use, whenever possible, n=3 of each factor. To demonstrate that none of these factors contribute significantly (at 95 % confidence) more than the remaining others, an ANOVA should be used to evaluate intermediate precision. If significant differences among factors are observed, a series of F-tests and Student's T-tests should be used to isolate not only particular factors, but also individual variability contributors, such as a particular operator.

However, statistics should only be applied when the conclusions are meaningful. Due to sample spiking and other non-routine sample handling during the validation execution, systematic errors may not be accounted for with simple statistics, since those assume random data distributions. Acceptance criteria should be derived thoroughly, connected to product specifications, sample size, and assay performance to clearly demonstrate test

Figure 3

Process Flow for Development and Validation of a New Analytical Method



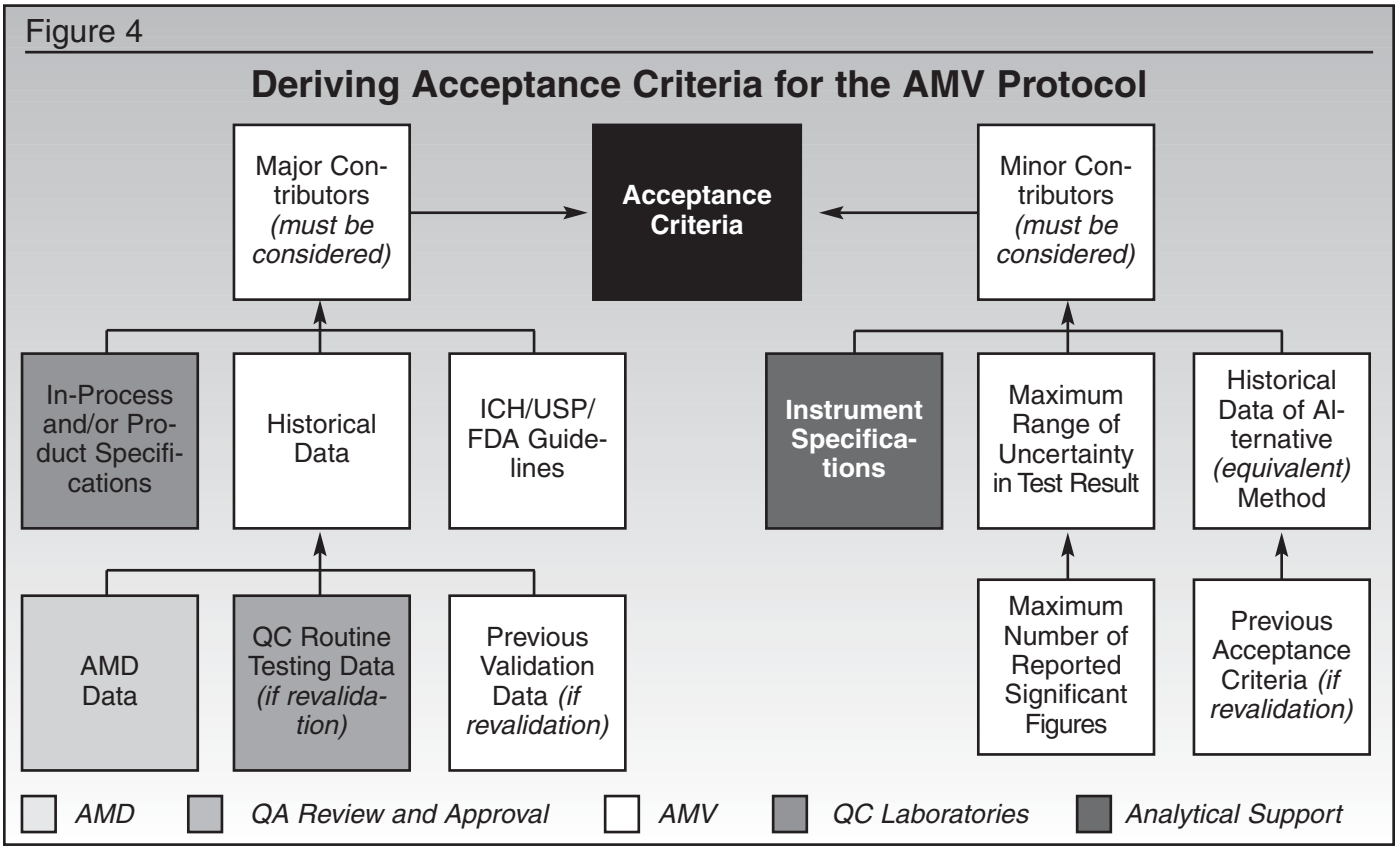
system suitability (see also *Figure 4* for more detail). These criteria may therefore be more convincingly derived from numerical limits readily available (i.e., instrument specifications set by the manufacturer). The SP should further include guidelines how the uncertainty of reported results (and therefore the number of significant digits used) is to be determined in AMVs. Whether to derive this uncertainty estimation from a propagation of error analysis, overall assay precision (intermediate precision), repeatability precision, or a combination of precision and accuracy (using an acceptable reference standard) may, at least for now, not be as important as the consistent implementation and adherence to this procedure. It may, therefore, be advisable to keep instructions simple, so that these can be consistently understood and followed by everyone involved. Management must realize that this procedure does impact how final release results and product specifications are to be reported. The author will provide more detail in Part II.

Requalifications and Revalidations

The SP should define when and to what extent test systems are to be requalified or revalidated (depending also on the initial definition for qualification versus validation). *Figure 1* and *Figure 2* provide applications for each of these two classifications. A good practice would be to set timelines for test methods to be requalified or revalidated even when test systems perform as expected, indicated by quarterly or annual QC trend reports. The requalifications or revalidations may then only be performed as partial requalifications and revalidations by demonstrating mainly accuracy and precision.

Whenever changes to the test system are required (i.e., increase in invalid tests), a full requalification or revalidation may be required. Once compendial (USP) methods are modified or automated, these methods should then be validated. Changes to the product, product formulation, or product matrix are considered major changes by the regulatory agencies and require test method validations, since these changes may impact test system suitability. It is, at least partially, a business decision how these conditions are defined in the SP, since any additional AMV projects will require more resources devoted from the AMV, QC, and QA staff to complete these projects, and may potentially interfere with a continuous production process. However, management should keep in mind that the expectations of regulatory agencies

Figure 4



are likely to be more demanding in the near future.

The AMV-AMD Interface

These two departments should be separate quality units, since one department alone should not be allowed to control development and validation. This is simply the same GMP principle, from a compliance standpoint, that QC validations must be approved by the QA department. Depending on the company size and requirements for AMVs, the AMV-AMD interface could be defined in the SP. Although it may not be necessary to include this in the SP, this interface should be carefully analyzed from a compliance and workflow (economical) point.

The AMV department should be involved in the review and approval process for developed methods, since they are expected to be the best scientific resource on which method can be validated, and will be sufficiently robust to withstand years of problem-free routine QC testing. Problematic assays should never reach the AMV department, since those will be difficult, if not impossible, to validate, and would cause many problems during routine QC testing. On the other hand, the AMD scientists are the method experts, and should be involved in the generation of acceptance criteria and other technical aspects of the validation protocol. Ultimately, the AMV department is responsible for the integrity of all data generated leading to the approved validations. A friendly and professional work environment between these departments is critical for efficiency, workflow, compliance, and overall quality of final method validation packages, thus ensuring product quality and safety to the patient.

The AMV-QA Interface

The QA department oversees and controls the AMV department by their approval of all AMV documents. Whenever the number of AMV projects increases, the staffing of delegated personnel from QA management should proportionally increase to avoid hold-ups of AMV approvals. Although this may be logically concluded, companies often overlook this fact, and must then consequently pay the price by having AMV projects not completed on time; therefore potentially operating out of compliance. The QA department should clearly communicate their expectations for the AMV program, since they have the final word on why, when, and how things are to be completed.

In case that the AMV department is permitted to run, according to the SP, with a relatively high level of autonomy, the department should educate QA on the principles of AMVs, as performed. In many companies, this may be more likely to be the case, since AMV scientists may be more experienced in how to practically plan and execute AMVs. The QA department will, and should always have, a certain level of control, as it naturally is the department that assures that activities are conducted with the uttermost adherence to compliance and integrity. For the AMV department to be efficient, it should establish and maintain a professional work atmosphere, staffed with qualified personnel where documents are reviewed timely, carefully, and objectively, since this is critical for timely project completion and compliance.

The validation protocol should be executed by trained QC operators who routinely perform this test procedure in the appropriate QC laboratory, using only qualified instruments and equipment. This ensures that validation samples are tested blindly, and that validation results are directly comparable to future routine testing results. Whenever validation projects lag behind manufacturing, or when revalidations are required (new formulation, matrix change etc.), historical QC assay performance must be integrated in the AMV protocols.

The AMV department should also be integrated in the analysis of timely (i.e., yearly) QC test method evaluations to ensure that assays maintain suitability, as demonstrated in the AMV report. This is important because not all QC test methods may contain appropriate system suitability criteria (i.e., blanks, standards, controls, etc.). In addition, assay controls may "drift" over time from the historical average towards the control limits (i.e., ± 3 standard deviations), therefore not only potentially increasing the number of invalid results, but also indicating uncertainty in the overall test system suitability. Whenever test system suitability becomes uncertain (i.e., loss of accuracy and/or precision), the overall test system must be re-evaluated. Assay control drifts, for example, may be caused by unstable control material, change in reagents, or testing equipment (i.e., chromatography columns). In either case, overall test system suitability may be affected, and any required modifications to

qualified or validated test methods should be supported by corresponding requalifications or revalidations.

The SP should also state which procedure to follow in case validation samples are Out-of-Specification (OOS). This is important since OOS results for validation samples or routine samples may be otherwise classified as (failing) retesting.

Process and method validation project planning and execution should be well-orchestrated, and be run chronologically parallel to ensure that at some point, the complete production and testing system is validated, and therefore compliant. The process to achieve overall compliance is somehow analogous to the 'chicken and egg' paradox. Technically, the complete process cannot be validated as run during large-scale production without validated test procedures for in-process and final container testing. At the same time, test procedures cannot be validated for routine production samples when those are not produced under final validated production conditions.

Inspections by Regulatory Agencies

The category of the drug product (i.e., biological) will determine which agency branch (i.e., Center for Biologics Evaluation and Research [CBER], Team Biologics) will be responsible for conducting inspections. The FDA usually sends a team of inspectors, which may include a local representative from the Agency, depending on the district and the company being inspected. When an inspection will take place, and by which team members, can be predicted to some degree. It is also important to realize that different agencies communicate their inspection notes, and schedule their audits among themselves. Finally, inspection records should be provided to the AMV department, so that it can anticipate follow-up requests. Once the inspectors are on-site, and have eventually devoted attention to QC laboratories, senior management should have organized and delegated AMV or QA personnel to defend AMVs at that time. Delegated personnel should be familiar with the content of the AMV SP, and should understand all definitions therein. The inspectors usually (but not always) proceed systematically, requesting documents and clarifications, starting from general to specific procedures. The personnel presenting documents and clarifications to the inspectors have an impact as to where, and to which detail, the inspection is heading. This is why it is so important to have in place a solid SP that is consistently followed. The inspectors may 'spot-check' particular AMVs once they have familiarized themselves with the SP(s). At this time, a list of all in-process and final container assays (customized for each particular agency), and their qualification/validation status (as defined in the SP) should be provided to the inspectors upon request.

"It is also important to realize that different agencies communicate their inspection notes, and schedule their audits among themselves."

Commitments to the agencies should be conservative, since these will usually require additional resources. Like the level of insurance invested through a solid AMV program, it is again a delicate balance to what degree companies should commit to complete AMVs, since an instant 'pleasing' of the inspectors may not add any value to the AMV program and, therefore, not add to the safety of the patient. Potential severe penalties (i.e., consent decree) may be impinged, and a loss of trust may substantiate if these commitments are not completed by the committed dates. With a solid SP in hand, the defending company delegate may not have to extensively commit to additional AMV projects.

Conclusions

GAMVP is a continuous improvement process. The industry must set new standards to cope with the increasing requirements set forth by regulatory agencies which have, and will themselves, become more educated and experienced. The AMV SP must ensure that all test methods provide accurate and reliable results. The SP will be reviewed by auditors from reg-

ulatory agencies. Auditors will identify any lack of detail and guidance, and evident inconsistencies between the SP and AMV package. The SP will demonstrate to the auditors that the AMV department can consistently control the quality of the validation process. This document directs the number and detail of qualifications and validations to be expected under normal operating conditions. Management must anticipate the level of AMVs required for continuous compliance. The validity and integrity of the test results must be properly demonstrated. Only then may companies produce and sell their product. ❑

About the Author

Stephan O. Krause is the manager of the QC Analytical Validation department within the Biological Products division of Bayer HealthCare Corporation. He received a doctorate degree in bioanalytical chemistry from the University of Southern California. Dr. Krause can be reached by phone 510-705-4191.

Suggested Reading

1. Code of Federal Regulations, 21 CFR Part 211. Current Good Manufacturing Practice for Finished Pharmaceuticals. 2001
2. FDA. *Guidelines for Validation*. 1994
3. FDA. *Guide to Inspections of Pharmaceutical Quality Control Laboratories*.
4. International Conference on Harmonization (ICH), Q2A. "Validation of Analytical Procedures." *Federal Register*. Vol. 60. 1995.
5. ICH, Q2B, "Validation of Analytical Procedures: Methodology." *Federal Register*. Vol. 62. 1996.
6. United States Pharmacopoeia, USP 25 <1225>. "Validation of Compendial Methods."
7. Pharmacopoeial Forum, PF <1223>. "Validation of Alternative Microbiological Methods." Vol. 28 No. 1. 2002.
8. Huber, L. "Validation and Qualification in Analytical Laboratories." *Interpharm Press.* Englewood, CO. 1999.

Article Acronym Listing

ANOVA:	Analysis Of Variance
AMD:	Analytical Method Development
AMV:	Analytical Method Validation
CBER:	Center for Biologics Evaluation and Research
CFR:	Code of Federal Regulations
C of A:	Certificate of Analysis
DQ:	Design Qualification
FDA:	Food and Drug Administration
ICH:	International Conference on Harmonization
IQ:	Installation Qualification
GAMVP:	Good Analytical Method Validation Practice
OOS:	Out-of-Specification
OQ:	Operational Qualification
PQ:	Performance Qualification
QA:	Quality Assurance
QC:	Quality Control
R&D:	Research & Development
SOP:	Standard Operating Procedure
SP:	Standard Practice/Procedure
USP:	United States Pharmacopoeia

Validating Immunoassays Using the Fluorescence Polarization Assay for the Diagnosis of Brucellosis

An Example and as an Application to ISO Standards 9000 and 17025

By David Gall
and Klaus Nielsen
Canadian Food Inspection Agency, Animal Disease Research Institute

The statistical methods presented in this paper are readily available in standard epidemiological texts. Unfortunately, many researchers, diagnosticians, and laboratorians either are unaware or do not use these statistical techniques. Their use in validating immunoassays would greatly enhance confidence in the acceptance of a new assay at national and international levels. For regulatory agencies to accept and approve a new serological test at national or international levels, they must be assured that a new assay can distinguish true positive from true negative samples with minimum false positive and false negative results, with more accuracy than the currently used test. The best way to accomplish this is by using the various statistics (i.e., sensitivity, specificity, confidence limits, sample size, and kappa that are available from the authors in an Excel® 97 template). For the purposes of this manuscript, *brucellosis* will be used as an example using the fluorescence polarization assay of a new test to be validated. A review of the literature suggests that most of the current in use assays for the diagnosis of *brucellosis*, which are often incorrectly called gold standards, have not been validated. Yet, new assays are often compared to these current in use assays resulting in misunderstanding or incorrect conclusions about the new assay. This may result in non acceptance of a perfectly useful assay, due to poor data or choice of statistics. This can be avoided with better test design and the use of appropriate statistical methods to compare and validate new tests with current in use tests. With access to computers and software, it is now possible and easier to use statistics such as Receiver Operating Characteristics (ROC).

"For regulatory agencies to accept and approve a new serological test at national or international levels, they must be assured that a new assay can distinguish true positive from true negative samples..."

Thousands of samples can be measured, tabulated, and analyzed. An example is the validation of the fluorescence polarization assay with a sensitivity (n = 1084), specificity (n = 23,754) of 99.4% (98.7 – 99.7 95% Confidence Limits [CL]), and 99.8% (99.8 – 99.9 95% CL), respectively. Meeting internationally accepted criteria for validation and measurements of uncertainty as defined in ISO standards 9000 and 17025 will increase the likelihood of acceptance by international and national regulatory agencies. Failure to properly validate a new test may increase the opportunity for litigation.

What is validation? A review of scientific literature regarding general principles of validation is not common, but papers discussing validation of specific procedures or tests are numerous.[1-7] In the International Organization for Standardization (ISO) standard 8402, validation is defined as "the confirmation by examination and the provision of objective evidence that the particular requirements for a specific intended use are fulfilled."[8] This definition is used in ISO standards 9000, 17025, and the earlier reference manual on validation of test methods proposed by the European Cooperation for Accreditation of Laboratories.[9] For serological assays, Jacobson defined validation as an assay that consistently provided test results identifying animals as positive or negative, and by inference, accurately determining the disease status of animals with statistical certainty.[10] In other words, using objective evidence, confirm through examination that the test has fulfilled its intended use.

The United States Animal Health Association (USAHA) proposed criteria for evaluating experimental *brucellosis* tests for specificity and sensitivity.[11] However, new tests would have been subjected to criteria more rigorous than official in use tests required by the United States Department of Agriculture (USDA). For the acceptance of new tests, the association recommended that new tests be more sensitive and/or specific than current in use tests. If the sensitivity and/or specificity were comparable to the in use tests, then they should be less costly, easier to conduct, and adaptable to automation. Similarly, Wright et al., suggested that a new test must be equal or superior concerning diagnostic performance.[12] Nielsen et al., validated the Fluorescence Polarization Assay (FPA) for detection of antibodies to *Brucella abortus* using defined reference positive and negative samples.[13] Previous to the FPA, the Indirect Enzyme Immunoassay (IELISA) and the Competitive Enzyme Immunoassay (CELISA) were similarly validated.[14] The data were compared with the current in use tests, thus meeting the ISO definition of examination and objective evidence, the USAHA criteria for better sensitivity, specificity, cost, performance, and adaptation to automation, and superior diagnostic performance as suggested by Wright et al.[12]

Before validation of an assay can commence, a developed test must undergo various verification procedures designed to provide objective evidence that specified requirements have been fulfilled.[8] Optimization and standardization of an assay would be an essential part of this verification.[14] Optimization of a serological assay would include optimizing concentrations or dilutions of reagents, determining variation between replicate samples (i.e., quality control and test sera), determining variation in background activity, determining an initial cutoff and finally, determining initial test performance.[14] Standardization would include a standardized assay format, standardized preparation of all reagents such as buffers, chemicals, and biological reagents, and strict adherence to protocols and procedures between laboratories using the same assay.[14] Laboratories accredited using ISO 9000 and ISO 17025 standards are required to produce and show compliance with approved technical and operating procedures, thus facilitating validation of serological assays and technology transfer.

Test performance can be objectively measured in terms of repeatability and accuracy. Repeatability (a form of precision) is the ability of an assay or procedure to produce consistent results in repeated tests. Consistent results with quality control sera in a serological assay would be a good example of repeatability. Variation between replicate test samples and quality control sera would result in poor repeatability. the accuracy of an assay is the ability to identify positive and negative samples correctly. Noting that assays can be repeatable without being accurate, but not the reverse, is important.[15] Using defined reference samples, the accuracy of an assay can be described as sensitivity and specificity.

Sensitivity is the ability of a test to produce a positive result when the sample is from a diseased animal. It is measured as a percentage of the population with the disease that has a positive result.[16] Similarly,

specificity is the ability of a test to produce a negative result when the sample is from a healthy animal, and being a percentage of the healthy population that has produced a negative result. The underlying prevalence of a disease does not influence sensitivity or specificity.[16]

Relative sensitivity and relative specificity should not be confused with sensitivity and specificity. Relative sensitivity is the ability of a test to produce a positive result compared with another test or series of tests producing a positive result. Relative specificity is the ability of a test to produce a negative result relative to another test or series of tests. The other test(s) sensitivities and specificities should approach 100% if used to classify animals in the positive category or negative category resulting in a close approximation of the actual sensitivity and specificity.[17, 18] If ascertaining the actual disease status is difficult or costly, relative sensitivity and specificity may be used.[18] In comparing a new test that may be superior to another test or series of tests, the resultant relative sensitivity or specificity may be lower, resulting in false or wrong conclusions.

A known number of reference positive and negative samples representative of the population under study must be assembled for estimation of sensitivity and specificity. Reference positive and negative samples can be defined in any way, if the definition is explicit.[19] In veterinary medicine, reference positive and negative samples are usually determined using a "gold standard" which could be another test, procedure or multiple tests or procedures.[16] Ideally, reference positive samples should be obtained from individual animals of known disease status, or from animals whose herd status was known through disease status.[12] However, obtaining samples of known disease status may not be possible, so the best available method, such as current in use tests, could be used.[16] For example, using immunologically based tests to examine a new immunologically based test is acceptable, but may introduce a bias against the new test because the evaluation of the new test was limited to those animals selected by the other test may not be measuring the same antibody population depending on the antigen used.[10, 20] Similarly, a combination of other tests or procedures could be used to define reference negative samples. Once assembled, testing of the samples should be conducted blindly.

How many samples are required when assembling reference samples for validation? Failure to consider sample size may affect sensitivity and specificity estimates. Often, more samples are required to demonstrate small differences in sensitivity or specificity.[19] For example, 15,000 samples would be required to distinguish between tests with specificity differences of 0.1%.[21] More important, sample size affects the confidence limits for calculated values of sensitivities and specificities. Too few samples result in wide confidence limits, negating any usefulness of the data or resulting in the wrong conclusions about the new test. The greater the sample size, the better the confidence limits for sensitivity and specificity. More data and tighter confidence limits raise the confidence in the test to distinguish between positive and negative, especially near the cutoff. As well, larger size increases the probability of being representative of the field population.

Associated with test performance is uncertainty of measurement which describes dispersion or variation of data about a value. Examples of measures of dispersion are range, standard deviation, and confidence intervals. Sensitivity and specificity estimates are subject to sampling variation, and as such should have CL which are a measure of uncertainty as mentioned in the ISO standard 17025.

Materials and Methods

Reference Samples

The FPA is an appropriate example of test validation.[22] The defined positives (n = 1084) were samples selected from animals from which *Brucella abortus* was isolated from various tissues or secretions (Gold Standard). The defined negatives (n = 23,754) were randomly selected Canadian samples submitted for routine testing from animals with no previous clinical or epidemiological evidence of Brucellosis. Canada was officially declared free of *bovine brucellosis* in 1985.

Statistical Data Analysis

Sensitivity, relative sensitivity, specificity, and relative specificity are determined as indicated in *Figure 1* and *Figure 2*. A template in Excel® 97 is available from the authors.

A simple formula for estimating the lower and

Figure 1

Determination of Sensitivity and Specificity (Accuracy)

Test/Disease Status	Disease (+)	Disease (-)
New Test Positive (+)	A (true positives)[a,e]	B (false positives)[c]
New Test Negative (-)	C (false negatives)[d]	D (true negatives)[b,f]

a. True positives are those samples testing positive and have the disease.
b. True negatives are those samples testing negative and do not have the disease.
c. False positives are those samples testing positive and do not have the disease.
d. False negatives are those samples testing negative and have the disease.
e. Sensitivity (in percent) = (A / A+C) x 100
f. Specificity (in percent) = (D / B+D) x 100

Figure 2

Determination of the Relative Sensitivity or Specificity

New Test/ In Use Test(s)	In Use Test(s) (+)	In Use Test(s) (-)
New Test Positive (+)	A (true positives)[a,e]	B (false positives)[c]
New Test Negative (-)	C (false negatives)[d]	D (true negatives)[b,f]

a. True positives are those samples testing positive and classified postive.
b. True negatives are those samples testing negative and are classified negative.
c. False positives are those samples testing positive and are classified negative.
d. False negatives are those samples testing negative and are classified positive.
e. Relative sensitivity (in percent) = (A / A+C) x 100
f. Relative specificity (in percent) = (D / B+D) x 100

upper confidence limits is presented in *Figure 3*.[19, 23] A template in Excel 97 is available from the authors. The formula is a normal approximation to calculate 95 percent confidence limits providing symmetrical confidence limits about the point estimate. When the point estimate approaches 100%, the confidence limit often exceeds 100%. This formula may be used when the sensitivity and specificity values are of moderate size ($0.3 \leq p \leq 0.7$).[24] A better estimate of confidence limits is the non symmetrical limits presented in *Figure 4*, especially if the sensitivity or specificity value is near zero or 100%.[24] A template in Excel 97 is available from the authors for the non symmetrical limits.

Frequently, results from known reference positive and negative samples overlap, creating a certain amount of uncertainty regarding the choice of a cutoff and the resultant sensitivity and specificity as depicted in *Figure 5*.[25] The exact placement of the cutoff is subjective. An otherwise excellent assay's chances for acceptance are reduced through a poor choice of cutoff. An ROC curve, plotting sensitivity against specificity results at various cutoff points (*Figure 6*), removes the subjectivity inherent in frequency distributions shown in *Figure 5*. This is the only measure available that is uninfluenced by decision biases and prior probabilities comparing different assays with a common easily understandable scale.[26] *Figure 6* is also a graphic representation of the relationship between sensitivity and specificity estimates that is easily determined using ROC software.[27] Each point on a ROC curve represents a two x two table of sensitivity and specificity estimates associated with a cutoff value from the lowest to the highest value (*Figure 6*). Along the diagonal line, a true positive response equals a false positive response and is often called the chance line. The greater the curve above this chance line, the better the discrimination between the reference positive and negative samples. As well, the Area Under the ROC Curve (AUC) is a good measure of detection, and is useful when distribution assumptions cannot be made or do not hold.[25, 28] An AUC of 0.91 implies that a randomly selected sample from the positive group will test higher than a randomly selected sample from the negative group 91% of the time with 95% certainty.[29] ROC curve analysis can also compare the assay performance of two or more tests.[30]

Sample Size

A simple formula for determining sample size for sensitivity or specificity is presented in *Figure 7*, and is available from the authors in an Excel 97 template.[10, 31] This formula is useful for surveys when an estimate for allowable error, sensitivity, or specificity is available, but could also be used to help a researcher in determining an appropriate sample size for validating

Figure 3

Calculation of Symmetrical 95% Confidence Limits for Sensitivity or Specificity

Using the Fluorescence Polarization Assay (FPA) as an Example

$$L = 1.96 \times \sqrt{([pxq]/n)}$$

L = the lower or upper 95% limit for p

p = the observed proportion (i.e., sensitivity or specificity value)

q = 1 – p

n = total number of samples

Example: In a reference positive population of 1084 animals infected with *Brucella abortus*, 1077 tested positive on the FPA while seven tested negative. The sensitivity point estimate (p) was 1077/1084 or 99.4%.

p = 0.994, q = 0.006, n = 1084

$$L = 1.96 \times \sqrt{([0.994 \times 0.006]/1084)}$$

Lower 95% for p: 0.994 - 0.005 = 0.989 or 98.9%

Upper 95% for p: 0.994 + 0.005 = 0.999 or 99.9%

Figure 4

Calculation of Non Symmetrical 95% Confidence Limits for Sensitivity or Specificity

Using the Fluorescence Polarization Assay (FPA) as an Example

$$P_L = \frac{(2np + C^2_{\propto/2} - 1) - C^2_{\propto/2}\sqrt{C^2_{\propto/2} - (2 + 1/n) + 4p(nq + 1)}}{2(n + C^2_{\propto/2})}$$

$$P_u = \frac{(2np + C^2_{\propto/2} + 1) + C^2_{\propto/2}\sqrt{C^2_{\propto/2} + (2 + 1/n) + 4p(nq - 1)}}{2(n + C^2_{\propto/2})}$$

P_L = the lower 95% limit for p

P_u = the upper 95% for p

p = the observed proportion (i.e., sensitivity or specificity value)

q = 1 – p

n = total number of samples

$C^2_{\propto/2}$ = 95% confidence limits for p of 1.96

Example: In a reference positive population of 1084 animals infected with *Brucella abortus*, 1077 tested positive on the new test while seven tested negative. The sensitivity point estimate (p) was 1077/1084 or 99.4%.

p = 0.994, q = 0.006, n = 1084

P_L = 98.7%

P_u = 99.7%

a new test.[31] The allowable error is a percentage error expressed as a decimal allowed for the estimate of sensitivity or specificity. Using the formula, the data presented in *Figure 8* are the estimates of the number reference positive animals required for each sensitivity listed at the top of the figure. The allowable error is listed on the left column of the table (i.e., the expected estimate for FPA sensitivity is within 1% of the true level 95% of the time). In brackets, are the calculated non symmetrical 95% confidence limits using the formula presented in *Figure 4* for the expected sensitivities or relative sensitivities. The limits get wider as the sample size decreases. Similarly, the formula can be used to estimate the sample size required for specificity or relative specificity as presented in *Figure 9*.[10, 20, 32] Since specificity or relative specificity is expected to be higher than sensitivity, the range of specificity and allowable errors are different. Due to the greater sample size, the confidence limits are narrow.

Results

The 95% CL presented in *Figure 3* are calculated using the template provided in Excel 97. The limits are symmetrical and should only be used when either sensitivity or specificity is of moderate value (between 30% and 70%). If 99% confidence limits are required, 2.56 can be substituted for 1.96. From the example for the FPA, the lower 99% for p (sensitivity point estimate) would be 98.8% versus 98.9% for 95% CL, and the upper 99% for p would be 100% versus 99.9% for 95% CL.

For the same data presented in *Figure 3*, the non symmetrical limits presented in *Figure 4* for p at 95%, CL is 98.7% for the lower limit, and 99.7% for the upper limit. When the sensitivity or specificity value approaches 0% or 100%, non symmetrical confidence limits should be used.

Each point on a ROC curve represents a two by (x) two table of true positive and false positive estimates associated with a cutoff value from the lowest to the highest value as presented by *Figure 6* for the FPA. The sensitivity and specificity for the optimal cutoff

Figure 5

Frequency Distribution of Reference Positive and Negative Samples

Exact placement of the cut off is subjective, affecting the sensitivity and specifity of the assay.

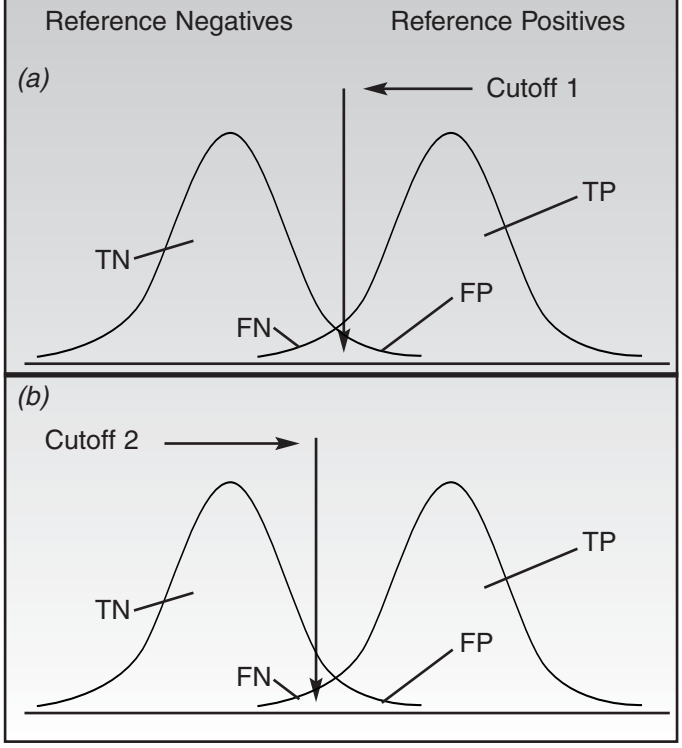
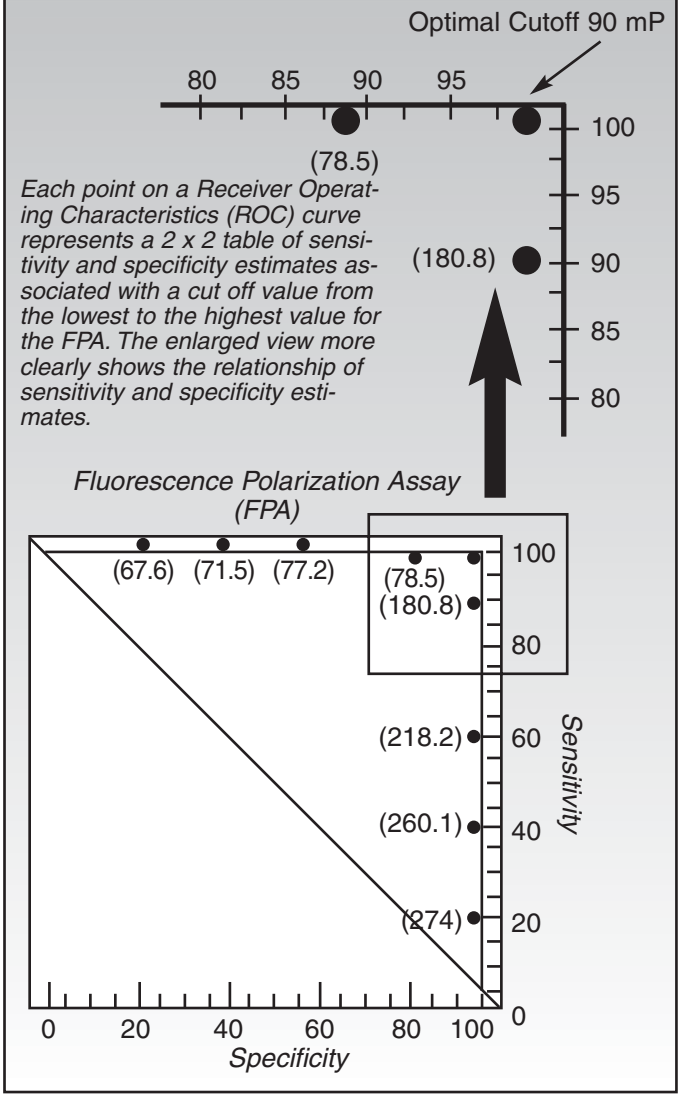


Figure 6

Relationship of Sensitivity and Specificity Estimates



values of 90 mP are 99.4% (98.7 - 99.7) and 99.8% (99.8 - 99.9), respectively. If sensitivity were wanted over specificity, then a lower cutoff value of 90 mP could be chosen as illustrated in *Figure 6,* removing any possibility of false negatives. The converse would be true for specificity. For instance, the sensitivity and specificity for cutoff value 67.6 mP are 100% (100 - 100) and 20% (20.3 - 21.3), respectively, while for cutoff value 274 mP, the sensitivity and specificity are 20% (17.7 - 22.5) and 100% (100 - 100), respectively. The numbers in brackets after each point estimate for sensitivity and specificity are the 95% confidence intervals as determined by software.[27] The AUC for the FPA, as presented in *Figure 6* and determined by software, was 0.999 (0.999 - 1.000) indicating that a randomly selected sample from the positive group will test higher than a randomly selected sample from the negative group 99.9% of the time with 95% certainty.

Using the data presented in *Figure 8*, the sample size required for the FPA (expected sensitivity equals 99 ± 1%) was 396 samples giving confidence limits of 97.3% to 99.7%. The actual size of 1084 samples exceeded the number calculated, resulting in a sensitivity of 99.4% (98.7% to 99.7% CL). Similarly, the sample size required with an allowable error of 0.1% for a specificity of 99% was 39,600 with confidence limits of 98.9% to 99.1% from *Figure 9*. The actual sample size was 23,754, giving a specificity of 99.8% with confidence limits of 99.8% to 99.9%.

As previously mentioned, sample size also affects the confidence limits for sensitivities and specificities. For example, using the formula illustrated in *Figure 3,* the range between the lower confidence limit (98.8%) and the upper confidence limit (100%) is 1.2% for a sample size of 600. Using the same

Figure 7

Formula for Calculation of the Number of Samples Required for Expected Sensitivity or Specificity

The Allowed Error is a Percentage Error Expressed as a Decimal for Each Estimate of Sensitivity or Specificity

$$n = ((4 \times p \times q)/L)^2$$

Where:

n = the number of samples required

p = the expected proportion

q = 1 – p

L = the allowable error

The number 4 = approximate square of Z = 1.96

which provides a 95% confidence level
for 99% confidence level the number 6.6
(z = 2.56) should be substituted

$$\text{ex. } n = ((4 \times .99 \times .01)/.01)^2 = 396$$

data, a sample size of 1084 results in a range of 1% between the lower (98.9%) and upper confidence limits (99.9%), resulting in a higher confidence in the data. For a sample size of 10,000, the range is further reduced to 0.4%. For the FPA, the range is very tight for sensitivities and specificities, since the sample sizes for the reference positive and reference negative are large. To decrease the range of the confidence limits from 1% to 0.4% required increasing the sample size approximately nine fold.

The flowchart presented in *Figure 10* summarizes the steps required to validate an assay such as the FPA.

Conclusion

Not covered in this article is another statistic known as "Kappa" used to determine agreement between tests. Kappa may be used when the disease status is unknown, and assembling known reference positive and negative samples is not possible. A kappa value of 0 suggests no agreement beyond chance while a kappa value of 1 reflects perfect agreement. Generally, excellent agreement is greater than 0.7.[16] The possibility exists, however, that agreement can occur by chance alone for two tests being compared when both exceed 50% in sensitivity and specificity.[10] As well, the kappa value is affected by the prevalence of the disease in the population of interest. Two poorly validated tests may have good agreement because they are measuring the same antibody or antigen. Consequently, kappa statistics should be used cautiously. A template in Excel 97 is available from the authors for determination of kappa.

The other statistical techniques presented in this paper are available in standard epidemiological texts.[31] However, the use of these techniques for validating immunoassays is sparse. Both Martin and Swets et al. have alluded to this.[18, 33] Use of these statistical techniques would increase the likelihood of successfully validating a new test, and acceptance by regulatory agencies responsible for approving new diagnostic assays. Most new assays or even current in use assays for the serological diagnosis of brucellosis, such as the Rose Bengal Plate Test (RBPT) and the Tube Agglutination Test (TAT), have not been or are not validated.

Of the statistical techniques presented in this paper, the formula for calculation of sample size is the most important. Insufficient sample size could result in wrong conclusions about a newly developed assay or comparisons to other assays. At the very minimum, 300 samples should be selected for estimates of sensitivity or specificity.[10] Obtaining the desired number or sample type is not always possible due to resource or logistical limitations, or prohibitive costs. Although the ideal number and type of samples are not available, most assays can still be validated if the sample definitions are clear. Conclusions can be drawn from ROC analysis of as few as 100 samples.[19] The CLs for the resulting sensitivities and specificities will, of course, be wider because of fewer samples.

Another longer term approach to validation of an assay is the banking of samples to obtain the number and type of samples.[20] This approach has the added advantage of linking the past, present, and future producing more reliable validation data, and may be the only method available for assays where collection of samples is difficult, hazardous, or costly.

Other factors that can influence validation of an assay are calibration and maintenance of equipment, analyst proficiency, training and laboratory conditions.

Figure 8

Expected Sensitivity or Relative Sensitivity

Number of reference positive samples required for the expected estimates for sensitivity or relative sensitivity of a new test with an allowable error for sensitivity or relative sensitivity within a percentage of the true level 95% of the time.

Allowable Error	85%[a]	90%	95%[a]	99%
1%[b]	5100[c] (83.98 – 85.96%)[d]	3600 (88.96 – 90.95%)	1900 (93.90 – 95.92%)	396[e] (97.27 – 99.68%)
2%	1275 (82.89 – 86.89%)	900 (87.81 – 91.84%)	475 (92.53 – 96.71%)	99 (93.71 – 99.95%)
5%	204 (79.18 – 89.46%)	144 (83.61 – 94.17%)	76 (86.71 – 98.44%)	16 (74.52 – 99.77%)
10%	51 (71.65 – 92.99%)	36 (74.34 – 96.98%)	19 (72.23 – 99.77%)	4 (38.75 – 98.09%)

a. Estimates of sensitivity.
b. Allowable error for estimate of sensitivity or relative sensitivity within a chosen percentage of the true level 95% of time.
c. Estimate of number of reference positive samples required to achieve sensitivity.
d. Non symmetrical 95% confidence limits for each estimate of sensitivity or relative sensitivity.
e. Example from *Figure 7*.

Figure 9

Expected Specificity or Relative Specificity

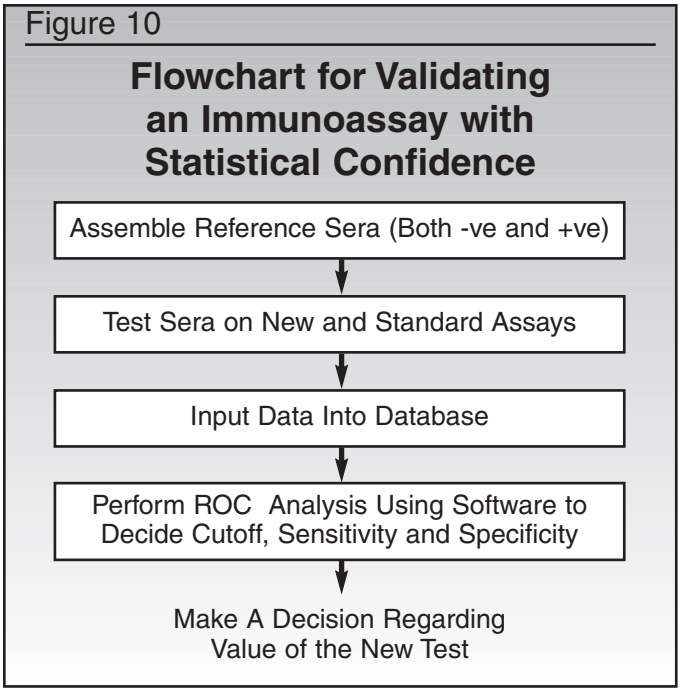
Number of reference negative samples required for the expected estimates for specificity or relative specificity of a new test with an allowable error for specificity or relative specificity within a percentage of the true level 95% of the time.

Allowable Error	90%[a]	95%	97%[a]	98%[a]	99%
0.01%[b]	36,000,000 (89.99 – 90.01)	19,000,000 (94.99 – 95.01)	11,640,000 (96.99 – 97.01)	7,840,000 (97.99 – 98.01)	3,960,000 (98.99 – 99.01)
0.1%	360,000 (89.90 – 90.10)	190,000 (94.90 – 95.10)	116,400 (96.90 – 97.10)	78,400 (97.90 – 98.10)	39,600 (98.90 – 99.09)
0.5%	14,400 (89.50 – 90.48)	7,600 (94.48 – 95.47)	4,656 (96.46 – 97.46)	3,136 (97.43 – 98.45)	1,584 (98.34 – 99.41)
1%	3,600[c] (88.96 – 90.95)[d]	1,900 (93.90 – 95.92)	1,164 (95.81 – 97.87)	784 (96.68 – 98.82)	396 (97.27 – 99.68)

a. Estimates of specificity.
b. Allowable error for estimate of specificity or relative specificity within a chosen percentage of the true level 95% of time.
c. Estimate of number of reference negative samples required to achieve specificity.
d. Non symmetrical 95% confidence limits for each specificity and allowable error.

Data produced from instrumentation that is not well maintained or calibrated is suspect and will affect the validation outcome. The estimates of sensitivity, specificity, and ROC analysis could be erroneous because of poorly maintained and calibrated equipment. In the wrong hands, a new test that performed well in the researcher's laboratory fails because the analyst who was evaluating the test was not properly trained re-

Figure 10



garding the test, or does not have the suitable background for evaluating the test. As a result, the test does not gain acceptance. The analyst may not be proficient due to poor laboratory techniques such as mixing of samples or pipetting of samples. This can be addressed by using proficiency panels to assist the analyst in becoming proficient. Laboratory conditions can affect the validation of assays, such as improper storage of samples, which affects the quality of the samples and the resultant data produced. The above factors are covered in ISO 9000 and ISO 17025 standards. Laboratories accredited through these standards are required to maintain records, thus improving the likelihood of proper validation.

Validation or validated assays elicit various interpretations and responses.[34-36] In veterinary medicine, it is an assay that consistently provides results that correctly identify samples as positive or negative.[10] Validation to other researchers is a time-limited process, or an ongoing process of the assay performance. Validation is also defined as "the confirmation by examination and the provision of objective evidence that the particular requirements for a specific intended use are fulfilled."[8, 37, 38] The Food and Drug Administration (FDA) defines the term as "Establishing documented evidence which provides a high degree of assurance that the specific process will consistently produce a product meeting its predetermine specifications and quality attributes."[39] Regardless of which definition is used, statistical analysis of the data increases the scientific integrity of the assay, therefore gain national and international acceptance. ❑

About the Author

David Gall and Klaus Nielsen of the Canadian Food Inspection Agency, Animal Disease Research Institute have been involved in the development, optimization, standardization, quality control, validation and technology transfer of primary binding assays such as enzyme immunoassays (ELISA), fluorescence polarization assay (FPA) for the past 20 years. Gall and Nielsen have been involved in the successful transfer and acceptance of immunoassays at the national and international levels. They can be reached by phone at 613-228-6698, and by fax at 613-228-6667. David Gall can be reached by e-mail at gall@em.agr.ca.

References

1. Biancifiori, F., Garrido, F., Nielsen, K., Moscati, L., Duran, M. and Gall, D. (2000) "Assessment of a Monoclonal Antibody-Based Competitive Enzyme Linked Immunosorbent Assay (CELISA) for Diagnosis of Brucellosis in Infected and REV 1 Vaccinated Sheep and Goats." Microbiologica 23: 2000. pp. 399-406.
2. Gall, D., Nielsen, K., Forbes, L., Davis, D., Elzer, P., Olsen, S., Balsevicius, S., Kelly, L., Smith, P., Tan, S. and Joly, D. (2000). "Validation of the Fluoresence Polarization Assay and Comparison to Other Serological Assays for the Detection of Serum Antibodies to *Brucella abortus* in Bison." *Journal of Wildlife Dis.* Vol. 36, No. 3. 2000. pp. 469-476.
3. Paulo, P. S., Vigliocco, A., Ramondino, R. F., Marticorena, D., Bissi, E., Briones, G., Gorchs, C., Gall, D. and Nieslen, K. "Evaluation of Primary Binding Assays for Presumptive Serodiagnosis of Swine Brucellosis in Argentina." *Clinical and Diagnostic Laboratory Immunology* Vol. 7 No. 5. 2000. pp. 828-831.
4. Samartino, L., Gregoret, R., Gall, D. and Nielsen, K. "Fluorescence Polarization Assay: Application to the Diagnosis of *Bovine Brucellosis* in Argentina." *Journal of Immunoassay*. Vol. 20 No. 3. 1999. pp. 115-126.
5. Gall, D., Colling, A., Marino, O., Moreno, E., Nielsen, K., Perez, B. and Samartino, L. "Enzyme Immunoassays for Serological Diagnosis of *Bovine Brucellosis*: A Trial in Latin America." *Clinical and Diagnostic Laboratory Immunology.* Vol. 5 No. 5. 1998. pp. 654-661.
6. Vanzini, V. R., Aguirre, N., Lugaresi, C. I., de Echaide, S. T., de Canavesio, V. G., Guglielmone, A. A., Marchesino, M. D. and Nieslen, K. (1998). "Evaluation of Indirect ELISA for the Diagnosis of *Bovine Brucellosis* in Milk and Serum Samples in Dairy Cattle in Argentina." *Preventative Veterinary Medicine*. Vol. 36. 1998. pp. 211-217.
7. Nielsen, K. H., Kelly, L., Gall, D., Balsevicius, S., Bossé, J., Nicoletti, P. and Kelly, W. Comparison of Enzyme Immunoassays for the Diagnosis of *Bovine Brucellosis*. *Preventative Veterinary Medicine*. 26: 1996b. pp. 17-32.
8. International Organization for Standardization (ISO 8402, 1994). Quality Management and Quality Assurance - Vocabulary.
9. European Cooperation for Accreditation of Laboratories

(1997). Validation of Test Methods: General Principles and Concepts EA-2/06. 37, rue de Lyon, FR-75012 Paris, France.
10. Jacobson, R. H. "Validation of Serological Assays for Diagnosis of Infectious Diseases." *Rev. Sci. Tech. Off. Int. Epiz.* Vol. 17 No. 2. 1998. pp. 469-486.
11. The United States Animal Health Association. Brucellosis Scientific Advisory Committee. Critique of Proposed "Criteria for Evaluating Experimental Brucellosis Tests for Specificity and Sensitivity." Appendix F: 1987. pp. 290-293.
12. Wright, P. F., Nilsson, E., Van Rooij, E. M. A., Lelenta, M and Jeggo, M. H. "Standardisation and Validation of Enzyme-Linked Immunosorbent Assay Techniques for the Detection of Antibody in Infectious Disease Diagnosis." *Rev. Sci. Tech. Off. Int. Epiz.* Vol. 12 No. 2. 1993. pp. 435-450.
13. Nielsen, K., Gall, D., Jolley, M., Leishman, G., Balsevicius, S., Smith, P., Nicoletti, P. and Thomas, F. "A Homogeneous Fluorescence Polarization Assay for Detection of Antibody to *Brucella Abortus*." *Journal of Immunological Methods*. 1996c. pp. 161-168.
14. Nielsen, K., Gall, D., Kelly, W., Vigliocco, A., Henning, D. and Garcia, M. (1996a). Immunoassay Development: Application to Enzyme Immunoassay for the Diagnosis of Brucellosis. Agriculture and Agri-Food Canada, 3851 Fallowfield Road, Nepean, Ontario, Canada K2H 8P9. ISBN 0-662-24163-0.
15. Cannon, R. M. and Roe, R. T. (1982). "Livestock Disease Surveys: A Field Manual for Veterinarians." Australia Government Publication Serv. Canberra, Australia. ISBN 0-644-02101-2.
16. Martin, S. W. "The Interpretation of Laboratory Results." *Veterinary Clinical American Food Animal Practice*. Vol. 4 No. 1. 1988. pp. 61-78.
17. Baldock, F. C. (1988). Epidemiological Evaluation of Immunological Tests. In: ELISA Technology in Diagnosis and Research, Ed. Graham Burgess, James Cook University of North Queensland, Townsville, Australia.
18. Martin, S. W. (1977). The evaluation of tests. *Canadian Journal Comparative Medicine.* Vol. 41 No.1. pp. 19-25.
19. Metz, C. "Basic Principles of ROC Analysis. Seminars in Nuclear Medicine." Vol. 3 No. 4. 1978. pp. 283-298.
20. McNab, B. (1991). General Concepts for Evaluating Test Performance and Making Decisions: Notes for Program Designers, Managers and Researchers. Agriculture and Agri-Food Canada, Animal and Plant Health Directorate, Nepean, Ontario, Canada (Unpublished).
21. Crofts, N., Maskill, W. and Gust, I.D. (1988). Evaluation of Enzyme-Linked Immunosorbent Assays: a Method of Data Analysis. *Journal of Virological Methods*. Vol. 22. pp. 139.
22. Nielsen, K. and Gall, D. "Fluorescence Polarization Assay for the Diagnosis of Brucellosis: a Review." *Journal of Immunoassay and Immunochemistry*. 2001. Vol. 22 No. 3. pp. 183-201 (In publication).
23. Remington, R.D. and Schork, M.A. "Statistics with Applications to the Biological and Health Sciences." 1970. Prentice-Hall Inc., Englewood Cliffs, N.J. USA.
24. Fleiss, J.L "Statistical Methods for Rates and Proportions." *2nd Edition, Wiley Series in Probability and Mathematical Statistics*. 1981. New York, USA.
25. Erdreich, L.S. and Lee, E.T. "Use of Relative Operating Characteristic Analysis in Epidemiology: A Method for Dealing with Subjective Judgement." *American Journal of Epid.* Vol. 114 No. 5. 1981. pp. 649-662.
26. Swets, J. A. "Measuring the Accuracy of Diagnostic Systems." *Science* Vol. 240 No.4857. 1988. pp. 1285-1293.
27. Schoojans, F., Zalata, A., Depuydt, C.E. and Comhaire, F.H. "MedCalc: A New Computer Program for Medical Statistics." *Computer Methods and Programs in Biomedicine*. Vol. 48. 1995. pp. 257-262.
28. Swets, J. A. "ROC Analysis Applied to the Evaluation of Medical Imaging Techniques." *Invest. Radiol*ogy 1979. Vol. 14. pp. 109-121.
29. Zweig, M. H. and Campbell, G. "Receiver-Operating Characteristic (ROC) Plots: a Fundamental Evaluation Tool in Clinical Medicine." *Clinical Chemistry*. Vol. 39. 1993. pp. 561-577.
30. Griner, P. F., Mayewski, R. J., Mushlin, A. I. and Greenland, P. "Selection and Interpretation of Diagnostic Tests and Procedures." *Annals of Internal Medicine*. Vol. 94 No.4. 1981. pp. 555-600.
31. Martin, S. W., Meek, A. H. and Willeberg, P. (1987). Veterinary Epidemiology, Principles and Methods. *Iowa State University Press*. Ames, Iowa, USA. ISBN 0-8138-1856-7.
32. McNab, B. (1997). Basic principles of evaluating test performance for making decisions: Ministry of Agriculture Food and Rural Affairs, Epidemiology and Risk Assessment, Guelph, Ontario, Canada. http://www.gov.ca/omafra/english/research/risk/frameworks/asum2.html.
33. Swets, J. A., Dawes, R. M. and Monahan. "Better Decisions Through Science." *Scientific American.* 2000. pp. 82-85.
34. Green, C. "Biological Method Validation: A Practical Approach." *Journal of Validation Technology.* 1998. Vol. 4 No. 3. pp. 217-22
35. Noy, R. J. "Quality Assurance of Validation Data." *Journal of Validation Technology.* 1997. Vol. 3 No. 2. pp. 148-156.
36. Possa, C. "How Much Validation is Enough? The Balancing Act." *Journal of Validation Technology.* 1997. Vol. 3 No. 2. p. 188.
37. Mathers, D. D., Marshall, R. T., Caillibot, P. F. and Phipps, M. J. (1994). The ISO 9000 Essentials: A Practical Handbook for Implementing the ISO 9000 Standards. Canadian Standards Association, Mississauga, Ontario, Canada. ISBN 0-921347-40-5.
38. General Requirements for the Competence of Testing and Calibration Laboratories (2000). Document CAN-P-4D (ISO 17025). Standards Council of Canada, Ottawa, Ontario, Canada.
39. Ferrante, M. "A Simple Way to Establish Acceptance Criteria for Validation Studies." *Journal of Validation Technology* 1999. Vol. 3 No. 4.

Article Acronym Listing

AUC:	Area Under the ROC Curve
CELISA:	Competitive Enzyme Immunoassay
CL:	Confidence Limits
FDA:	Food and Drug Administration
FPA:	Fluorescence Polarization Assay
IELISA:	Indirect Enzyme Immunoassay
ISO:	International Organization for Standardization
RBPT:	Rose Bengal Plate Test
ROC:	Receiver Operating Characteristics
TAT:	Tube Agglutination Test
USAHA:	United States Animal Health Association
USDA:	United States Department of Agriculture

PO Box 6004, Duluth, MN 55806
Phone: 218-723-9477 or 888-524-9922 U.S. Only
Fax: 218-723-9308
E-Mail: info@ivthome.com Web Site: www.ivthome.com